Mechanistic and Structural Analysis of a Family 31 α-glycosidase and its Glycosyl-Enzyme Intermediate

Andrew L. Lovering¹, Seung Seo Lee², Young-Wan Kim², Stephen G. Withers², Natalie C. J. Strynadka¹

¹Department of Biochemistry and Molecular Biology, University of British Columbia, 2146 Health Sciences Mall, Vancouver, BC, V6T 1Z3, Canada

²Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, BC, V6T 1Z1, Canada

Corresponding Author: Natalie Strynadka

Tel: 604-822-0789
Fax: 604-822-5227
natalie@byron.biochem.ubc.ca

Running Title: Structure of a family 31 α-glycosidase
SUMMARY

We have determined the first structure of a family 31 α-glycosidase, that of YicI from *E.coli*, both free and trapped as a 5-fluoroslyropyranosyl enzyme intermediate via reaction with 5-fluoro-α-D-xylopyranosyl fluoride. Our 2.2 Å resolution structure shows an intimately associated hexamer with structural elements from several monomers converging at each of the 6 active sites. Our kinetic and mass spectrometric analysis verifies several of the features observed in our structural data including a covalent linkage from the carboxylate side chain of the identified nucleophile D416 to the C1 of the sugar ring. Structure-based sequence comparison of YicI with the mammalian α-glucosidases lysosomal α-glucosidase and sucrase-isomaltase predicts a high level of structural similarity and provides a foundation for understanding the various mutations of these enzymes that elicit human disease.
ABBREVIATIONS

PNP, para-nitrophenol; ax-5αXF, (5S)-5-fluoro-α-D-xylopyranosyl fluoride; eq-5αXF, (5R)-5-fluoro-α-D-xylopyranosyl fluoride; TOF, time of flight; PEG, polyethylene glycol; MPD 2-methyl-2,4-pentanediol; PCMB, para-chloromercuribenzoate; MAD, multiwavelength anomalous dispersion; MIR, multiple isomorphous replacement; NCS, non-crystallographic symmetry; GH, glycosyl hydrolase; RMS, root mean squared; CAZY, Carbohydrate Active enzymes; ASU, asymmetric unit.
INTRODUCTION

Glycoside hydrolases play critical roles in biology ranging from digestion and decomposition of polysaccharides to the biosynthesis of glycoproteins. Gene sequences of many thousands of these important enzymes have now been determined and the corresponding enzymes grouped into families on the basis of sequence similarity(1-5). These are widely available at the following website (http://afmb.cnrs-mrs.fr/CAZY/).

The α-glucosidases are a particularly important sub-set of these enzymes, both in primary metabolism and in glycoconjugate biosynthesis and processing. These enzymes are principally found within families 13 and 31, and to a lesser extent in families 4 and 63. Family 13 contains a wide range of glucoside processing enzymes, including the α-amylases and cyclodextrin glucanotransferases. Correspondingly, enzymes of this family have attracted considerable attention, with numerous mechanistic studies and with 3-dimensional structures having been known for some 20 years. Family 31, by contrast, has received relatively little attention despite its importance and the number of different activities represented from a range of organisms including animals, plants and microorganisms(6,7). This family contains such important α-glucosidases as the human lysosomal α-glucosidase whose deficiency results in Pompe’s disease (also known as glycogen storage disease type 2 or acid maltase deficiency, the marked feature of which is the accumulation of glycogen in heart and skeletal muscle cells), the endoplasmic reticulum glucosidase II that plays a key role in glycoprotein processing and folding, and the digestive enzyme sucrase-isomaltase that is the target of inhibition by the anti-diabetes drugs acarbose and miglitol. It also contains α-xylosidases, isomaltosyl
transfertases and the mechanistically interesting α-glucan lyases, which carry out an elimination reaction rather than hydrolysis.

Despite the importance of the family, mechanistic insights are limited. The enzymes are known to be retaining α-glycosidases, which hydrolyse the glycosidic bond with net retention of anomeric configuration via an acid/base catalysed mechanism involving a covalent glycosyl-enzyme intermediate. Through a range of studies involving affinity labeling, trapping of reaction intermediates and site-directed mutagenesis the catalytic nucleophile has been identified as an aspartic acid within the consensus sequence WIDMNE (D224 for the *A. niger* α-glucosidase,(6-14)). On the basis of sequence comparisons and kinetic analysis of mutants the acid/base catalyst has been tentatively assigned as an aspartic acid residue (D647 in SPGase, the *Schizosaccharomyces pombe* α-glucosidase(14)). Highly oxocarbenium ion-like transition states are suggested by the large α-secondary kinetic isotope effects seen for both α-glucosidases and the α-glucan lyase(15-17), as well as by the tight binding observed for azasugar transition state analogue inhibitors such as nojirimycin and acarbose(17-21). However, a key missing component, which would be particularly important for inhibitor design, is the 3-dimensional structure of any member of this family.

Inspection of the complete genome sequence of *Escherichia coli* revealed it contains two genes encoding enzymes from Family 31, *yicl* and *yihQ*, but no information was available concerning the function of each gene product. These proteins should be good model enzymes for both mechanistic and structural analysis, being devoid of the post-
translational modifications that can complicate such studies in proteins from higher order organisms. They are therefore attractive targets for investigation.

This manuscript describes the cloning, expression, kinetic and mechanistic characterization and also the 3-dimensional structural analysis of one of these two enzymes, YicI. The results obtained confirm previous predictions of key catalytic residues in this family and provide a structural framework both for the design of new α-glucosidase inhibitors and for the understanding of mutations to the lysosomal α-glucosidase that result in Pompe’s disease.\(^1\)

Footnote 1. During the preparation of this manuscript another paper describing the cloning, expression and preliminary characterisation of YicI appeared, but included no structural or mechanistic analysis(22).
MATERIALS AND METHODS

General Procedures

All $^1$H nuclear magnetic resonance (NMR) spectra were recorded at either 200 or 300 MHz and $^{19}$F NMR spectra at 188 MHz or 282 MHz using a Bruker AC-200 or AV-300 spectrometer. $^{19}$F NMR spectra were referenced to CF$_3$COOH. Mass spectra of small molecules were recorded using a PE-Sciex API 300 triple quadrupole mass spectrometer (Sciex, Thornhill, Ont, Canada) equipped with an electrospray ionization ion source. The Bradford assay (23) was used for the determination of protein concentration. Agarose (Bio-Rad, CA, USA) gels (0.8% w/v) were used for all electrophoresis of DNA. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE(24)) was performed to analyze the homogeneity of proteins.

$\alpha$-D-Xylopyranosyl fluoride was synthesized according to the published method(25). 4-Nitrophenyl $\alpha$-D-xylopyranoside (PNP$\alpha$Xyl) was purchased from Aldrich Chemical Company. Syntheses of (5S)- and (5R)-5-fluoro-$\alpha$-D-xylopyranosyl fluorides (ax-5F$\alpha$XF and eq-5F$\alpha$XF, respectively) were achieved via radical bromination of per-O-acetyl-$\alpha$-D-xylopyranosyl fluoride followed by displacement with fluoride using AgBF$_4$, then deprotection using NH$_3$/MeOH. Full details will be provided elsewhere. Pepsin (from porcine mucosa) was purchased from Boehringer Mannheim. Yeast extracts and Tryptone were purchased from Difco Laboratory, MI, USA. 40% Acrylamide/N,N’-methylene bisacrylamide and N,N,N’,N’-tetramethylethylenediamine (TEMED) were purchased.
from Bio-Rad, CA, USA. All other chemicals and reagents were purchased from Sigma Chemical Co. unless otherwise noted.

**Cloning of the yicI Gene**

The genomic DNA of *E. coli* K12 was isolated using a QIAmp tissue kit (Qiagen). The *yicI* gene was amplified by PCR with *Pwo* polymerase (Roche) using the chromosomal DNA of *E. coli* as a template and the two synthetic primers [YicITOP (5’- CAG AAC TAA GGA ACG CAT ATG AAA ATT AGC -3’) and YicIEND (5’- ATC AAG CTC GAG CAA CGT AAT TGT CAG CGC -3’)] designed based on the sequence of the *yicI* gene (Genbank number AE000443). The PCR product was digested with *Nde*I and *Xho*I and ligated with pET29a (Novagen) that had been digested with the corresponding restriction enzymes. After sequence analysis one clone without any PCR error was designated as pET29YICI and was used for the production of 6×His-tagged *YicI* protein. The synthesis of PCR primers and the analysis of DNA sequences were carried out by the Nucleic Acid and Peptides Service Unit in the Biotechnology Laboratory at the University of British Columbia.

