

Divergence of Catalytic Mechanism within a Glycosidase Family Provides Insight into Evolution of Carbohydrate Metabolism by Human Gut Flora

Tracey M. Gloster,^{1,*} Johan P. Turkenburg,¹ Jennifer R. Potts,^{1,2} Bernard Henrissat,³ and Gideon J. Davies¹

¹York Structural Biology Laboratory, Department of Chemistry

²Department of Biology

University of York, Heslington, York, YO10 5YW, UK

³Architecture et Fonction des Macromolécules Biologiques UMR6098 CNRS, Universités Aix-Marseille I & II, case 932, 163 Avenue de Luminy, 13288 Marseille cedex 9, France

*Correspondence: gloster@ysbl.york.ac.uk

DOI 10.1016/j.chembiol.2008.09.005

SUMMARY

Enzymatic cleavage of the glycosidic bond yields products in which the anomeric configuration is either retained or inverted. Each mechanism reflects the dispositions of the enzyme functional groups; a facet of which is essentially conserved in 113 glycoside hydrolase (GH) families. We show that family GH97 has diverged significantly, as it contains both inverting and retaining α -glycosidases. This reflects evolution of the active center; a glutamate acts as a general base in inverting members, exemplified by *Bacteroides thetaiotaomicron* α -glucosidase BtGH97a, whereas an aspartate likely acts as a nucleophile in retaining members. The structure of BtGH97a and its complexes with inhibitors, coupled to kinetic analysis of active-site variants, reveals an unusual calcium ion dependence. ¹H NMR analysis shows an inversion mechanism for BtGH97a, whereas another GH97 enzyme from *B. thetaiotaomicron*, BtGH97b, functions as a retaining α -galactosidase.

INTRODUCTION

Glycoside hydrolases (GHs), the enzymes responsible for cleavage of glycosidic bonds in mono-, di- and polysaccharides, are ubiquitous throughout all kingdoms of life, where they play myriad biological roles. GHs catalyze cleavage of the glycosidic bond with either net retention or inversion of anomeric configuration. The majority of the enzymes characterized to date tend to adopt the “classical” mechanisms seminally first proposed by Koshland (1953). Inversion of stereochemistry at the anomeric position involves a single-step mechanism (Figure 1A), in which protonation of the glycosidic oxygen by the general acid and leaving group departure is concomitant with nucleophilic attack of a water molecule that has been deprotonated by the general base. Typically, the general acid and base are enzymatic residues bearing a carboxylate group. Retention of stereochemistry involves a double displacement mechanism via a covalent intermediate (Figure 1B). During the first step, protonation of

the glycosidic oxygen by the acid/base (acting as an acid) to aid leaving group departure is concomitant with nucleophilic attack by the nucleophile residue at the anomeric carbon, which leads to the formation of the covalent intermediate. Subsequently, the acid/base (acting as a base) deprotonates a nucleophilic water molecule (Zechel and Withers, 2000). A particularly rich source of glycosidases, within which one can study the evolution of carbohydrate active enzyme systems and their mechanisms, is the bacterium *Bacteroides thetaiotaomicron*. Analysis of the *B. thetaiotaomicron* genome has demonstrated that the organism possesses a significant number of genes (6.6% of the entire genome) that encode carbohydrate-processing enzymes, including many GHs; this is the largest proportion of genome devoted to carbohydrate metabolism of any organism sequenced to date (Davies et al., 2005).

B. thetaiotaomicron, a Gram-negative anaerobe, is a symbiotic component of the microflora in the human gut; the flora perform a range of beneficial metabolic tasks that are not encoded by our own genome (Comstock and Coyne, 2003; Xu et al., 2003; Zocco et al., 2007). *B. thetaiotaomicron* plays a major role in the breakdown of polysaccharides ingested in the diet into a form that could otherwise not be absorbed and utilized by the host (Comstock and Coyne, 2003; Xu et al., 2003; Zocco et al., 2007). Perhaps the most well-characterized polysaccharide utilization machinery in *B. thetaiotaomicron* is that encoded by the *sus* operon. The *sus* operon contains seven genes (*susA*–*susG*) that encode proteins involved in the utilization and metabolism of starch (D’Elia and Salyers, 1996; Reeves et al., 1997), and is regulated by the activator SusR in response to high levels of glucose oligomers (Cho et al., 2001). *susA* encodes a neopullanase, which hydrolyzes the α -1,4 linkages in starch to produce smaller oligosaccharides; *susB* encodes an α -glucosidase, which is able to break down these smaller oligomers into glucose; *susC*–*susF* encode outer membrane proteins involved in starch binding; and *susG* encodes a protein, the role of which is unclear, but which has high sequence similarity to amylases (D’Elia and Salyers, 1996; Reeves et al., 1997).

Carbohydrate active enzymes have been classified into families based upon their amino acid sequence similarities (Coutinho and Henrissat, 1999). Currently there are 113 sequence-distinct glycosidase families. A feature of almost all classical CAZY families is that, since sequence dictates structure, and structure

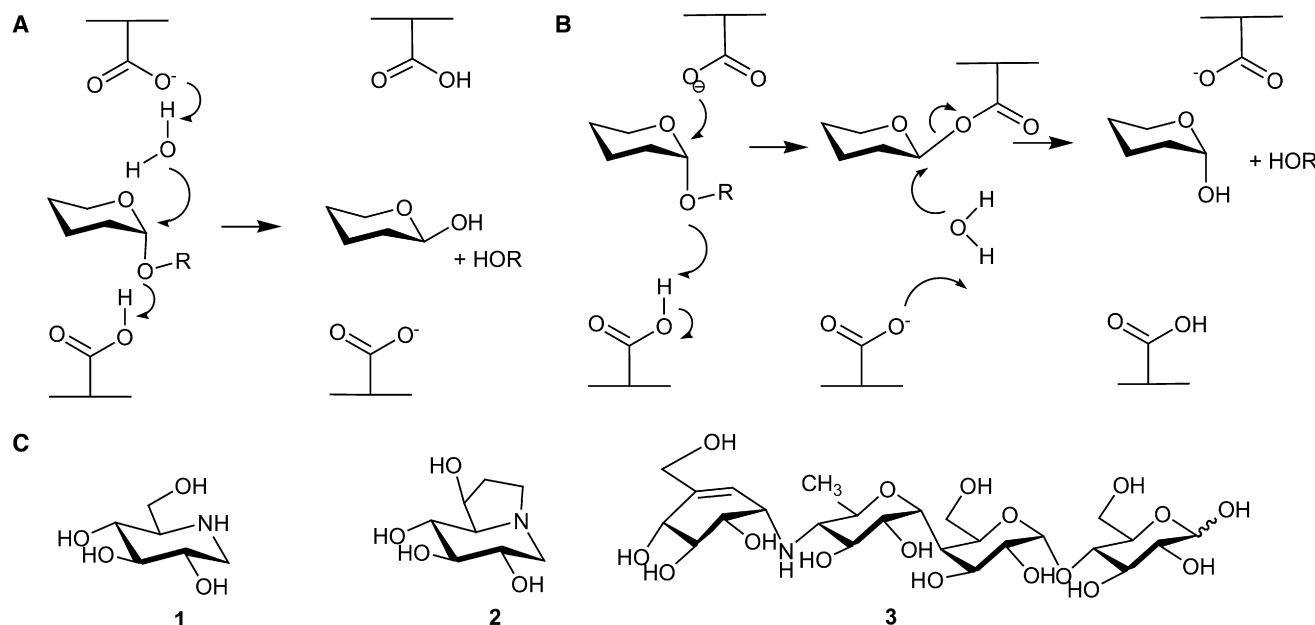


Figure 1. GH Mechanisms and α -Glucosidase Inhibitors

(A and B) Hydrolysis with (A) inversion and (B) retention of anomeric configuration.
(C) Inhibitors deoxynojirimycin (1), castanospermine (2), and acarbose (3).

determines function, the catalytic mechanism is conserved within a sequence-based family (Henrissat and Davies, 1997). Exceptions to this rule are rare and unusual: GH4 and GH109 enzymes are not classical hydrolases, but instead use NAD^+ in a transient reduction/oxidation reaction with leaving group elimination (Rajan et al., 2004), and GH23 is a family with both inverting and retaining hexosaminidases, but the catalytic mechanism of neither is understood, and may involve substrate participation in catalysis. The α -glucosidase encoded by the *susB* gene belongs to family GH97 (Hughes et al., 2003) of this Carbohydrate Active enZymes (CAZy) classification (<http://www.cazy.org/>), the catalytic mechanism of which was unknown prior to this work, but had been predicted to be retaining in an insightful bioinformatics analysis (Naumoff, 2005). *B. thetaiotaomicron* possesses a total of 10 GH97 members, and there are currently 69 other bacterial GH97 open reading frames, and one from the archaea *Haloarcula marismortui*. The goal of our work was to probe the structural and functional diversity within this family in order to shed light on the evolution and diversity of carbohydrate metabolism in the human gut.