**Expression and Purification of YicI**

Plasmids containing the *yicI* gene (pET29YICI) were transformed into *E. coli* BL21 (DE3) and the resulting cells cultured at 37˚C in 1 liter of Luria-Bertani (LB) media containing 20 µg/ml kanamycin overnight. The overnight culture was induced by the addition of 0.3 mM isopropyl β-D-thiogalactopyranoside directly and grown for a further 5 hrs. The cell pellet was harvested and resuspended in 50 mM Tris-HCl (pH 8.0)
containing 5 mM imidazole and 300 mM NaCl. The cell suspension was passed twice through a French press at 5°C, centrifuged at 10,000 g for 30 min and the soluble extract was purified by Ni chelation chromatography. The protein was then dialyzed against 50 mM phosphate buffer (pH 7.0) and stored at 4°C.

The tryptic fragments of the protein were analyzed by electrospray ionization mass spectrometry to confirm the identity of the protein. The protein (1 mg/mL) was incubated with 10 units of trypsin in pH 7.0 phosphate buffer (1 mL) for 1 hour. The resulting digest was subjected to liquid chromatography (LC)/electrospray ionization mass spectrometry (ESI MS) analysis and masses of fragments were analyzed.

**Enzyme Kinetics**

All experiments were carried out at 37 °C in pH 7.0, 0.05 M phosphate buffer containing 0.1% BSA unless otherwise noted. Michaelis-Menten parameters for the p-nitrophenyl glycosides were measured by monitoring the release of p-nitrophenol in the presence of enzyme at 400 nm with a Varian Cary 300 Bio UV/VIS spectrometer equipped with a circulating water bath. Cuvettes of either 1 cm or 0.1 cm path length were used. The typical substrate concentration range employed was 0.3-3 × Kₘ. The difference in extinction coefficient, Δε, between p-nitrophenyl glycoside and p-nitrophenol at 400 nm at pH 7.0 and 37 °C was determined by measuring the difference in absorbance between fixed equal concentrations of each pair of p-nitrophenols and PNP glycosides. Kinetic parameters were obtained by direct fit of the data to the Michaelis-Menten equation using the software GraFit 4.0 (Leatherbarrow, R.J., Erithacus Software Ltd., Staines, UK).
Kinetic parameters for α-D-xylopyranosyl fluoride were determined by monitoring the release of fluoride from a range of substrate concentrations using an Orion 96-09 combination fluoride ion electrode interfaced to a computer running the LoggerPro software (Vernier Software Ltd.). Initial rates were used for the determination of kinetic parameters, which were obtained by direct fit of the data to the Michaelis-Menten equation using GraFit 4.0. Initial rates of hydrolysis of ax-5FαXF and eq-5FαXF by YicI were identical at all concentrations of ax-5FαXF (0.1 – 2 mM) and eq-5FαXF (0.1 – 1.2 mM) assayed. This rate was taken as the V\text{max} value for each substrate and each k\text{cat} value was calculated from these V\text{max} values by dividing each V\text{max} by the corresponding enzyme concentration.

**pH-Dependence Studies**

Measurement of the pH-dependent activity of YicI was carried out using α-D-xylopyranosyl fluoride as substrate and monitoring fluoride ion release. The following buffers were used: sodium acetate (pH 4.0 - 5.5), MES (pH 6.0), MOPS (pH 6.5 - pH 7.5) and Gly-Gly (pH 8.0 – pH 8.5). All buffers were 0.05 M in strength and contained 0.05 M NaCl. The substrate-depletion method was employed as follows. A solution of substrate (0.02 mM, ~50 fold lower than the value of K\text{m}) was preincubated at 37 °C, then the release of fluoride ion after addition of the enzyme was monitored using an Orion 96-09 combination fluoride ion electrode interfaced to a computer running the LoggerPro software (Vernier Software Ltd.) until at least 80% depletion of substrate had occurred. Appropriate controls confirmed that, at 37 °C, YicI is stable over the reaction time at each pH value studied. Fitting of the data to a first order rate equation (GraFit 4.0),
yielded an apparent rate constant for the reaction, from which the $k_{cat}/K_m$ value for each substrate was calculated by dividing that rate constant by the concentration of YicI. Obtained $k_{cat}/K_m$ values were then plotted versus pH and fitted to the appropriate curve using GraFit 4.0, thereby yielding apparent $pK_a$ values.

**Inhibition Studies**

Inhibition studies were performed by measuring enzyme activity in the presence of various concentrations of each inhibitor, using PNPαXyl as substrate. YicI, preincubated at 37 °C, was added to 200 µL of buffer solution containing PNPαXyl and varying amounts of inhibitors, also preincubated at the corresponding temperature. The release of p-nitrophenol was monitored spectrophotometrically at 400 nm. The experiments were repeated at different concentrations of PNPαXyl. A Dixon plot of 1/v versus inhibitor concentration for each substrate concentration intersects a line given by 1/$V_{max}$ at an inhibitor concentration equal to -$K_i$.

**Inactivation Kinetics**

For analysis of inactivation kinetics, the enzyme (final concentration 0.2 – 3.6 mg/ml) was pre-incubated with a range of concentrations of ax-5FαXF or eq-5FαXF at 37 °C or 10 °C. 10 µl aliquots of the sample were withdrawn at time intervals and added to 500 µl of 3 mM PNPαXyl pre-equilibrated at the appropriate temperature in the UV/Visible spectrometer. The residual enzyme activity at each time interval at each concentration of inactivator was measured in this way. Pseudo-first order rate constants at each inactivator concentration ($k_{obs}$) were determined by fitting each curve to a first order rate equation.
Values for the inactivation rate constant ($k_i$) and the inactivator dissociation constant ($K_i$) were determined by fitting $k_{obs}$ values and inhibitor concentrations to the following equation:

$$k_{obs} = k_i [I] / (K_i + [I]) .$$

Reactivation of the inactivated enzyme was carried out as follows. Enzyme (100 µL, 0.30 mg/mL) fully inactivated by the inactivator at 10 ºC was concentrated using 10 kDa nominal cut-off centrifugal concentrators (Amicon Corp., Danvers, MD) to a volume of approximately 40 µL and diluted with 1000 µL of buffer. This was repeated twice, and the retentate was diluted to a final volume of 100 µL. The inactivated enzyme was then incubated at 37 ºC and reactivation was monitored by removal of aliquots (10 µL) at appropriate time intervals and assayed as described above. Measured activities were corrected for decreases in activity due to denaturation over this time course using data for noninactivated control samples. The reactivation rate constant, $k_{react}$, was determined by fitting the data to a first order rate equation, as described above.

**Crystallization, Data Collection and Processing**

All YicI crystals used in this study were grown using a protein concentration of 10 mg/ml, with 1 µl of purified protein mixed with 1 µl of reservoir solution in a hanging drop experiment. Crystal form 1 was obtained with Crystal Screen II (26) solution 47 (100 mM sodium acetate pH 4.6, 2 M ammonium sulfate) at 18 ºC. Crystal form 2 was obtained with the Zeelen Screen (27) solution 39 (100 mM Tris pH 8.5, 15 % w/v PEG 1500, 5 % v/v MPD) at 8 ºC. During optimization of the growth of crystal form 1 it was
realized that the original screening solution had undergone a pH shift during storage, and that the crystals grew at pH 6.5. Both crystal forms were flash frozen in liquid nitrogen for data collection at 100 K. Crystal form 1 was cryo-protected using mother liquor supplemented with 100 mM MOPS pH 6.5 and 16 % w/v glycerol. Crystal form 2 was cryo-protected using mother liquor supplemented with a final concentration of 25 % v/v MPD. Data for all native datasets and the Ho derivative were collected on an ADSC Q210 detector at beamline 8.2.1 at the Advanced Light Source (Berkeley, CA.). Data for the La and PCMB datasets were collected on a MAR 345 detector using an in-house CuKα source. Diffracted intensities were evaluated using DENZO and SCALEPACK (28), and MOSFLM and SCALA (29). Various data manipulations were accomplished using the CCP4 program suite (29). Data collection statistics and derivatization details are given in Tables 2 and 3. Crystals of the YicI: eq-5FαXF complex were obtained in a short 3 minute soak using crystal form 1 and 0.5 mM eq-5FαXF in cryo-buffer.

**Structure Phasing and Refinement**

The structure was solved using both MAD and MIR phases with the COMBINE procedure in SOLVE (30). Briefly, MAD phases from the Ho derivative were combined with MIR phases obtained using the highly redundant La derivative as an anomalously scattering “native” dataset and the PCMB crystal soak as an isomorphous derivative. The mean figure of merit following this procedure was 0.39. Density modification using DM (29) improved the map enough to be able to assign a rough C-alpha trace for one monomer, which was then utilized to find the NCS operators. Phase extension to 2.2 Å with 6-fold averaging was performed using RESOLVE (30). The model output from
RESOLVE contained approximately 85% of the structure, with matched sequence assignment for 60% of sidechains. The remainder of the model was built manually using XFIT (31). Crystal form 2 was solved by molecular replacement with MOLREP (32), using the crystal form 1 hexamer as the search model. Refinement of all datasets was achieved using CNS (33) and REFMAC (34), with full statistics provided in Table 3. Ligand topologies and restraints were generated using the PRODRG server (35). No NCS constraints or restraints were used during refinement.