Here we describe the three-dimensional structure and mechanistic characterization of the enzyme encoded by *susB*, a GH97 α -glucosidase from *B. thetaiotaomicron* (BtGH97a). The native structure, together with three-dimensional complexes with the inhibitors deoxynojirimycin (1) and castanospermine (2), in addition to kinetics of wild-type and mutant enzymes, reveals an unusual metal ion-dependent catalysis. Structural similarity searches confirm the earlier sequence predictions (Naumoff, 2005) that the enzyme is related to family GH27 retaining α -galactosidases (see Fujimoto et al., 2003; Garman et al., 2002). Close inspection of the structural overlap, however, revealed no obvious candidate nucleophile, and ^1H NMR

analysis showed that BtGH97a is an inverting enzyme, consistent with the observed disposition of catalytic groups and in contrast to earlier predictions made *in silico*. Examination of GH97 sequences, in light of the structure, suggests an early evolutionary divergence into two main subfamilies, in only one of which the catalytic constellation presented by BtGH97a is conserved. The second subfamily is shown to be a subfamily of retaining enzymes, demonstrated here for the *B. thetaiotaomicron* α -galactosidase BtGH97b by ^1H NMR analysis. Family GH97 shows a clear evolutionary divergence into both inverting and retaining enzymes, which now, by virtue of the three-dimensional structure and chemical analysis, allows prediction of catalytic mechanism for other family members.

RESULTS

Family GH97 comprises around 80 members at present, 10 of which are from *B. thetaiotaomicron*. Two members, SusB (hereafter, BtGH97a) from *B. thetaiotaomicron* (Smith and Salyers, 1991) and an enzyme from *Tannerella forsythensis* (Hughes et al., 2003) have been investigated previously, and both have been shown to possess α -glucosidase activity. BtGH97a was shown to hydrolyze a range of substrate sizes from maltose to maltoheptaose, and both α -1,4 and α -1,6 linkages (Smith and Salyers, 1991).

BtGH97a Three-Dimensional Structure and Activity

The structure of BtGH97a was solved using X-ray crystallography data to 1.90 Å resolution (Table 1). Phasing was performed using single-wavelength anomalous dispersion data collected at the peak wavelength (0.9795 Å) for selenium, and the majority of the protein was built automatically. The tertiary structure of

Table 1. Data Processing and Refinement Statistics for *BtGH97a*

Data Collection ^a	<i>BtGH97a</i> (SeMet)	<i>BtGH97a</i> (Native)	<i>BtGH97a</i> in Complex with 1	<i>BtGH97a</i> in Complex with 2
Resolution (Å)	15–2.50 (2.59–2.50)	15–1.90 (1.97–1.90)	15–1.66 (1.72–1.66)	20–2.00 (2.07–2.00)
Space group	<i>P</i> 2 ₁	<i>P</i> 2 ₁	<i>P</i> 2 ₁	<i>P</i> 2 ₁
Unit cell parameters				
<i>a</i> , <i>b</i> , <i>c</i> ; β	75.6, 111.9, 102.5; 100.8°	75.6, 111.6, 102.4; 100.9°	75.7, 112.2, 102.5; 100.9°	74.6, 112.0, 104.1; 100.9°
<i>R</i> _{merge}	0.084 (0.16)	0.084 (0.28)	0.075 (0.35)	0.076 (0.35)
Mean <i>I</i> /σ	15.9 (8.6)	8.9 (2.3)	20.0 (3.4)	13.3 (2.2)
Completeness (%)	100 (99)	91 (57)	99 (93)	95 (69)
Multiplicity	5.3 (5.0)	2.3 (1.7)	5.4 (3.3)	3.5 (2.5)
No. unique reflections	57,815	120,776	195,947	108,971
<i>R</i> _{cryst}	—	0.15	0.15	0.17
<i>R</i> _{free}	—	0.20	0.19	0.23
Rmsd bonds (Å)	—	0.014	0.014	0.015
Rmsd angles (°)	—	1.41	1.41	1.50
Rmsd chiral volume (Å ³)	—	0.103	0.109	0.105
PDB ID code	—	2JKA	2JKE	2JKP

Values in parentheses represent outer shell. Rmsd, root-mean-square deviation.

^a Beamline (ESRF) ID23-1.

BtGH97a reveals three domains (Figure 2A); an N-terminal β-super-sandwich domain (residues 1–317), followed by a canonical (β/α)₈ barrel (residues 318–636) and a C-terminal β sheet domain (residues 637–738). A calcium ion is present in the (β/α)₈ barrel domain; it is octahedrally coordinated by four glutamate residues (Glu194, Glu508, Glu526, and Glu532) and both oxygen atoms of a molecule of ethylene glycol, which was sequestered from the cryoprotectant. The calcium ions in each molecule of the asymmetric unit in the native structure (and the complexes described below) yield atomic B factors of 14–16 Å² when refined with unitary occupancy. Superposition with other (β/α)₈ barrel GHs (data not shown), as well as the presence of the four glutamates, suggests that the calcium ion is, in fact, bound in the active-site region of the enzyme. We initially assumed that it would be inhibitory, as is often seen when cations bind in place of the cationic transition state (Brzozowski et al., 2000; Tailford et al., 2007).

The calcium ion displays high apparent affinity and may only partially be removed using EDTA or Chelex. Partial removal of the calcium causes a substantial reduction in activity, but the enzyme maintains a residual activity, confirming high metal ion affinity. Other divalent ions, including magnesium and zinc, caused inactivation of the enzyme. The addition of calcium to the EDTA-treated enzyme resulted in restoration of full activity; all further kinetics experiments were thus conducted in the presence of 10 mM CaCl₂, at which the enzyme had full activity.

The pH dependence of *k*_{cat}/*K*_M for *BtGH97a* was fitted to a bell-shaped ionization curve (see Supplemental Data available online), which gave *pK*_a values of 5.5 and 7.7; all subsequent kinetics were performed at the enzyme optimum, pH 6.6. Kinetic parameters for *BtGH97a* were determined with the aryl α-glycosides dinitrophenyl α-D-glucopyranoside (DNP-Glc) and *p*-nitrophenyl α-D-glucopyranoside (pNP-Glc), and with the α-1,4 linked disaccharide maltose, Table 2. The *k*_{cat}/*K*_M was highest for the most activated substrate, DNP-Glc (~2000 s⁻¹ mM⁻¹), followed by pNP-Glc (620 s⁻¹ mM⁻¹), but only modest for maltose (0.0072 s⁻¹ mM⁻¹). The lower activity on maltose may reflect

the enzyme's preference for longer substrates (Smith and Salyers, 1991).

Probing *BtGH97a* through Inhibition

The inhibition of *BtGH97a* using classical α-glycosidase inhibitors yielded *K*_i values of 59 μM for deoxynojirimycin **1**, 59 μM for castanospermine **2**, and 108 nM for acarbose **3**. Isofagomine, calystegine B₂, and isofagomine lactam did not inhibit the enzyme at 100 μM. The structures of *BtGH97a* in complex with **1** and **2** were determined at 1.7 and 2.0 Å resolution, respectively. Clear 2*F*_{obs} – *F*_{calc} electron density in the active site (of both molecules of the asymmetric unit) allowed facile building of the complexes (Figures 2B and 2C). Both **1** and **2** are found in a slightly flattened (toward ⁴E) ⁴C₁ chair conformation. Each of the inhibitors and the surrounding active-site residues superpose perfectly, so only the interactions for *BtGH97a* in complex with **1** are described (Figure 2E). Hydrogen bonds are formed between the hydroxyl group at C6 and Glu439, between the hydroxyl at C4 and Glu391 and Trp331, and between the hydroxyl at C3 and Lys467. The hydroxyl group at the C2 position interacts with the calcium ion and with His507. The ring nitrogen atom of **1** interacts with a water molecule, which is also within hydrogen bonding distance of Glu439 and Glu508, and hence may mimic the catalytic water molecule in an inverting mechanism (described below in light of ¹H NMR experiments). The calcium ion assumes an octahedral coordination, with six ligands, as was observed in the native structure. It coordinates the same four glutamate residues (Glu194, Glu508, Glu526, and Glu532), but instead of interacting with an ethylene glycol molecule, it now coordinates the hydroxyl group at C2 of **1** and a water molecule (all at a distance of between 2.3 and 2.5 Å; see Supplemental Data).

Dissection of *BtGH97a* Mechanism Using Active-Site Variants

Each of the glutamate residues in the unusual active-site constellation of *BtGH97a* (Glu194, Glu439, Glu508, Glu526,

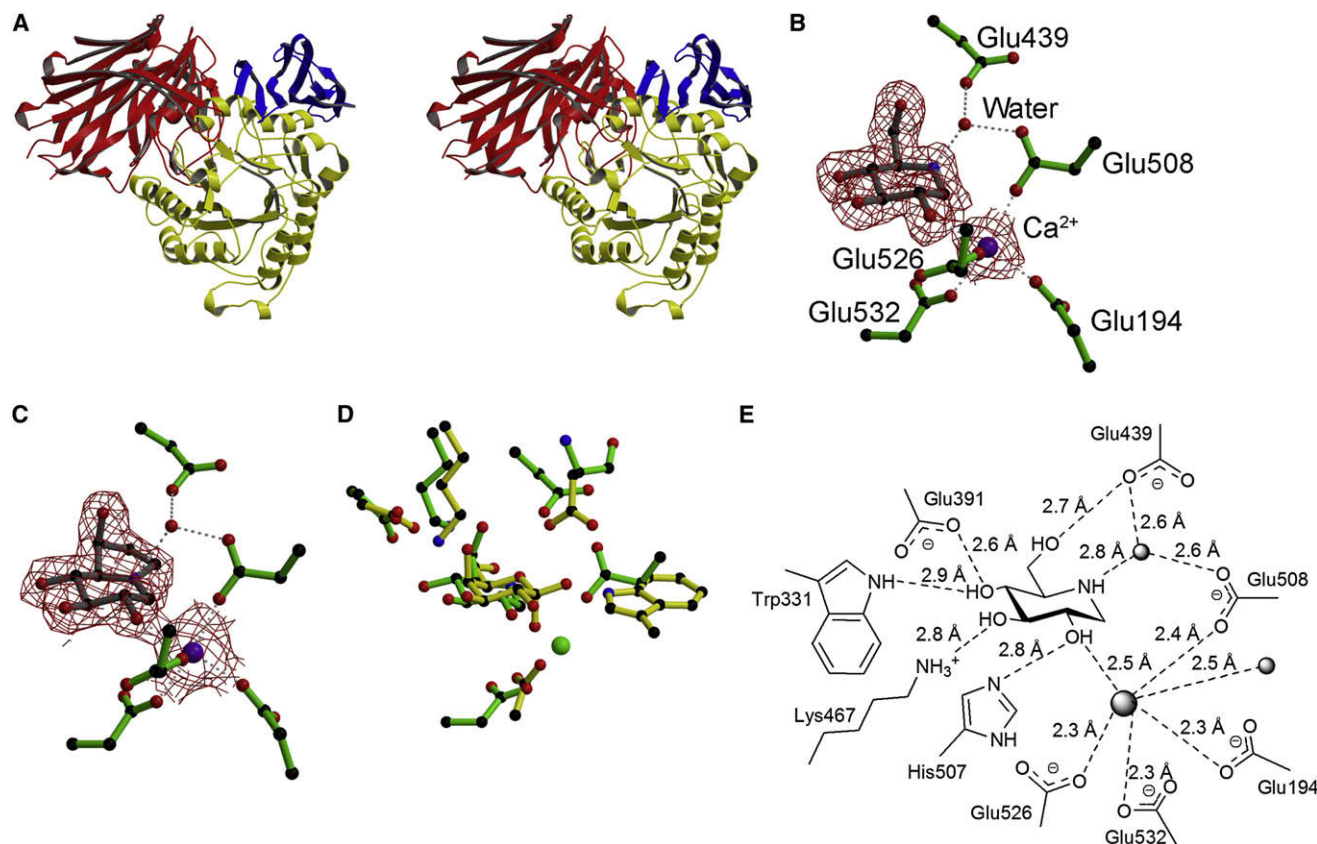


Figure 2. Structural Insights into *BtGH97a*

(A) Divergent stereo ribbon representation of *BtGH97a*; the N-terminal domain is in red, the core (β/α)₈ domain in yellow, and the C-terminal domain in blue. (B and C) Ball-and-stick representation of *BtGH97a* in complex with (B) **1** and (C) **2**; observed electron density for the maximum likelihood weighted $2F_{\text{obs}} - F_{\text{calc}}$ map is contoured at 1σ . The purple spheres represent calcium ions. (D) Ball-and-stick representation of the active-site overlap between *BtGH97a* and a GH27 enzyme (PDB ID code **1UAS**). Figures were drawn using BOBSCRIPT (Esnouf, 1997). (E) Interactions between *BtGH97* and **1**.

and Glu532) were mutated in turn to alanine. Where activity allowed, full Michaelis-Menten kinetics were performed using *p*NP-Glc as a substrate. Additionally, substrate depletion methods, using DNP-Glc as a substrate, were used to determine the $k_{\text{cat}}/K_{\text{M}}$ directly (Table 2). Mutations of Glu194 and Glu532 allowed measurable activity with *p*NP-Glc; in both cases, the K_{M} was higher and the k_{cat} lower than for the wild-type enzyme. Overall the $k_{\text{cat}}/K_{\text{M}}$ was reduced around 70-fold for E194A and 80-fold for E532A. Activity using DNP-Glc as a substrate could be measured for all mutants except E508A. The $k_{\text{cat}}/K_{\text{M}}$ was reduced around 7-fold for E194A, 250,000-fold for E439A, 40-fold for E526A, and 4-fold for E532A. These values are discussed below in light of further mechanistic investigations.

GH97 Structural and Sequence Alignments

As predicted in an insightful bioinformatics analysis (Naumoff, 2005), searches for the nearest structural homologs to *BtGH97a* using SSM (Krissinel and Henrick, 2004) revealed the closest matches were to GHs from family 36 (an α -galactosidase from *Thermotoga maritima*; PDB ID code, 1ZY9 [unpublished]) and family 27 (including an α -galactosidase from rice; PDB ID code,

1UAS [Fujimoto et al., 2003]). Although *BtGH97a* shares low sequence similarity with the sequences from GH27 and GH36, *BtGH97a* superposes with a root-mean-square deviation between $C\alpha$ atoms of 3.3 Å over 391 residues with the closest GH36 structure (1ZY9) and 3.3 Å over 277 residues with the closest GH27 structure (1UAS). GH27 and GH36 both belong to clan GH-D in the CAZy classification (Coutinho and Henrissat, 1999), and are known to perform catalysis with net retention of anomeric configuration (Comfort et al., 2007).