Coordinates: Coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 1XSI, 1XSJ and 1XSK.
RESULTS AND DISCUSSION

Sequence Alignment Analysis

The hypothetical protein encoded by the yicI gene from E. coli shows sequence similarity of 50% or more with hypothetical proteins from a range of organisms as well as with known glycosyl hydrolase (GH) family 31 enzymes (Figure 1). It also shows 20 – 29% similarity with mammalian, plant and fungal α-glucosidases, and plant α-xylosidases and 20% similarity with α-1,4 glucan lyases (GLases) from family 31. The consensus sequence surrounding the catalytic aspartate nucleophile in this region of plant, mammalian and fungal enzymes, including α-glucosidases and α-xylosidases, is WiDMnE, with a slight variation for GLases to WiDMnX (V or T). However, bacterial proteins are shown to have the sequence KTDFGE and are absolutely invariant. These clear distinctions between amino acid sequences of family 31 enzymes of higher organisms and bacteria suggest that the bacterial proteins took a different path early in the evolution of GH family 31(36,37).

Substrate Specificity

To investigate the enzymatic activity of the YicI protein (see methods for cloning, overexpression and purification), 12 different glycoside substrates were incubated with the protein in pH 7.0, 0.05 M phosphate buffer at 37 °C. As expected from the high sequence similarity with the GH family 31 α-xylosidase from L. pentosus, the YicI protein rapidly hydrolyses p-nitrophenyl α-D-xylopyranoside (PNPαXyl) and α-D-xylopyranosyl fluoride (αXylF) as substrates. Lower, but significant activity was
observed with $p$-nitrophenyl $\alpha$-D-glucopyranoside, while neither the aryl $\alpha$-glycosides of sugars such as galactose, mannose and arabinose, or any $\beta$-glucosides, were hydrolyzed at all (Table 1). Similarly, neither maltose or maltotriose were cleaved, though the recent paper of Okuyama et al (22) showed that isoprimeverose ($\text{Xyl} \alpha(1,6) \text{Glc}$) is an excellent substrate, consistent with the expected role of this enzyme in xyloglucan degradation. Okuyama et al also report the ability of YicI to bind disaccharide acceptors in the aglycone-binding site, raising the possibility of potential substrates with at least 3 sugar units. Values of $k_{\text{cat}}$ and $K_m$ for the hydrolysis of PNP$\alpha\text{Xyl}$ by this protein are in the same range as those for two other bacterial $\alpha$-xylosidases ($k_{\text{cat}} = 0.21 \text{ s}^{-1}$ and $K_m = 1 \text{ mM}$ for the enzyme from Bacillus sp. and $k_{\text{cat}} = 0.1 \text{ s}^{-1}$ and $K_m = 1.3 \text{ mM}$ for that from L. pentosus(38,39) When YicI was incubated with either PNP$\alpha\text{Xyl}$ or PNP$\alpha\text{Glc}$ and products were checked by TLC, only the hydrolysis products, xylose and glucose were detected, indicating the absence of lyase activity. As a result, the YicI protein has been designated as an $\alpha$-xylosidase and very similar conclusions were reached in the aforementioned work by Okuyama et al.

The Effects of Temperature, Metal Ions and pH on Activity

Hydrolysis of PNP$\alpha\text{Xyl}$ was measured at various temperatures (data not shown), with activity increasing up to 50°C and being rapidly lost at higher temperatures. YicI stored in pH 7.0, 0.05 M phosphate buffer retains full activity for 48 hours at 37 °C and for several months at 4°C. The effects of various metal ions including Ca$^{2+}$, Mg$^{2+}$, Mn$^{2+}$, Zn$^{2+}$, Ni$^{2+}$, Cu$^{2+}$ and Co$^{2+}$ on the activity of YicI were examined and none of these affected the rate of hydrolysis of PNP$\alpha\text{Xyl}$. Measurement of the pH-dependent activity of YicI was
performed using $\alpha$-d-xylopyranosyl fluoride ($\alpha$XylF) as a substrate. A classical bell-shaped dependence of $k_{cat}/K_m$ upon pH was observed as expected, indicating at least two essential, ionizable groups in the free enzyme (Figure 2). The two apparent $pK_a$ values are $pK_{a1} = 4.9 \pm 0.2$ and $pK_{a2} = 7.9 \pm 0.1$, with an estimated pH optimum of pH 6.4. These presumably correspond to the nucleophile, D416 and the acid/base catalyst, D482, respectively.

**Reversible Inhibitors of YicI**

Three different reversible inhibitors were tested with YicI (Figure 3). Among these 1-deoxyxylonojirimycin (XN) has the highest potency ($K_i = 0.9$ $\mu$M), with the gluco-analogue, 1-deoxynojirimicin (130 $\mu$M) binding almost 50-fold more weakly, consistent with the much higher rates of hydrolysis of xylosides than glucosides. Indeed the $\alpha$-glucosidase inhibitor acarbose was completely non-inhibitory. The 1-azasugar xyloisofagomine (Xyl-1-N) binds more weakly (176 $\mu$M) than the 5-azasugar, consistent with previous observations on azasugar inhibitory preferences(40).

**The Inactivation of YicI with Mechanism Based Inactivators, ax- and eq-5-Fluoro-$\alpha$-D-Xylopyranosyl Fluorides**

Two new mechanism-based inactivators, ax-5F$a$XF and eq-5F$a$XF were designed, synthesized and tested (Figure 3). Compounds of this class are designed to react as substrates, but form a relatively stable 5-fluoroxylosyl-enzyme intermediate, which hydrolyses only slowly, thus they often behave as time-dependent inactivators. However, no time-dependent inactivation was observed when ax-5F$a$XF was incubated with YicI at
37 °C, and aliquots were removed for assay of activity. However, the activity measured was lower than that in the control reaction containing no inhibitor – and was concentration-dependent. The inhibition observed must be due to ax-5FαXF carried over in the aliquot for assay – suggesting that ax-5FαXF is acting as a tight binding, reversible inhibitor. This behavior is reminiscent of that seen upon the reaction of 5-fluoro-α-glycosyl fluorides with α-glycosidases, where apparent tight binding was also observed(12,41-43), but was found to be due to the accumulation of a glycosyl-enzyme intermediate that turns over on a timescale shorter than the assay time. Thus, ax-5FαXF was tested as an apparent competitive reversible inhibitor and the (apparent) $K_i'$ value was determined to be 9.8 μM. The equatorial epimer eq-5FαXF behaved likewise, with a $K_i'$ value of 0.45 μM revealing even tighter apparent binding.

To confirm that this rapid inactivation is due to the accumulation of the covalent intermediate, reaction mixtures of both samples were subjected to analysis by electrospray ionization mass spectrometry, which revealed that the mass of the protein (89,094±8) increased by a mass of 157 (reaction with ax-5FαXF) and 158 (reaction with eq-5FαXF). This is in agreement, within error, with the expected difference for addition of the mass of a 5-fluoro-xylosyl unit (mass of 151) from ax-5FαXF and eq-5FαXF. The labeling of YicI by these two reagents indicates that the inhibition is indeed due to the accumulation of the covalent intermediate.

Time-dependent inactivation was indeed measured when reactions were carried out at a lower temperature, where turnover was slowed (Figure 4). Even though the inactivation
was not complete at low concentrations of inactivators, the initial phase of inactivation followed pseudo first-order kinetics in both cases. Inactivation kinetics were fitted to the simple model:

\[
E + 5F\alpha XF \xrightarrow{K_i} E\cdot 5F\alpha XF \xrightarrow{k_i} E-5F\alpha X + F^- \]

The apparent first order rate constants, \(k_{obs}\) at each concentration of each inactivator were fitted to the expression \(k_{obs} = k_i[I]/([I] + K_i)\), yielding \(k_i\) and \(K_i\) values for both compounds. For ax-5F\(\alpha\)XF, values of \(k_i = 0.17 \pm 0.01\) s\(^{-1}\) and \(K_i = 2.4 \pm 0.3\) mM were measured, while values of \(k_i = 0.16 \pm 0.02\) s\(^{-1}\) and \(K_i = 0.31 \pm 0.06\) mM were measured for eq-5F\(\alpha\)XF. The kinetic competence of the intermediates formed was assessed by dialyzing excess inactivator from each inactivated enzyme sample at 10 °C, then monitoring the reactivation of each inactivated enzyme by taking aliquots and measuring activity at a series of time intervals at 37 °C. The reactivation for both samples followed pseudo first-order kinetics (Figure 4 C+D) and yielded rate constants of \(k_{react} = 3.2 (\pm 0.4) \times 10^{-3}\) s\(^{-1}\) for ax-5F\(\alpha\)XF and \(3.2 (\pm 0.1) \times 10^{-4}\) s\(^{-1}\) for eq-5F\(\alpha\)XF at 37 °C, proving that inactivation occurred through the normal enzymatic reaction mechanism.

Rates of enzyme-catalysed hydrolysis of each 5F\(\alpha\)XF substrate at 37 °C were measured using a fluoride electrode at a series of substrate concentrations. In each case the enzymatic reaction continued at a constant rate until almost all the substrate had been consumed, indicating very low \(K_m\) values for each compound. These observations are reminiscent of what was seen during the reaction of \(\alpha\)-glycosidases with 5-fluoro-\(\alpha\)-
glycosyl fluorides (12,41-43) and consistent with the low $K_i$' values measured, since the $K_m$ value of each compound as substrate should be equal to its $K_i$' value as inhibitor. The turnover number, $k_{cat}$, was determined from the slope of these plots of fluoride release versus time, yielding $k_{cat} = 2.16 \pm 0.02 \times 10^{-3} \text{ sec}^{-1}$ for ax-5FαXF and $1.22 \pm 0.08 \times 10^{-4} \text{ sec}^{-1}$ for eq-5FαXF. The similarity of these values to the reactivation rate constants confirms that the same process is being monitored in the two cases.

The catalytic nucleophile was identified as Asp 416 within the sequence FKTDFG by LC/MS-MS analysis of proteolytic digests of YicI trapped as its glycosyl enzyme intermediate with eq-5FαXF. Full experimental details are provided in Supplementary Information.

**Overall architecture and oligomerization of YicI**

Crystallization trials yielded two different crystal forms of YicI, each possessing 6 molecules of YicI per asymmetric unit, related by point group 32 non-crystallographic symmetry. The hexamer is formed from two trimeric structures stacked slightly out of register, as is seen from the two views in Figure 5. This observation of a hexameric YicI is in agreement with results from native PAGE and dynamic light scattering analysis (data not shown), and consistent with the results of Okuyama *et al* based on gel filtration chromatography (22). Structure refinement statistics are listed in Table 3. The rms deviation between the form 1 and form 2 hexamers is 0.46 Å² for 18,552 common backbone atoms (0.29 Å² between both form 1 structures), as calculated using the “magic fit” and “refine fit” procedure in swiss PDB viewer (44). Using an identical procedure,
the range of rms deviations between chain A of form 2 and the other chains of the hexamer ranges between 0.2 and 0.28 Å² for backbone atoms. These values are consistent with YicI adopting a highly similar conformation in all the protomers in this study with the only significant exception being due to a partial disorder of some active site loops (induced by crystal packing) in monomer D of crystal form 1. The total buried surface area per monomer upon oligomerization was calculated to be 3913 Å² using CNS (33), a significantly high value in keeping with a physiologically relevant oligomer. Remarkably, although YicI contains 13 Cys residues per monomer, the structure has no disulphide bonds (although a distance of 3.8 Å between the SG atoms of C343 and C412 suggests that one may be possible in structurally related homologues).

The structure of a YicI monomer can be divided into five distinct domains: the N-terminal domain (domain N, residues 1-245), the catalytic domain (245-349 and 387-588), an insert into the catalytic domain (349-387), a proximal C-terminal domain (588-665) and a distal C-terminal domain (665-772). A representation of these individual domains is provided in Figure 6.

The N-terminal Domain

Domain N is predominantly composed of β-strands, folding together to form an extensive, antiparallel β-sandwich. Loops protruding from Domain N form several key intermolecular interactions within the hexamer. Residues 1-13 form a small U-shaped loop that protrudes into the active site of an adjacent monomer, while a loop formed by residues 37-52 contacts a second distinct active site in the hexamer. The largest loop of
domain N, residues 71-90, contacts the catalytic domain of both its own and other monomers, and appears to be important for hexamer formation. The region formed by residues 164-197 of Domain N makes several contacts to the active site of its own monomer. Firstly, a short loop formed by residues 164-172 packs against α6 of the catalytic domain, and the following residues, 176-182, then forms the end strand of the domain N β-sandwich. Lastly, a loop comprising residues 183-197 contacts the catalytic domain and forms part of the edge of the active site. Domain N shows structural similarity with domains from other CAZY entries, with DALI (45) returning significant Z-scores of 5.8 and 5.6, respectively (12 & 8 % sequence identity), for domains within glucoamylase (PDB 1lf6, (46)) and chondroitin AC lyase (PDB 1rw9, (47)). These structurally similar domains have been proposed to play a role in carbohydrate binding (48). An interesting feature of these homologous domains is that they are placed at different points in the protein sequence, preceding the catalytic domain in glucoamylase, and following it in AC lyase. This domain is also found in other members of GH13 and, like YicI, can be seen to associate with the active site of an opposing monomer when present in a multimeric structure (49). Any role of domain N in substrate binding may come from direct effects of loop residues contributing to the various active sites of the hexamer, or more indirect structural effects due to its close interaction with regions of the catalytic domain.

**The Catalytic Domain**

The catalytic domain is composed of a β8α8 barrel, very similar to that of the GH13 family members, with an additional mixed α+β domain inserted between β3 and α3. The
secondary structure of the catalytic domain, and a sequence alignment with selected
GH31 family members, is provided in Figure 1. CAZY clans GH-A, D, H and K have
also been shown to contain $\beta_8\alpha_8$ catalytic domains (1), all of which place the active site
nucleophile on $\beta_4$, and $YicI$ appears to be consistent with this (see below). The $\beta_8\alpha_8$
barrel is typical of this fold, although $\alpha_5$ is replaced by a loop that contacts domain N,
identical to some members of GH13 (50). Helix $\alpha_8$ continues into another $\alpha$-helix, with
the conserved residue P570 forming a $\sim$60° kink between the two helices. In keeping with
the $YicI$ substrate specificity for $\alpha$-glycosidic bonds, the catalytic domain structure
matches best with GH13 $\alpha$-amylase II (PDB 1bvz, (49)), with a DALI score of 17.3 for a
253 amino acid overlap (12% identity).

Both $YicI$ and $\alpha$-amylase II possess an inserted domain between $\beta_3$ and $\alpha_3$ (residues 349-
387 in $YicI$), with the $\beta$-sheet of both inserts overlaying well, but showing structural
differences in the loop regions. This domain is known to vary with substrate specificity
(51). As well as contributing residues to the active site of $YicI$, this insert is also involved
in monomer:monomer packing, and forms part of the central “pore” of the hexamer.
Conserved sequence regions I to VII of GH31 family members (52) are shown mapped to
the $YicI$ structure in Figure 1, one of which, region II, is in the $\beta_3$-$\alpha_3$ insert. Region I and
regions III-VI are located in the main $\beta_8\alpha_8$ fold, and contribute some of the active site
residues. Region VII is located on the $\alpha_8$/continuing helix interface, and is involved in
packing of the catalytic domain against the C-terminal domains.

The C-terminal Domains
Running Title: Structure of a family 31 α-glycosidase

The proximal C-terminal domain possesses a mixed α+β fold, consisting of a central antiparallel β-sheet (that also harbors a long, twisted β-hairpin), and a small loop with two short helical turns. The loop region packs against the start of the catalytic domain, whilst the sheet contacts α-helices 6, 7 and 8 of the β₈α₈ barrel. The contacts between this region and the barrel are particularly hydrophobic, and may indicate a role for the proximal C-terminal domain in protecting the barrel from bulk solvent. The fold of this domain also appears to share structural homology with other CAZY entries, DALI returning Z-scores of 5.4 with trehalose synthase (PDB 1iv8, 60AA overlap, 10% identity, (53)) and 5.3 with α-amylase II (PDB 1bvz, 62 AA overlap, 10% identity, (49)). This domain plays a role in forming the Yicl hexamer, associating with the same domain from the opposing monomer in the dimeric ring. The homologous domain from α-amylase II is not responsible for dimer formation, and is solvent-exposed.

The distal C-terminal domain is located at the points of the triangle-shaped hexamer, and makes its major interaction with the rest of the structure at the β₈α₈ conserved region VII, as noted above. This domain plays no role in hexamer formation, and may be responsible for binding carbohydrates, due to its similarity to carbohydrate binding domains. The fold consists of an antiparallel β-sandwich, comprising 2 sheets of 5 β-strands each, strands 2 and 6 running parallel to each other. The highest DALI scores obtained were 4.8 for an α-L-arabinofuranosidase (PDB 1pz2, 77AA overlap, 6% identity, (54)), 4.8 for a β-xylosidase (PDB 1px8, 69 AA overlap, 10% identity, (55)) and interestingly, 4.7 for the xylan binding domain of a xylanase (PDB 1xbd, 61 AA overlap, 20% identity, (56)). Despite the similar Z-scores, the structural homology with 1pz2 is best, as this overlaps
with all ten β-strands matching, and also shows 2 parallel strands in the same location as those of YicI. Carbohydrate binding domains often bind divalent cations for structural stability, and the location of the Ho binding site in this domain in our heavy atom soaking trials may be indicative of the ancestry of the distal C-terminal domain, although it appears that only water is bound in this pocket in the native electron density maps.