A global overlap of *BtGH97a* with the GH27 α -galactosidase homolog from rice, for which there is a ligand complex (PDB ID code, 1UAS), reveals partial conservation of the active-site residues (Figure 2D). The galactose present in the GH27 structure overlays well with **1** bound in the active site of *BtGH97a*. Asp391, Lys467, and Glu532 from *BtGH97a* superpose well with Asp51, Lys128, and Asp185 (the catalytic acid/base), respectively, from the GH27 enzyme. There are no reports of a metal ion bound in the active site of the GH27 or GH36 enzymes, and, although Glu532 appears in a similar position to the acid/base residue from the GH27 enzyme, the other three glutamate residues that coordinate the calcium ion are not conserved.

Table 2. Kinetic Data for BtGH97a with DNP-Glc, pNP-Glc and Maltose, BtGH97b with pNP-Gal and Melibiose, and BtGH97a Mutants with pNP-Glc and DNP-Glc

Enzyme/Mutant	Substrate	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M ($s^{-1} \text{ mM}^{-1}$)
BtGH97a	DNP-Glc	0.14 ± 0.01	280 ± 7	1997
	pNP-Glc	0.085 ± 0.009	53 ± 2	620
	Maltose ^a	2.0 ± 0.4	0.015 ± 0.002	0.0072
BtGH97b	pNP-Gal	0.23 ± 0.01	91 ± 2	391
	Melibiose	1.5 ± 0.17	0.015 ± 0.0004	0.011
BtGH97a				
E194A	pNP-Glc	0.31 ± 0.02	2.8 ± 0.07	9.2
E532A	pNP-Glc	0.14 ± 0.02	1.1 ± 0.04	7.7
Wild-type	DNP-Glc	—	—	1807 ± 68
E194A	DNP-Glc	—	—	255 ± 3
E439A	DNP-Glc	—	—	0.007 ± 0.001
E508A	DNP-Glc	—	—	ND
E526A	DNP-Glc	—	—	48 ± 4
E532A	DNP-Glc	—	—	402 ± 29

Kinetic data for BtGH97a with DNP-Glc, pNP-Glc and maltose, BtGH97b with pNP-Gal and melibiose, and BtGH97a mutants with pNP-Glc were determined using full Michaelis-Menten kinetics, and those for BtGH97a mutants with DNP-Glc were determined using substrate depletion methods. Where no data are reported for BtGH97a mutants with pNP-Glc, no activity could be measured. ND, none detected.

^aSubstrate inhibition was observed. The K_i value for maltose was 17.2 mM.

There is also a large difference in the region surrounding the catalytic nucleophile (Asp130) in the GH27 enzyme, where, in the equivalent position in BtGH97a was, to our surprise, a glycine residue (Gly469).

This surprising lack of a candidate catalytic nucleophile for BtGH97a made the prediction of the stereochemical outcome unclear. The distance between the two key catalytic residues in a three-dimensional structure (if identified) are often useful in suggesting the mechanism adopted by the enzyme; in the case of retaining enzymes, the distance between the two residues is, on average, 5.5 Å, whereas it is often larger for inverting enzymes, as both the substrate and water need to fit into the active site simultaneously (Zechel and Withers, 2000). However, the number of glutamate residues in the active site of BtGH97a makes it difficult to predict the mechanism for this family with certainty; hence, the stereochemistry was determined unambiguously by NMR.

Determination of BtGH97a Mechanism

The stereochemistry of BtGH97a hydrolysis using ^1H NMR was determined by following the product of the enzymatic reaction and its subsequent mutarotation to the thermodynamically favored ratio of α - and β -anomers. BtGH97a hydrolysis of pNP-Glc was monitored over 17 hr (Figure 3A). Only the β -anomer of glucose was observed to accumulate initially; as mutarotation occurred, however, the amount of the β -anomer began to decline and the α -anomer appeared, until the standard equilibrium of around 64% β -anomer/36% α -anomer was reached. The data unequivocally demonstrate that the reaction hydrolyzed by BtGH97a proceeds with inversion of stereochemistry at the anomeric position to generate the β -D-glycoside as product. The three-dimensional structure had already suggested the likely candidate for the BtGH97a catalytic base as Glu439. Glu439 sits

on the “ β -face,” and, in the complexes with **1** and **2**, it coordinates a water molecule in a position for in-line attack at the anomeric center (Figure 2E). In addition, the E439A mutant was 250,000-fold less efficient at hydrolyzing DNP-Glc than the wild-type enzyme, which is consistent with loss of base catalytic assistance to water attack. The comparison with the GH27 retaining enzymes shows that Glu439 is ~ 1.6 Å “further away” from the anomeric center compared with the aspartate nucleophile (Asp130) in the GH27 enzyme. Thus, this simple shift in the position of the carboxylate provides the extra space required for a single-step inverting mechanism.

Assignment of the catalytic acid is more troublesome, given the constellation of glutamates in the active site. Glu532 is isostructural with the known acid/base residue in the GH27 enzyme, suggesting that it may act as the catalytic acid. Furthermore, the E532A mutant had a far greater effect on pNP-Glc hydrolysis (80-fold) compared with the 4-fold change on DNP-Glc; given the greatly reduced need for protonic assistance with the latter substrate in these kinetics, and in light of the overlap with the GH27 enzyme, there is compelling evidence for the role of Glu532 as the catalytic acid.

Relating Sequence to Mechanism in GH97

Sequence alignments of the whole GH97 family gave rather intriguing results when the positions of the five active-site glutamate residues were examined. The residues equivalent to Glu526 and Glu532 in BtGH97a (both coordinate the calcium ion, and Glu532 is predicted to be the catalytic acid) were absolutely conserved throughout all members. The residue equivalent to Glu194, which is also important in coordinating the calcium ion, is also highly conserved, implying that calcium dependence may be universal in this family. Intriguingly, comparison of the residues equivalent to Glu439 in BtGH97a,

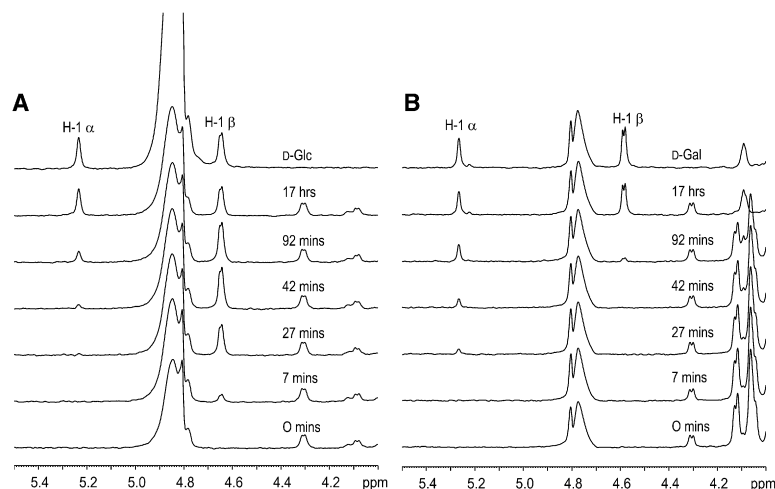


Figure 3. ^1H NMR Analysis of Stereochemical Outcome

(A and B) Hydrolysis using (A) *BtGH97a* with pNP-Glc and (B) *BtGH97b* with pNP-Gal. The times indicate the beginning of the spectra acquisition, and peaks for the α - and β -anomers are shown (assigned according to Roslund et al. [2008]).

predicted to be the catalytic base, shows that, in around half of the family members, this is in a conserved HHET motif; however, in the other members, there is a highly variable sequence in this region. Given our proposal that Glu439 is the key catalytic base in the inversion mechanism, and it is almost essential for catalysis in *BtGH97a*, this observation suggests a different catalytic mechanism for some of the other family members. Furthermore, the vast majority of these other GH97 members, which lack the putative base, have their own signature motif, (K/M)(I/V)DF, with the aspartate lying in exactly the same position as the nucleophilic aspartate (Asp130 in the PDB ID code 1UAS) of their GH27 relatives. A phylogenetic tree derived from the multiple-sequence alignment of 79 family GH97 members corroborates this finding with the clear definition of two subfamilies. In the *BtGH97a*-like members (which possess a putative base for inversion), a glycine (Gly469 in *BtGH97a*) in the motif K(T/S)GY is found where the aspartate residue is found in the second subfamily (Figure 4). Analysis of the whole family thus suggests that around half of the GH97 members possess a catalytic base, and would likely act with inversion by virtue of identity to *BtGH97a*, but that an Gly-Asp change in the other half of the family would generate, by analogy to GH27, a catalytic constellation consistent with retention via a covalent glycosyl-enzyme intermediate.

Mechanistic Studies on a Second GH97 Enzyme

To ascertain whether an enzyme that possessed the “opposite” set of motifs to *BtGH97a* did in fact proceed with retention of configuration, as sequence alignments would suggest, we cloned and overexpressed another open reading frame encoding a GH97 enzyme from *B. thetaiotaomicron* (*BtGH97b*). This enzyme, however, did not hydrolyze DNP-Glc or pNP-Glc, but was shown to be active on *p*-nitrophenyl α -D-galactopyranoside (pNP-Gal), revealing that it is an α -galactosidase. It did, however, display an identical metal ion dependency profile to *BtGH97a*. The pH dependence of k_{cat}/K_M for *BtGH97b* was more difficult to ascertain than for *BtGH97a*, as activity levels appeared to change very little with pH values between 4 and 8; kinetics were therefore performed at pH 6.6 (the optimum for *BtGH97a*). Although it is often dangerous to interpret pH rate profiles (Knowles, 1976), the flattening here may be related to

the unusual presence of a catalytic calcium ion, which will impact on the pK_a of the acid/base and nucleophile and any potential Lewis-acid assistance to leaving group departure. Kinetic parameters for *BtGH97b* were determined with the aryl α -glycoside pNP-Gal and the α -1,6 disaccharide melibiose (Table 2). The catalytic efficiency, reflected by the k_{cat}/K_M , was higher for pNP-Glc ($391 \text{ s}^{-1} \text{ mM}^{-1}$) than for melibiose ($0.011 \text{ s}^{-1} \text{ mM}^{-1}$).

BtGH97b hydrolysis of pNP-Gal was monitored over 17 hr by ^1H NMR (Figure 3B). During the first hour, the α -anomer of galactose accumulated, with no evidence of any β -anomer, demonstrating that α -galactose was the initial product from the α -galactoside substrate, and that the enzyme indeed acted with retention of anomeric configuration.

DISCUSSION

GH97 enzymes can be classified into two broad subfamilies of inverting and retaining enzymes (Figure 4; see also the Supplemental Data), breaking the dogma that stereochemistry is conserved within a classical GH family; this also contradicts previous bioinformatics studies, which predicted a retaining mechanism for hydrolysis (Naumoff, 2005). Our studies also demonstrate an expansion in the substrate specificity previously determined for GH97 enzymes, as previously only α -glucosidase activity had been reported (Hughes et al., 2003; Smith and Salyers, 1991). As only two were tested here, and *B. thetaiotaomicron* alone possesses 10 GH97 enzymes, it suggests that there may be further substrate diversity to discover within this fascinating family. Each of the enzymes was shown to be extremely specific for either glucose- (*BtGH97a*) or galactose- (*BtGH97b*) containing substrates, demonstrating absolute selectivity at the C4 position. Furthermore, GH97 enzymes appear to possess strong metal ion dependence, which is rare among GHs. GH2 *Escherichia coli* β -galactosidase (*lacZ*) is magnesium ion-dependent (Juers et al., 2001; Richard et al., 1996; Selwood and Sinnott, 1990; Sinnott and Withers, 1978), and family 38 and 47 α -mannosidases are zinc and calcium ion dependent, respectively. These α -mannosidases are observed to coordinate a metal ion to the hydroxyl groups at the C2 and C3 positions of inhibitors, which is thought to play a role in substrate specificity and/or distortion (Karaveg et al., 2005; Vallée et al., 2000; van den Elsen et al., 2001). The calcium ion in GH97 enzymes, however, may provide acid assistance to the catalytic acid during hydrolysis, as has been suggested for the magnesium ion in the *E. coli* β -galactosidase. The presence of interactions between the calcium ion and four glutamate residues in *BtGH97a*, as well as the hydroxyl group at C2 of the inhibitors, suggests

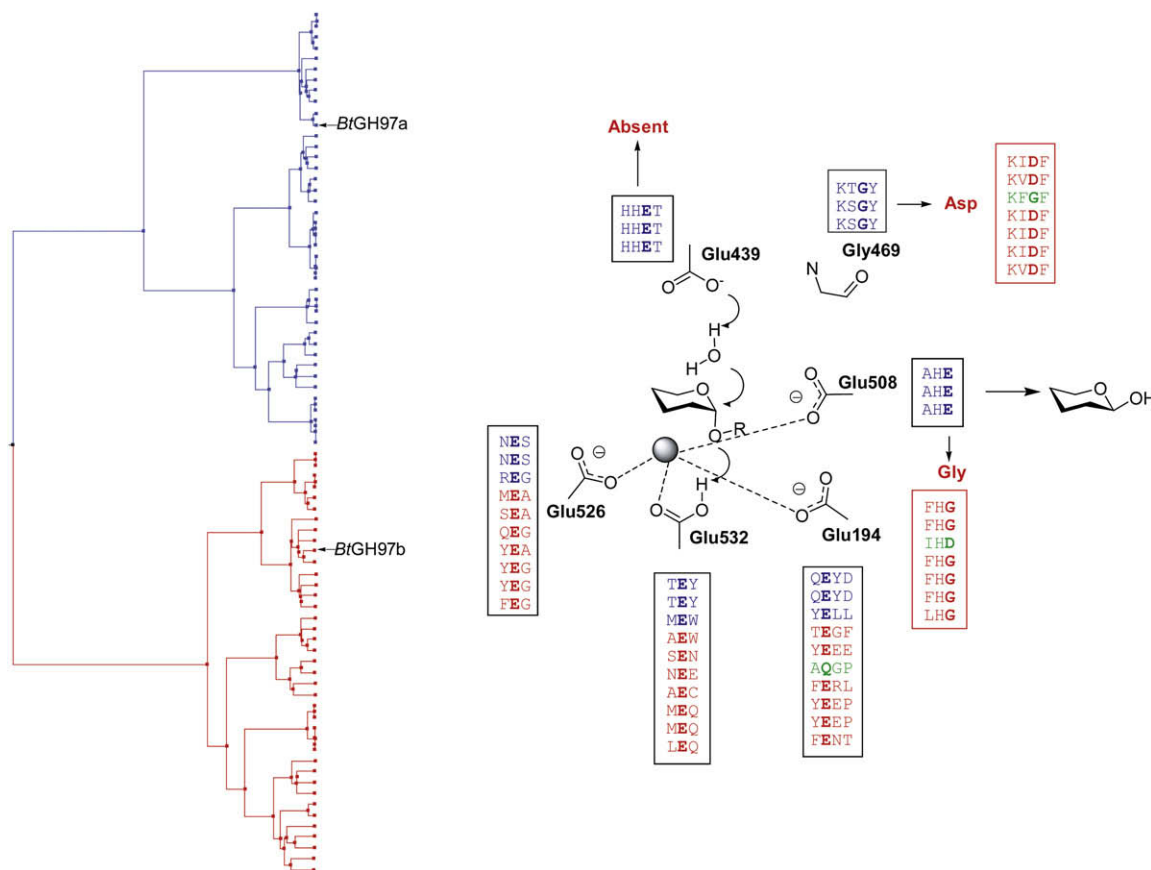


Figure 4. GH97 Family Tree, Active Center Geometry, and Sequence Alignments for the Key Active-Site Residues of the *B. thetaiotaomicron* Representatives

Inverting subfamily sequences are shown in blue and retaining subfamily sequences in red. The inverting-retaining transition reflects a loss of the inverting base and its replacement by a Gly-Asp shift elsewhere in the sequence. GH97 has six unusual sequences, which appear to contain neither inverting nor retaining catalytic signatures; the *B. thetaiotaomicron* outlier is shown in green. The full annotated family tree is given in [Supplemental Data](#).

that the enzyme has evolved a way to keep the ion bound tightly and that it is vital to function.