Structure of the Active Site

As is typical of a ββαβαβ structure, the active site is positioned at the C-terminal end of the barrel (Figures 6 and 7). The ends of the barrel β-strands and the loops that follow them contribute the major part of the active site (residues 271-280, 301-316, 342-352, 413-420, 462-468, 477-485, 508-518 and 537-544), with additional regions from the barrel insert (378-382) and three different N-terminal domains (183-195 from the corresponding monomer, and 4-10 and 37-51 from two distinct monomers of the hexamer). The formation of the active site from several of the monomers suggests that oligomerisation of YicI may play a functional role in substrate binding, as well as in stabilization of the protein fold. The active site pocket is easily accessible, largely via a cavity in the middle of the long edge of the hexamer (where the two trimeric rings meet), and also, to a lesser extent, from a cavity extending from the central pore of the hexamer.

Members of GH families with a ββαβαβ fold are known to place the active site nucleophile on β4 and the acid/base on differing β-strands of the barrel (57), e.g. β5 for GH13. Indeed, the proposed nucleophile in YicI is residue D416, located on β4. Although the catalytic domain of YicI matches well with those from GH13, and the reaction catalysed
is similar (α-retaining), the acid/base in \textit{YicI} is residue D482 on β-strand 6, whereas for GH13 members the acid/base is a glutamic acid and is located on β5. This placement of the acid/base on β6 is more similar to that of enzymes from GH18 (chitinases), which catalyse a β-retaining reaction. However, the important functional characteristic is the distance between the two catalytic carboxylic acid residue oxygens – ~6 Å in \textit{YicI} and ~6 Å in GH13 (PDB 1UH2, (48)). The longer distance of at least 7 Å in GH18 (PDB ICTN, (58)) is consistent with the fact that GH18 chitinases follow a different mechanism in which the substrate amide functions as the nucleophile, with the adjacent carboxylate assisting it in that role.

The nucleophile D416 makes two hydrogen bonds, one to the backbone N of F417 (2.8 Å), and another to Ne of W345 (2.8 Å). The acid/base D482 makes hydrogen bonds to two well-coordinated water molecules (all water molecules referenced in this section refer to chain A, form 2 for the description of active site residues and form 1 for sugar binding), one to Wat39 (2.55 Å, also contacted by W8, D185 and Y194) and another to Wat596 (2.9 Å, also contacted by W479 and D511). The conserved residue R466 is placed between the two catalytic residues but makes no direct contacts to either, hydrogen bonding instead to E419 and D185. Residue E419 in particular is well conserved in GH31 proteins, and may serve to fix the orientation of R466. Residues H304, D306 and K414 form a hydrogen bonding network at the bottom of the active site pocket. Typical of a sugar-binding enzyme, there are many hydrophobic residues lining the active site, including W8, F277, C307, W315, W345, W380, F417, W479 and F515. Interestingly, approximately half of these active site hydrophobic residues originate from
elements outside the $\beta_8\alpha_8$ fold, a fact not immediately obvious when looking at sequence alignments containing the seven classical GH31 motifs (52). This may have substantial consequences for specificity within homologues of different oligomeric structure, such as human sucrase-isomaltase, which is predicted to be a dimer (59).

**Binding of eq-5FαXF**

Soaking of the form 1 crystals in 0.5 mM eq-5FαXF for 3 min resulted in clear ring-shaped electron density (Figure 7a), visible at full occupancy in 3 of the active sites (monomers A, B and C) and at partial occupancy in monomer E. Modeling of the sugar analog into this density confirms that the compound has indeed formed a covalent bond between C1 of the sugar and O\(\delta\)2 of D416, with loss of the fluoride from C1. The sugar forms a covalent bond to the enzyme, with a bond distance of 1.45 Å, and a C\(\gamma\):O\(\delta\)2:C1 bond angle of 116° modeled during refinement. Whereas the free eq-5FαXF in solution has been shown to adopt a 4\(C_1\) chair conformation, the covalently bound β-linked sugar is found in a 1\(S_3\) skew boat conformation (Figure 8). This flipped conformation could simply be a response to the strong anomeric effects at C5 and C1 from F5 and the enzyme carboxyl group, respectively – both wanting to adopt axial placements. However, precedent with a similar intermediate trapped in an α-mannosidase would suggest that this conformation is determined by interactions with the enzyme and is mechanistically relevant (60). It is very interesting that this conformation for the β-D-xylopyranosyl enzyme intermediate is identical to the conformation seen for the bound β-glycoside substrate in the uncleaved Michaelis complexes of several β-glycosidases (61-63). Such structures of Michaelis complexes have typically been achieved by soaking or co-
crystallization of a non-cleavable substrate analogue such as a thioglycoside with wild type enzyme, or a natural substrate with an inactive mutant enzyme. In each case the β-glycoside substrate has been found in a 1S3 conformation, consistent with the dictates of stereoelectronic theory (64) or simply minimizing steric repulsive interactions during nucleophilic attack as discussed in (61). The observation of this same conformation for our β-linked glycosyl-enzyme intermediate implies that the same stereoelectronic forces are coming into play as in the cleavage of the β-glycoside substrate. It therefore appears that the same conformational itinerary (4C1 → 1S3 → 4H → 1S3 → 4C1) is followed by α-glycosidases and β-glycosidases in this case, but starting at different points on the path.

The eq-5F-xylosyl adduct makes several important contacts with the protein and conserved water molecules. Sugar atom O2 hydrogen bonds to D482 Oδ1 (2.4 Å) and R466 Nη1 (3.2 Å), O3 to Wat816 (2.5 Å, also coordinated by D482, W479 and D511), K414 Nζ (2.9 Å) and H540 Ne2 (3 Å), and atom O4 to D306 Oδ2 (3.1 Å) and H540 Ne2 (3.2 Å). Atom F5 is located in a hydrophobic environment composed of F277, C307, W315, W345, W380 and F417. Additionally, the peptide bond between W315 and C316 adopts a cis conformation, orienting the sidechain of W315 to closely contact C307 and direct its sidechain towards the sugar-binding site.

The presence of the bulky F277 near C5 of the adduct may provide an explanation for the very low α-glucosidase activity of YicI. This phenylalanine residue is situated at the end of β1, and appears to be present in an insert that is not found in the GH31 family members that possess α-glucosidase activity (Figure 1). This insertion may sterically
crowd this region of the active site, preventing the extra CH$_2$OH group of glucose (c.f. xylose) from binding. Other residues contributing to the overall shape of the sugar-binding site include W345, which may assist in distortion of the substrate, and F515, which is positioned over the hydroxyl groups O3 and O4. An active site tryptophan is commonly observed in glycosidases and has been postulated to serve as a platform for substrate distortion (65).

Of particular interest is the orientation of the catalytic nucleophile in this structure. In the other $\alpha$-glycosidases for which the structure of a glycosyl enzyme intermediate has been determined, the carbonyl oxygen of the nucleophile is located in close proximity to sugar ring O5. This led to the suggestion that O:O interactions may play an important catalytic role, either via ground state destabilization or transition state stabilization. In YicI the nucleophile is twisted substantially away from this region, with the carbonyl oxygen located in close proximity to the sugar H2. While this may not be catalytically relevant for YicI, it does suggest that, in the structurally related $\alpha$-glucan lyases, the C2 proton may well be abstracted by the departing carboxylate, as had been previously suggested (17).

An additional interesting interaction network between the substrate and the enzyme is that involving the sugar 2-hydroxyl, which interacts with E419 via R466. Previous kinetic studies on SPGase and its mutants (13) had shown that mutation of the equivalent conserved Glu (E484 in SPGase and E419 of YicI) eliminated its ability to hydrolyse maltose, but had lesser effects on rates of hydration of D-glucal, a substrate that has no 2-
Running Title: Structure of a family 31 α-glycosidase

hydroxyl. The structure of YicI is therefore completely consistent with this finding, and suggests that E419 orients R466 for optimal interaction and transition state stabilization. Interestingly glucan lyases, which utilize an elimination mechanism rather than hydrolysis for enzymic deglycosylation (17) have Val or Thr at the corresponding position. This implies that the E419-mediated hydrogen bonding network may play a more important role in effecting deglycosylation through hydrolysis than through elimination.