The power of sequence alignments, in conjunction with the structure of *BtGH97a* and experimental demonstration of reaction stereochemistry, has demonstrated beautifully how the consensus motifs surrounding a few key residues can be used to predict the mechanism for hydrolysis for the majority of the GH97 enzymes. Our analysis, however, indicates there are six sequence “outliers” which fall into the part of the tree containing the enzymes which proceed with retention, but where the sequences appear not to contain the consensus motifs consistent with retaining or inverting enzymes; one of these sequences is from *B. thetaiotaomicron*, as highlighted in green in Figure 4. At the position equivalent to Gly469 in *BtGH97a*, inverting enzymes possess the motif K(T/S)GY, whereas retaining enzymes contain (K/M)(I/V)DF. The six outliers possess the motif KXG(Y/F), which implies that they lack the catalytic nucleophile. Similarly, at the position equivalent to Glu508 in *BtGH97a*, inverting enzymes have a conserved HE motif, whereas retaining enzymes have HG; in the six exceptions, one possesses an HE motif, while the other five sequences have an HD motif. Examination of the residues equivalent to Glu439 in *BtGH97a* shows that the inverting

enzymes possess an absolutely conserved HHET, whereas the sequence is apparently random for the retaining enzymes at this position. Strangely, however, of the six exceptions, although one possesses an HNET motif, the rest have a variable sequence, suggesting that they also lack the residue equivalent to the catalytic base for the inverting mechanism; hence, they appear to lack both the base and the enzyme-derived nucleophile. In addition, while the glutamate equivalent to Glu194 in *BtGH97a* is a highly conserved glutamate residue in both inverting and retaining enzymes, the six outlier sequences all possess a glutamine residue at this position. These intriguing observations suggest that these members may either be inactive or that they have evolved yet another catalytic apparatus, adding to the remarkable plasticity of family GH97. Inactive glycosidase variants that have evolved to perform other activities have been described in other families, most notably GH18 (Durand et al., 2005) and GH13 (Bröer and Wagner, 2002; Janeček et al., 1997), and bioinformatics analysis has predicted that the same may be true of GH57 (Zona et al., 2004).

Simple repositioning of a base or nucleophile has been reported as a route to “man-made” changes in reaction stereochemistry, notably by Wang et al. (1994) and Kuroki et al. (1995).

These authors insightfully foresaw a route to mechanistic change that only now has been shown to have evolved in nature by these GH97 enzymes. One further case in which two families, related by sequence and structure, showed different stereochemistries was reported by [Alberto et al. \(2004\)](#); the positioning of the catalytic apparatus, however, is invariant, and the change in mechanism appears to be achieved by subtle repositioning of the sugar itself.

GH97 is thus an extremely rare family of GHs in which enzymes have unambiguously been shown to proceed with both inversion and retention of stereochemistry. In light of the three-dimensional structure and in conjunction with superpositions with other clan members, careful dissection of nuances in the sequence alignments has aided this discovery. In all other families, with the exception of GH4 (which are NAD⁺ dependent and not classical hydrolases) and GH23 (where the basis for retention is not understood), the mechanism is conserved. The metal dependence for the GH97 enzymes indicates a special feature that possibly helps to create an environment that supports both the retaining and inverting mechanisms. Evolutionary pressure must have conferred an advantage to similar enzymes with the ability to utilize mechanisms with different stereochemical outcomes. It is only as we begin to unlock these subtleties and secrets that nature has imposed during evolution that we can begin to predict enzyme specificity and mechanism with any certainty in the future.

SIGNIFICANCE

Cleavage of the glycosidic bond, catalyzed by glycoside hydrolases (GHs), yields one of two stereochemical outcomes in which configuration of the reactive anomeric center is either retained or inverted. These two reactions follow one of two canonical mechanisms featuring double displacement via a covalent intermediate, or a single nucleophilic displacement by water, respectively. The difference in mechanism reflects the dispositions of the important functional groups within the active site of the enzyme, which is essentially conserved within the 113 sequence-based GH families. Our studies demonstrate that family GH97 enzymes have diverged significantly from this general model, as it contains both inverting and retaining α -glycosidases. This divergence of mechanism reflects a comparatively small change in the active center, in which a glutamate plays the role as a general base in inverting members, whereas, in retaining members, an aspartate residue from a different place in the sequence likely acts as the nucleophile. The three-dimensional structure and kinetic analyses of *BtGH97a* reveal an unusual calcium ion-dependent catalytic center in which a glutamate residue is poised to activate the nucleophilic water molecule with both general acid and calcium ion assistance to leaving group departure. Bioinformatics analysis of the whole sequence family reveals an evolutionary split between inverting and retaining family members, reflecting subtle changes to the active-center environment, which is possibly aided by the metal ion. In addition, these sequence alignments reveal an intriguing six sequences that appear to possess neither the base for inversion nor the nucleophile for retention of configuration.

Evolutionary pressures must have encouraged these subtle differences between closely related enzymes to confer a biological advantage on those able to hydrolyze substrates to both α - and β -anomer products. It is only as we dissect these variations imposed by nature that mechanisms and specificity can be predicted with certainty in the future.

EXPERIMENTAL PROCEDURES

Gene Cloning, Protein Production, and Protein Purification

The genes encoding *BtGH97a* and *BtGH97b* (GenBank accession numbers [AA078808](#) and [AA076978](#), respectively) were amplified using the polymerase chain reaction from *B. thetaiotaomicron* VPI-5482 genomic DNA using primers that gave ligation-independent cloning (LIC) compatible ends. These were ligated into an LIC-modified pET28a vector using standard procedures ([Bonsor et al., 2006](#)). Protein production and purification was identical for each protein. Plasmid containing the gene of interest was transformed into *E. coli* BL21 (DE3) cells, and cultured in 0.5 L autoinduction media ([Studier, 2005](#)) supplemented with 50 $\mu\text{g ml}^{-1}$ kanamycin at 37°C for 8 hr. Protein production was induced at 30°C overnight. Cells were harvested and resuspended in 20 mM HEPES, pH 7, 150 mM NaCl, and lysed by sonication. The supernatant was applied to a 5 ml HisTrap nickel-Sepharose column (GE Healthcare), preequilibrated in the same buffer, and the protein eluted from an imidazole gradient. The protein was dialyzed to remove the imidazole, concentrated, and purified further on an S200 16/60 gel filtration column, preequilibrated in 20 mM HEPES, pH 7, 150 mM NaCl. Selenomethionine-containing *BtGH97a* was obtained and purified in the same way as described for the native protein, except that the autoinduction media contained selenomethionine (SeMet).

Crystallization

BtGH97a, at 10 mg ml⁻¹, was crystallized from 18%–22% polyethylene glycol 3350 and 0.02 M sodium/potassium phosphate. Crystals were cryoprotected in a solution containing the relevant mother liquor with the addition of 25% ethylene glycol, and were flash frozen in liquid nitrogen. Crystals were also grown in the presence of 2.5 mM **1** or **2** using the same crystallization conditions.

Data Collection, Structure Solution, and Model Refinement

Data for both the native and the SeMet *BtGH97a* were collected from single crystals at 100K on beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Data were integrated and scaled using HKL2000 ([Otwinowski and Minor, 1997](#)), and all computing used the CCP4 suite of programs (CCP4, 1994), unless otherwise stated. Phasing from SeMet residues was performed in SOLVE ([Terwilliger and Berendzen, 1999](#)), and initial model building was carried out with RESOLVE ([Terwilliger, 2003](#)). The output model from RESOLVE was entered as a starting model in ARP/wARP ([Lamzin and Wilson, 1993](#)), which was used to trace the majority of the protein chain. Subsequent refinement involved alternate rounds of model building in COOT ([Emsley and Cowtan, 2004](#)) and likelihood-based refinement in REFMAC ([Murshudov et al., 1997](#)). Subsequent to our structure determination, the structure of native SusB was also released (PDB ID code [2D73](#)) by M. Kitamura, M. Okuyama, Y. Kitago, F. Tanzawa, H. Mori, A. Kimura, M. Yao, and I. Tanaka. Data for *BtGH97a* in complex with **1** and **2** were also collected on beamline ID23-1 at the ESRF, and were handled in the same way as described for the native data.

Site-Directed Mutagenesis of *BtGH97a*

Mutations (E194A, E439A, E508A, E526A, and E532A) were introduced into the gene encoding *BtGH97a* using a Quikchange site-directed mutagenesis kit (Stratagene) by the standard protocol described in the kit manual. Resulting plasmids were sequenced (Applied Biosystems 3130 sequencer) to check that the desired mutation had been incorporated. Mutant proteins were produced and purified using the same protocol as described for the native protein.

Enzyme Kinetics

Methods for the enzyme kinetics are described in full in the [Supplemental Data](#). Briefly, kinetic studies involving a spectrophotometric product were

conducted by monitoring the change in UV/visible absorbance with a Cintra 10 spectrophotometer. All experiments were carried out at 37°C, in a total volume of 1 ml and in the presence of 10 mM CaCl₂. Enzyme dependence on pH, ranging from pH 4 to 8, was carried out using substrate depletion methods to determine the k_{cat}/K_M directly with pNP-Glc for BtGH97a or pNP-Gal for BtGH97b as the substrate. *p*-Nitrophenolate release was monitored continuously at 400 nm and the data fitted to a first-order rate equation using GRAFIT (Leatherbarrow, 2001). k_{cat}/K_M values at different pH values were fitted to a bell-shaped ionization curve to determine the pK_a values of the ionizable groups. Individual kinetic parameters were determined at the pH optimum for catalysis using DNP-Glc or pNP-Glc for BtGH97a and the E194A and E532 mutants, and pNP-Gal for BtGH97b. 2,4-Dinitrophenolate/*p*-nitrophenolate release was monitored at 400 nm and the rates determined. Kinetic parameters were determined by fitting the data to the Michaelis-Menten equation in GRAFIT. Substrate depletion methods were used to determine the k_{cat}/K_M values for each of the five mutants with DNP-Glc. K_i values for 1–3 with BtGH97a were determined under steady-state conditions at low substrate concentration using pNP-Glc. Activity was monitored at 400 nm for 300 s, and the rates determined. The fractional decrease of the rate in the absence and presence of inhibitor was used to determine the K_i , which was taken as the mean of triplicate readings.

Kinetics were also performed using a stopped assay, where glucose was detected using glucose oxidase/peroxidase linking enzymes (Megazyme). Measurements were made with maltose as the substrate for BtGH97a and melibiose for BtGH97b. Aliquots were taken at time intervals over 10 min, boiled to inactivate the enzyme, and 1 ml of the glucose oxidase/peroxidase solution was added. Assays were otherwise performed according to the manufacturer's instructions. Reaction rates were determined and the data fitted to the Michaelis-Menten equation in GRAFIT. Substrate inhibition was observed with BtGH97a and fitted appropriately.

NMR

¹H NMR spectra were acquired on a Bruker 700 MHz spectrometer at 25°C, with presaturation for water suppression. Samples had been lyophilized and diluted in 100% ²H₂O twice prior to the experiment. Substrates pNP-Glc for BtGH97a and pNP-Gal for BtGH97b were present at a final concentration of 0.5 mM. Assays were run in 10 mM maleate buffer, pH 6.6, 10 mM CaCl₂, with the addition of 75 μ M trimethylsilylpropionic acid as an internal chemical shift reference. Substrate hydrolysis was initiated by the addition of enzyme (in the same buffer) at a final concentration of approximately 15 nM. Spectra (with 16 scans, 16,384 data points, and a 16.5 s relaxation delay between scans) were collected at regular intervals for 17 hr, at which time hydrolysis appeared to be complete. Control spectra of the substrate in the absence of enzyme, and of 0.5 mM glucose and 0.5 mM galactose in 100% ²H₂O, were also acquired.

Bioinformatics

Sequences were retrieved from the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>) from their accession number found in CAZy family GH97 (<http://www.cazy.org/fam/GH97.html>). The amino acid sequences were aligned using MUSCLE (Edgar, 2004). The aligned sequences were clustered using the SECATOR algorithm (Wicker et al., 2001), which relies on BIONJ (Gascuel, 1997) to build a tree from the multiple-sequence alignment, and subsequently collapses the branches from subtrees after identification of the nodes joining different subtrees (Wicker et al., 2001). The neighbor-joining tree was made from the resulting distance matrix using Blossum62 substitution parameters. Visualization of the tree was accomplished with an in-house program.

ACCESSION NUMBERS

Coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 2JKA, 2JKE, and 2JKP.

SUPPLEMENTAL DATA

Supplemental Data include three figures and Supplemental Experimental Procedures and can be found with this article online at <http://www.chembiol.com/cgi/content/full/15/10/1058/DC1/>.

ACKNOWLEDGMENTS

We thank the Biotechnology and Biological Sciences Research Council for funding. T.M.G. is a Sir Henry Wellcome postdoctoral fellow and G.J.D. is a recipient of a Royal Society/Wolfson Research Merit Award.

Received: June 23, 2008

Revised: September 1, 2008

Accepted: September 11, 2008

Published online: October 9, 2008

REFERENCES

- Alberto, F., Bignon, C., Sulzenbacher, G., Henrissat, B., and Czjzek, M. (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.
- Bonsor, D., Butz, S.F., Solomons, J., Grant, S., Fairlamb, I.J., Fogg, M.J., and Grogan, G. (2006). Ligation independent cloning (LIC) as a rapid route to families of recombinant biocatalysts from sequenced prokaryotic genomes. *Org. Biomol. Chem.* 4, 1252–1260.
- Bröer, S., and Wagner, C.A. (2002). Structure-function relationships of heterodimeric amino acid transporters. *Cell Biochem. Biophys.* 36, 155–168.
- Brzozowski, A.M., Lawson, D.M., Turkenburg, J.P., Bisgaard-Frantzen, H., Svendsen, A., Borchert, T.V., Dauter, Z., Wilson, K.S., and Davies, G.J. (2000). Structural analysis of a chimeric bacterial alpha-amylase. High-resolution analysis of native and ligand complexes. *Biochemistry* 39, 9099–9107.
- Cho, K.H., Cho, D., Wang, G.-R., and Salyers, A.A. (2001). New regulatory gene that contributes to control of *Bacteroides thetaiotaomicron* starch utilization genes. *J. Bacteriol.* 183, 7198–7205.
- CCP4 (Collaborative Computational Project, Number 4). (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 50, 760–763.
- Comfort, D.A., Bobrov, K.S., Ivanen, D.R., Shabalin, K.A., Harris, J.M., Kulminskaya, A.A., Brumer, H., and Kelly, R.M. (2007). Biochemical analysis of *Thermotoga maritima* GH36 alpha-galactosidase (TmGalA) confirms the mechanistic commonality of clan GH-D glycoside hydrolases. *Biochemistry* 46, 3319–3330.
- Comstock, L.E., and Coyne, M.J. (2003). *Bacteroides thetaiotaomicron*: a dynamic, niche-adapted human symbiont. *Bioessays* 25, 926–929.
- Coutinho, P.M., and Henrissat, B. (1999). Carbohydrate-active enzymes: an integrated database approach. In *Recent advances in carbohydrate bioengineering*, H.J. Gilbert, G.J. Davies, B. Henrissat, and B. Svensson, eds. (Cambridge: Royal Society of Chemistry), pp. 3–12.
- D'Elia, J.N., and Salyers, A.A. (1996). Contribution of a neopullulanase, a pullulanase, and an alpha-glucosidase to growth of *Bacteroides thetaiotaomicron* on starch. *J. Bacteriol.* 178, 7173–7179.
- Davies, G.J., Gloster, T.M., and Henrissat, B. (2005). Recent structural insights into the expanding world of carbohydrate-active enzymes. *Curr. Opin. Struct. Biol.* 15, 637–645.
- Durand, A., Hughes, R., Roussel, A., Flatman, R., Henrissat, B., and Juge, N. (2005). Emergence of a subfamily of xylanase inhibitors within glycoside hydrolase family 18. *FEBS J.* 272, 1745–1755.
- Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.
- Esnouf, R.M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* 15, 132–134.
- Fujimoto, Z., Kaneko, S., Momma, M., Kobayashi, H., and Mizuno, H. (2003). Crystal structure of rice alpha-galactosidase complexed with D-galactose. *J. Biol. Chem.* 278, 20313–20318.