**Binding of Tris and ordered water at the active site: implications for possible multivalent sugar binding**

Following molecular replacement of form 2 using form 1, it was apparent that there was some unexplained electron density at the active site. This was accounted for by modeling in two molecules of Tris per active site (denoted TRS1 and 2; Figure 7b), Tris being a well-known glycosidase inhibitor and a common ligand in many glycosyl hydrolase structures. Indeed, Tris has been shown to inhibit the sucrase-isomaltase of GH31 (66). The N of TRS1 (the molecule deepest in the active site pocket) hydrogen bonds to D416 Oδ2 (2.6 Å), mimicking the position of a δ+ ring O atom of the sugar substrate in subsite -1. All three hydroxyl groups of TRS1 are involved in hydrogen bonds to the protein resembling interactions seen with the sugar substrate: O1 to D482 Oδ1 (2.5 Å) and R466 Nη1 (3.2 Å), O2 to D482 Oδ2 (2.5 Å) and O3 to D306 Oδ2 (2.5 Å) and H540 Nη2 (3.2 Å). Atom O2 of TRS1 also hydrogen bonds to both O1 (3 Å) and O3 (3 Å) of the second Tris molecule, TRS2. Compared to TRS1, TRS2 has fewer direct interactions with the enzyme, O3 hydrogen bonding to R466 Nη1 (3.1 Å) and O2 to D49 Oδ2 (2.8 Å). TRS1
interacts with the same aromatic-rich region as the sugar adduct, whilst TRS2 contacts residues W8, W380 and F417. It is interesting to speculate that the two bound Tris molecules may be representative of the binding of physiological substrate, particularly as a cluster of water molecules (Wat152, 167, 189, 452, 500 and 1120) near TRS2 possess interatomic distances consistent with those between ring hydroxyls of a pyranose sugar. This group of water molecules is located in a cleft formed between the barrel insert domain and barrel β4. This cleft is rich in aromatic residues, including W375, W377, W380, F417 and Y349 which are typically observed in sugar binding sites. In comparison to our sugar adduct, TRS1 O1 and O3 are placed in the same position as sugar ring hydroxyls O2 and O4, respectively. Atom O2 of TRS1 would also appear to be in an ideal position to represent the O of the α1,6-glycosidic bond of a substrate, able to be directly protonated by D482. Modeling of a polysaccharide upon the two Tris molecules and the observed cluster of bound water molecules would suggest a binding site accommodating at least 4 sugar units.

**Importance of the YicI Structure in Relation to the Human GH31 Family Members**

The YicI structure provides a strong foundation for the modeling of the catalytic β8α8 domain of the human GH31 enzymes (lysosomal α-glucosidase and sucrase-isomaltase) (Figure 1). Intriguingly several clinically relevant isolates of the lysosomal α-glucosidase encode for mutations that lie directly within the β8α8 region, all of which result in Pompe’s disease, a devastating glycogen storage disease in humans. The most common mutant, D645N/E, is located in the conserved region V, with its counterpart D511 in YicI involved in sugar binding and the maintenance of active site architecture. Other disease-
inducing mutations that may affect substrate binding are W402R (YicI H304), L405P (C307), M519T (F417), E521K (E419), R600C/H (R466), D645E/H (D511) and R672Q/W (R538). Several mutants are located in the barrel insert – Y455F (S354), G478R (D379) and W481R (W381) and may play more subtle roles in ligand binding. Other mutations may affect the general folding of the $\beta_8\alpha_8$ barrel elements, e.g. S566P (D429), G615R (G481), C647W (G513) and E689K (D553). The structure of YicI thus provides a useful molecular basis for understanding the effects of these mutations in individuals with Pompe’s disease and may help in directing potential therapeutic approaches.

The extra amino acids (in comparison to YicI) in the catalytic domain of the human GH31 members follow the sequence alignment and $\beta_8\alpha_8$ structure well, lying between secondary structure elements, and so probably represent longer loop regions. The region of sequence between $\beta_4$ and $\alpha_4$ is of particular interest, containing conserved cysteines and possible N-glycosylation sites. N-glycosylation may also occur in the loop between the barrel insert and $\alpha_3$ of the sucrase structure. Two defective mutants of human sucrase-isomaltase have mutations that occur in the $\beta_8\alpha_8$ domain, neither of which are close to catalytic residues: L620P, which probably acts to destabilize $\alpha_6$, and S542L, present in the loop between $\beta_4$ and $\alpha_4$. A large number of clinically important sucrase-isomaltase mutants are present outside this domain, and most likely affect protein targeting (67).

It becomes more difficult to predict the expected similarity of the human $\alpha$-glucosidases outside the catalytic domain. Sequence analysis using PFAM (68) indicates the presence
of a β-trefoil domain on the N-terminal side of the barrel in only the human enzymes, a prediction that would seem to preclude comparison with the bacterial enzymes in this region. However, the length of the amino acid region from the start of the predicted β-trefoil region to the start of the catalytic β8α8 barrel (61-323 in sucrase-isomaltase, and 81-340 in α-glucosidase) is roughly consistent with the length of the domain N motif preceding the catalytic domain in YicI (240 amino acids). Indeed, modeling of this region in sucrase-isomaltase based upon the YicI domain N places cysteines 62, 76, 87 and 92/93 of the former at positions in YicI (I3, L16, Q27 and M32 respectively) which would support the formation of 2 disulfide bonds with little perturbation of the observed domain structure. The situation between the YicI C-terminal domains and those from the human enzymes is less clear, especially when considering that in sucrase-isomaltase this region also has to be responsible for leading into another “N-terminal domain”. It is interesting to speculate that this “repeat fold” of sucrase-isomaltase could assemble as a trimer, thereby both fulfilling the hexameric shape of YicI, and presenting a singular face to the membrane (the isomaltase domain containing a putative transmembrane motif).

CONCLUSION

The cloning and kinetic analysis of YicI, in conjunction with the work of Okuyama et al, confirms that YicI is a hexameric α-xylosidase with low α-glucosidase activity, and is most likely responsible for xyloglucan degradation. In common with other GH31 hydrolases it is a retaining enzyme that uses a double displacement mechanism in which D482 acts as an acid/base catalyst to assist the formation of a covalent β-xylosyl-enzyme
that involves D416 as the catalytic nucleophile. The trapping of this intermediate with 5-fluoro-α-xylosyl fluorides and peptide mapping of labeled digests confirms the identity of the nucleophile and provides additional support for a mechanism involving the formation of a distorted (1S3) intermediate via oxocarbenium ion-like transition states.

The three-dimensional structure of YicI clearly shows that the catalytic domain thereof has a classic β8α8 fold, just as is the case for GH13 α-glucosidases. Armed with this structure, useful models of the two key human GH31 enzymes have been constructed thereby providing the first structural insights into the molecular basis for genetic diseases involving these enzymes. The structures presented here may therefore provide useful guidance in the selection of therapeutic approaches to these conditions, as well as in the design of novel inhibitors for interference with sucrose degradation.

ACKNOWLEDGEMENTS and FUNDING

We thank the Department of Energy (U.S.), the Advanced Light Source (Berkeley, CA) and staff of beamline 8.2.1 for access to data collection facilities. We thank the Howard Hughes Medical Institute and the Canadian Institute of Health Research (to N.S.) and the Natural Sciences and Engineering Research Council of Canada and the Protein Engineering Network of Centres of Excellence of Canada (to S.W.) for financial support. Graduate fellowships from the University of British Columbia (to S.L.) and postdoctoral fellowships from the Michael Smith Foundation for Health Research (to A.L.) and the Korean Science and Engineering Foundation (to Y-W.K.) are also gratefully acknowledged. N.S. is a HHMI International Scholar and CIHR Investigator. We also
thank Shouming He for assistance with mass spectrometry analysis, Richard Pfuetzner for light scattering experiments and Dr. David Vocadlo for assistance with inhibitor synthesis.
REFERENCES

1. Henrissat, B. (1991) *Biochem. J.* **280**, 309 - 316
2. Henrissat, B., and Bairoch, A. (1993) *Biochem. J.* **293**
3. Henrissat, B., and Bairoch, A. (1996) *Biochem. J.* **316**, 695 - 696
4. Henrissat, B. (1998) *Bioche. Soc. Trans.* **26**, 153 - 156
5. Campbell, J. A., Davies, G. J., Bulone, V., and Henrissat, B. (1997) *Biochem. J.* **326**
6. Frandsen, T. P., and Svensson, B. (1998) *Plant Mol. Biol.* **37**, 1 - 13
7. Yu, S., Bojen, K., Svensson, B., and Marcussen, J. (1999) *Biochim. Biophys. Acta* **1433**, 1 - 15
8. Quaroni, A., and Semenza, G. (1976) *J. Biol. Chem.* **251**, 3250 - 3253
9. Hermans, M. M. P., Kroos, M. A., van Beeumen, J., Oostra, B. A., and Reuser, A. J. J. (1991) *J. Biol. Chem.* **266**, 13507 - 13512
10. Iwanami, S., Matsui, H., Kimura, A., Ito, H., Mori, H., Honma, M., and Chiba, S. (1995) *Biosci. Biotech. Biochem.* **59**, 459 - 463
11. Kimura, A., Takata, M., Fukushima, Y., Mori, H., Matsui, H., and Chiba. (1997) *Biosci. Biotech. Biochem.* **61**, 1091 - 1098
12. Lee, S. S., He, S., and Withers, S. G. (2001) *Biochem. J.* **359**, 381 - 386
13. Okuyama, M., Okuno, A., Shimizu, N., Mori, H., Kimura, A., and Chiba, S. (2001) *Eur. J. Biochem.* **268**, 2270 - 2280
14. Lee, S. S., Yu, S., and Withers, S. G. (2002) *J. Am. Chem. Soc.* **124**, 4948 - 4949
15. Cogoli, A., and Semenza, G. (1975) *J. Biol. Chem.* **250**, 7802 - 7809
16. Igaki, S., and Chiba, S. (1989) in *Abstracts of the Annual Meeting of the Japan Agricultural Society, Niigata, Japan* Vol. 3Hp16, pp. 292, Agricultural Chemical Society of Japan, Tokyo