- Garman, S.C., Hannick, L., Zhu, A., and Garboczi, D.N. (2002). The 1.9 Å structure of alpha-N-acetylgalactosaminidase: molecular basis of glycosidase deficiency diseases. *Structure* 10, 425–434.
- Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* 14, 685–695.
- Henrissat, B., and Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* 7, 637–644.
- Hughes, C.V., Malki, G., Loo, C.Y., Tanner, A.C.R., and Ganeshkumar, N. (2003). Cloning and expression of alpha-D-glucosidase and N-acetyl-beta-glucosaminidase from the periodontal pathogen, *Tannerella forsythensis* (*Bacteroides forsythus*). *Oral Microbiol. Immunol.* 18, 309–312.
- Janeček, S., Svensson, B., and Henrissat, B. (1997). Domain evolution in the alpha-amylase family. *J. Mol. Evol.* 45, 322–331.
- Juers, D.H., Heightman, T.D., Vasella, A., McCarter, J.D., Mackenzie, L., Withers, S.G., and Matthews, B.W. (2001). A structural view of the action of *Escherichia coli* (*lacZ*) beta-galactosidase. *Biochemistry* 40, 14781–14794.
- Karaveg, K., Siriwardena, A., Tempel, W., Liu, Z.-J., Glushka, J., Wang, B.-C., and Moremen, K.W. (2005). Mechanism of class I (glycosylhydrolase family 47) alpha-mannosidases involved in N-glycan processing and endoplasmic reticulum quality control. *J. Biol. Chem.* 280, 16197–16207.
- Knowles, J.R. (1976). The intrinsic pK_a -values of functional groups in enzymes: improper deductions from the pH-dependence of steady-state parameters. *CRC Crit. Rev. Biochem.* 4, 165–173.
- Koshland, D.E., Jr. (1953). Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev. Camb. Philos. Soc.* 28, 416–436.
- Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2256–2268.
- Kuroki, R., Weaver, L.H., and Matthews, B.W. (1995). Structure-based design of a lysozyme with altered catalytic activity. *Nat. Struct. Biol.* 2, 1007–1011.
- Lamzin, V.S., and Wilson, K.S. (1993). Automated refinement of protein models. *Acta Crystallogr. D Biol. Crystallogr.* 49, 129–147.
- Leatherbarrow, R.J. (2001). GraFit Version 5 (Horley, UK: Erithacus Software Ltd.).
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53, 240–255.
- Naumoff, D.G. (2005). GH97 is a new family of glycoside hydrolases, which is related to the alpha-galactosidase superfamily. *BMC Genomics* 6, 112–123.
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307–326.
- Rajan, S.S., Yang, X., Collart, F., Yip, V.L., Withers, S.G., Varot, A., Thompson, J., Davies, G.J., and Anderson, W.F. (2004). Novel catalytic mechanism of glycoside hydrolysis based on the structure of an NAD^+/Mn^{2+} -dependent phospho-alpha-glucosidase from *Bacillus subtilis*. *Structure* 12, 1619–1629.
- Reeves, A.R., Wang, G.-R., and Salyers, A.A. (1997). Characterization of four outer membrane proteins that play a role in utilization of starch by *Bacteroides thetaiotaomicron*. *J. Bacteriol.* 179, 643–649.
- Richard, J.P., Huber, R.E., Lin, S., Heo, C., and Amyes, T.L. (1996). Structure-reactivity relationships for beta-galactosidase (*Escherichia coli*, *lacZ*). 3. Evidence that Glu-461 participates in Brønsted acid-base catalysis of beta-D-galactopyranosyl group transfer. *Biochemistry* 35, 12377–12386.
- Roslund, M.U., Tähtinen, P., Niemitz, M., and Sjöhol, R. (2008). Complete assignments of the 1H and ^{13}C chemical shifts and $J_{H,H}$ coupling constants in NMR spectra of D-glucopyranose and all D-glucopyranosyl-D-glucopyranosides. *Carbohydr. Res.* 343, 101–112.
- Selwood, T., and Sinnott, M.L. (1990). A solvent-isotope-effect study of proton transfer during catalysis by *Escherichia coli* (*lacZ*) beta-galactosidase. *Biochem. J.* 268, 317–323.
- Sinnott, M.L., and Withers, S.G. (1978). The necessity of magnesium cation for acid assistance of aglycone departure in catalysis by *Escherichia coli* (*lacZ*) beta-galactosidase. *Biochem. J.* 175, 539–546.
- Smith, K.A., and Salyers, A.A. (1991). Characterization of a neopullulanase and an alpha-glucosidase from *Bacteroides thetaiotaomicron* 95-1. *J. Bacteriol.* 173, 2962–2968.
- Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* 41, 207–234.
- Tailford, L.E., Money, V.A., Smith, N.L., Dumon, C., Davies, G.J., and Gilbert, H.J. (2007). Mannose foraging by *Bacteroides thetaiotaomicron*: structure and specificity of the β -mannosidase, Man2a. *J. Biol. Chem.* 282, 11291–11299.
- Terwilliger, T.C. (2003). Improving macromolecular atomic models at moderate resolution by automated iterative model building, statistical density modification and refinement. *Acta Crystallogr. D Biol. Crystallogr.* 59, 1174–1182.
- Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 55, 849–861.
- Vallée, F., Karaveg, K., Herscovics, A., Moremen, K.W., and Howell, P.L. (2000). Structural basis for catalysis and inhibition of N-glycan processing class I alpha-1,2-mannosidases. *J. Biol. Chem.* 275, 41287–41298.
- van den Elsen, J.M.H., Kuntz, D.A., and Rose, D.R. (2001). Structure of Golgi alpha-mannosidase II: a target for inhibition of growth and metastasis of cancer cells. *EMBO J.* 20, 3008–3017.
- Wang, Q., Graham, R.W., Trimbur, D., Warren, R.A.J., and Withers, S.G. (1994). Changing enzymatic reaction mechanisms by mutagenesis: conversion of a retaining glucosidase to an inverting enzyme. *J. Am. Chem. Soc.* 116, 11594–11595.
- Wicker, N., Perrin, G.R., Thierry, J.C., and Poch, O. (2001). Secator: a program for inferring protein subfamilies from phylogenetic trees. *Mol. Biol. Evol.* 18, 1435–1441.
- Xu, J., Bjursell, M.K., Himrod, J., Deng, S., Carmichael, L.K., Chiang, H.C., Hooper, L.V., and Gordon, J.I. (2003). A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science* 299, 2074–2076.
- Zechel, D.L., and Withers, S.G. (2000). Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc. Chem. Res.* 33, 11–18.
- Zocco, M.A., Ainora, M.E., Gasbarrini, G., and Gasbarrini, A. (2007). *Bacteroides thetaiotaomicron* in the gut: molecular aspects of their interaction. *Dig. Liver Dis.* 39, 707–712.
- Zona, R., Chang-Pi-Hin, F., O'Donohue, M.J., and Janeček, S. (2004). Bioinformatics of the glycoside hydrolase family 57 and identification of catalytic residues in amylopullulanase from *Thermococcus hydrothermalis*. *Eur. J. Biochem.* 271, 2863–2872.