17. Lee, S. S., Yu, S., and Withers, S. G. (2003) *Biochemistry* **42**, 13081 - 13090

18. Sinnott, M. L. (1990) *Chem. Rev.* **90**, 1171 - 1202

19. Legler, G. (1990) *Adv. Carbohydr. Chem. Biochem.* **48**, 319 - 385

20. Chambers, J. P., Elbein, A. D., and Williams, J. C. (1982) *Biochem. Biophys. Res. Commun.* **107**, 1490 - 1496

21. Hanozet, G., Percher, H.-P., Vanni, P., Oesch, B., and Semenza, G. (1981) *J. Biol. Chem.* **256**, 3703 - 3711

22. Okuyama, M., Mori, H., Chiba, S., and Kimura, A. (2004) *Protein Expr. Purif.* **37**, 170 - 179

23. Bradford, M. (1976) *Anal. Biochem.* **72**, 248 - 254

24. Laemmli, U. K. (1970) *Nature* **227**, 680 - 685

25. Hayashi, M., Hashimoto, S., and Noyori, R. (1984) *Chem. Lett.*, 1747 - 1750

26. Cudney, R. (1994) *Acta Crystallogr D Biol Crystallogr* **50**, 414-423.

27. Bergfors, T. M. (1999) *Protein Crystallization*, 1 Ed., International University Line, La Jolla, CA

28. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol* **276**, 307-326

29. Collaborative Computational Project, N. (1994) *Acta Crystallogr D Biol Crystallogr* **50**, 760-763.

30. Terwilliger, T. C. (2003) *Methods Enzymol* **374**, 22-37.

31. McRee, D. E. (1999) *J Struct Biol* **125**, 156-165.
32. Vagin, A., and Teplyakov, A. (2000) *Acta Crystallogr D Biol Crystallogr* 56, 1622-1624.

33. Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr D Biol Crystallogr* 54, 905-921.

34. Murshudov, G. N. (1997) *Acta Crystallogr D Biol Crystallogr* 53, 240-255.

35. Schuttelkopf, A. W., and van Aalten, D. M. (2004) *Acta Crystallogr D Biol Crystallogr* 60, 1355-1363. Epub 2004 Jul 1321.

36. Sampedro, J., Sieiro, C., Revilla, G., Gonzalez-Villa, T., and Zarra, I. (2001) *Plant Physiol.* 126, 910 - 920.

37. Monroe, J. D., Garcia-Cazarin, M. L., Poliquin, K. A., and Aviano, S. K. (2003) *Plant Physiol. Biochem.* 41, 877 - 885.

38. Zong, N., and Yasui, T. (1989) *Agric. Biol. Chem.* 53, 187 - 195.

39. Chaillou, S., Lokman, B. C., Leer, R. J., Posthuma, C., Postma, P. W., and Pouwels, P. (1998) *J. Bacteriol.* 180, 2312 - 2320.

40. Zechel, D. L., and Withers, S. G. (2000) *Acc Chem Res* 33, 11-18.

41. McCarter, J. D., and Withers, S. G. (1996) *J. Biol. Chem.* 271, 6889 - 6894.

42. Ly, H. D., Howard, S., Shum, K., He, S., Zhu, A., and Withers, S. G. (2000) *Carbohydr. Res.* 329, 539 - 547.

43. Numao, S., He, S., Evjen, G., Howard, S., Tollersrud, O. K., and Withers, S. G. (2000) *FEBS Lett.* 484, 175 - 178.

44. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis* 18, 2714-2723.
45. Holm, L., and Sander, C. (1997) *Nucleic Acids Res* **25**, 231-234.
46. Aleshin, A. E., Feng, P. H., Honzatko, R. B., and Reilly, P. J. (2003) *J Mol Biol* **327**, 61-73.
47. Lunin, V. V., Li, Y., Linhardt, R. J., Miyazono, H., Kyogashima, M., Kaneko, T., Bell, A. W., and Cygler, M. (2004) *J Mol Biol* **337**, 367-386.
48. Abe, A., Tonozuka, T., Sakano, Y., and Kamitori, S. (2004) *J Mol Biol* **335**, 811-822.
49. Kamitori, S., Kondo, S., Okuyama, K., Yokota, T., Shimura, Y., Tonozuka, T., and Sakano, Y. (1999) *J Mol Biol* **287**, 907-921.
50. Katsuya, Y., Mezaki, Y., Kubota, M., and Matsuura, Y. (1998) *J Mol Biol* **281**, 885-897.
51. MacGregor, E. A., Janecek, S., and Svensson, B. (2001) *Biochim Biophys Acta* **1546**, 1-20.
52. Kashiwabara, S., Azuma, S., Tsuduki, M., and Suzuki, Y. (2000) *Biosci Biotechnol Biochem* **64**, 1379-1393.
53. Kobayashi, M., Kubota, M., and Matsuura, Y. (2003) *J.Appl.Glyosci.* **50**, 1
54. Hovel, K., Shallom, D., Niefind, K., Belakhov, V., Shoham, G., Baasov, T., Shoham, Y., and Schomburg, D. (2003) *Embo J* **22**, 4922-4932.
55. Yang, J. K., Yoon, H. J., Ahn, H. J., Lee, B. I., Pedelacq, J. D., Liong, E. C., Berendzen, J., Laivenieks, M., Vieille, C., Zeikus, G. J., Vocadlo, D. J., Withers, S. G., and Suh, S. W. (2004) *J Mol Biol* **335**, 155-165.
56. Simpson, P. J., Bolam, D. N., Cooper, A., Ciruela, A., Hazlewood, G. P., Gilbert, H. J., and Williamson, M. P. (1999) *Structure Fold Des* **7**, 853-864.
Running Title: Structure of a family 31 α-glycosidase

57. Nagano, N., Porter, C. T., and Thornton, J. M. (2001) *Protein Eng* **14**, 845-855.

58. Perrakis, A., Tews, I., Dauter, Z., Oppenheim, A. B., Chet, I., Wilson, K. S., and Vorgias, C. E. (1994) *Structure* **2**, 1169-1180.

59. Hunziker, W., Spiess, M., Semenza, G., and Lodish, H. F. (1986) *Cell* **46**, 227-234.

60. Numao, S., Kuntz, D. A., Withers, S. G., and Rose, D. R. (2003) *J Biol Chem* **278**, 48074-48083. Epub 42003 Sep 48075.

61. Davies, G. J., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, A. M., Schulein, M., and Withers, S. G. (1998) *Biochemistry* **37**, 11707-11713.

62. Tews, I., Perrakis, A., Oppenheim, A., Dauter, Z., Wilson, K. S., and Vorgias, C. E. (1996) *Nat Struct Biol* **3**, 638-648.

63. Sulzenbacher, G., Driguez, H., Henrissat, B., Schulein, M., and Davies, G. J. (1996) *Biochemistry* **35**, 15280-15287.

64. Deslongchamps, P. (1983) *Stereoelectronic Effects In Organic Chemistry*, Pergamon Press, Oxford

65. Nerinckx, W., Desmet, T., and Claeysens, M. (2003) *FEBS Lett* **538**, 1-7.

66. Kano, T., Usami, Y., Adachi, T., Tatematsu, M., and Hirano, K. (1996) *Biol Pharm Bull* **19**, 341-344.

67. Spodsberg, N., Jacob, R., Alfalah, M., Zimmer, K. P., and Naim, H. Y. (2001) *J Biol Chem* **276**, 23506-23510. Epub 22001 May 23504.

68. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C., and Eddy, S. R. (2004) *Nucleic Acids Res* **32**, D138-141.
69. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., and Thompson, J. D. (2003) *Nucleic Acids Res* 31, 3497-3500.

70. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* 24, 946-950

71. Merritt, E. A. (1994) *Acta Crystallogr D Biol Crystallogr* 50, 869-873.

72. Esnouf, R. M. (1999) *Acta Crystallogr D Biol Crystallogr* 55, 938-940.
FIGURE LEGENDS

Figure 1. Sequence alignment of E.coli YicI and YihQ with Solfolobus solfataricus MalA (α-glucosidase), and the human lysosomal α-glucosidase (Hlys) and sucrase-isomaltase (HsiS, HsiI) catalytic domains. Structural features are represented by rectangles (α-helices) and arrows (β-strands), with elements of the β8α8 barrel core numbered and colored black, and other elements outside this fold colored grey. The seven classical GH31 sequence motifs (52) are labelled I-VII, and the amino acids within displayed in bold text. The catalytic nucleophile is in lower case. Strictly conserved amino acids are labeled by an asterisk, with strongly similar and weakly similar amino acids labeled by a colon and period, respectively. The sequence alignment was prepared using ClustalW (69).

Figure 2. pH-Dependence of the hydrolysis of αXylF by YicI.

Figure 3. Chemical structures and $K_i/K_i'$ values of inhibitors (A) and mechanism-based inactivators (B)

Figure 4. Inactivation of YicI from E. coli with ax-5FαXF and eq-5FαXF at 10°C and reactivation at 37°C. The enzyme was incubated with the following concentrations of ax-5FαXF and eq-5FαXF at 10°C and assayed with 3 mM PNPαXyl at 10°C. A. Inactivation by ax-5FαXF, σ 0.1 mM, 9 0.2 mM, + 0.3 mM, † 0.5 mM, – 0.8 mM, ‡ 2.5 mM (inset, replot of pseudo first order kinetic constants, $k_{obs}$ from A vs concentration of
ax-5FαXF); B. Inactivation by eq-5FαXF, 0.05 mM, 0.1 mM, 0.12 mM, 0.15 mM, 0.25 mM, 0.3 mM, 0.45 mM (inset: replot of $k_{obs}$ values from B vs concentration of eq-5FαXF); C. reactivation at 37 °C of YicI from *E. coli* inactivated (at 10 °C) with ax-5FαXF; D. reactivation of the enzyme inactivated with eq-5FαXF.

Figure 5 (a, b). Oligomeric structure of YicI. The two trimeric rings of the hexamer are colored red and blue, progressing from light shades at the N-terminus of the enzyme to darker shades at the C-terminus. (a) View down the hexamer three-fold axis. (b) View 90° from (a), down the hexamer two-fold axis. Figure prepared using Molscript (70) and Raster3D (71).

Figure 6. Domain structure of a YicI monomer. Individual domains are colored thus: N-terminal domain (domain N), yellow; β8α8 catalytic domain, red; insert into catalytic domain, purple; proximal C-terminal domain, blue; distal C-terminal domain, green. The active site nucleophile, D416, is shown in ball-and-stick form, colored according to atom type (C, yellow; O, red). Figure prepared using Molscript (70) and Raster3D (71).

Figure 7 (a, b). SigmaA-weighted electron density maps of the active site for (a) the form 1 eq-5FαXF sugar adduct, at 2.2 Å resolution, and (b) form 2, with two Tris molecules, at 2.1 Å resolution. Both maps show bias-free fo-fc omit density for the ligand, contoured at $3\sigma$. Selected active site residues are shown in ball-and-stick form, colored according to atom type (C, yellow; O, red; N, blue; S, green). Atoms in the ligand are colored in the same fashion, but with C represented using light grey, and F by magenta. In both (a) and
(b) the active site \( \beta_8\alpha_8 \) fold is shown in blue. Loops contributing to the binding pocket from other monomers in (b) are shown in red and green. Figure prepared using Bobscript (72) and Raster3D (71).

Figure 8 (a, b). (a) Schematic of the sugar adduct interactions with YicI. Hydrogen bonds are represented by dashed lines (and the distances given nearby, in units of Å). (b) Comparison between the YicI sugar adduct structure and the Cel5A Michaelis complex (61), including the respective nucleophile residues D417 and E139. The two structures are represented in ball-and-stick format and colored thus: YicI C atoms, yellow; Cel5A C atoms, blue; O, red; F, magenta. The two sugar molecules can be seen to overlap well, both adopting a \( 1S_3 \) skew boat conformation. Figure (b) prepared using Molscript (70) and Raster3D (71).
### TABLES

| Substrate               | $k_{cat}$ (s$^{-1}$) | $K_m$ (mM) | $k_{cat}/K_m$ (s$^{-1}$M$^{-1}$) |
|-------------------------|----------------------|------------|---------------------------------|
| PNPα-D-glucopyranoside  | 0.046 ± 0.001$^c$    | 7.7 ± 0.4  | 6.0                             |
| PNPα-D-xylopyranoside   | 0.13 ± 0.01$^c$      | 0.71 ± 0.04| 180                             |
| α-D-xylopyranosyl fluoride| 75 ± 3              | 0.97 ± 0.13| $7.7 \times 10^4$               |

$^a$ PNP: $p$-nitrophenyl; $^b$ Not Detected; $^c$ $\Delta \varepsilon$ of 4-nitrophenol at pH 7.0 was measured to be 9.13 (±0.02) mM$^{-1}$ cm$^{-1}$. Rates were calculated using this value.

Table 1. The Hydrolysis of Various Substrates by YicI protein from *E. coli*
### Crystal Parameters

| Dataset        | Ho Edge       | Ho Peak       | Ho remote     | La    | PCMB  |
|----------------|---------------|---------------|---------------|-------|-------|
| Unit Cell (Å)  | 162.01 × 175.10 × 210.71 | 161.97 × 175.30 × 211.69 | 161.51 × 175.02 × 211.56 |

### Soak details

- 20 mM HoCl₃ for 10 min
- 20 mM LaCl₃ for 10 min
- sat. PCMB for 5 min

### Site details

- 6 sites, (pocket of Q694, N666, D665, E632, H697) × 6
- 6 sites, C215 × 3
- C395 × 2, C556

### Data Collection Statistics

|                        | Wavelength (Å) | Resolution (Å) | No. reflections | No. unique reflections | I/σI | Completeness (%) | Rsym | FOM  |
|------------------------|----------------|----------------|-----------------|------------------------|------|------------------|------|------|
|                        | 1.5360         | 3-3.4 (3.52-3.4) | 253137          | 141469                 | 10.7 (4.9) | 89.3 (88.0)       | 7.8 (18.9) | 0.39 (0.22) |
|                        | 1.5352         |                | 252549          | 140746                 | 10.0 (4.3) | 89.0 (88.1)       | 8.4 (21.8) |      |
|                        | 1.5120         |                | 250236          | 139301                 | 10.9 (5.2) | 89.4 (89.3)       | 7.7 (18.1) |      |
|                        | 1.5418         |                | 939630          | 153081                 | 17.7 (7.7) | 96.2 (93.3)       | 10.3 (23.4) |      |
|                        | 1.5418         |                | 607394          | 156472                 | 6.9 (2.7)  | 98.2 (82.0)       | 19.8 (46.9) |      |

**Table 2.** Data collection and phasing statistics for the YicI derivative datasets.

---

*a* Values in parentheses are for the outer shell.

*b* \( R_{sym} = \frac{\Sigma |I_{hkl}| - <I>| \Sigma |I_{hkl}|. \)
Running Title: Structure of a family 31 α-glycosidase

| Dataset | Crystal Form 1 | Form 1 eq-5FαXF soak | Crystal Form 2 |
|---------|----------------|----------------------|----------------|
| Crystal Parameters | | | |
| Spacegroup | P2₁2₁2₁ | P2₁2₁2₁ | P2₁ |
| Unit Cell (Å) | 161.69 × 174.93 × 209.51 | 162.14 × 175.47 × 210.85 | 83.53 × 224.44 × 154.12 |
| β = 101.75° |
| Data Collection Statistics¹ | | | |
| Resolution (Å) | 2.2 (2.32-2.2) | 2.2 (2.32-2.2) | 2.1 (2.21-2.1) |
| Solvent content (%) | 56 | 54 | |
| No. reflections | 992974 | 930994 | 562187 |
| No. unique reflections | 295845 | 297554 | 299802 |
| Redundancy | 3.4 (2.2) | 3.1 (2.6) | 1.9 (1.6) |
| Completeness (%) | 98.9 (92.9) | 97.9 (94.6) | 93.2 (88.1) |
| I/σI | 12.7 (2.3) | 9.4 (2.4) | 10.4 (5.4) |
| R_sym² | 6.7 (35.0) | 9.6 (40.6) | 6.9 (10.4) |
| Refinement Statistics | | | |
| ASU Contents | 6 × 773 AA³, 18 SO₄, 3 Acetate, 4 MOPS, 980 H₂O | 6 × 773 AA³, 13 SO₄, 4 MOPS, 4 eq-5FαXF, 848 H₂O | 6 × 773 AA³, 12 Tris, 1438 H₂O |
| Rwork (%) | 20.0 | 23.6 | 18.0 |
| Rfree⁴ (%) | 23.9 | 27.2 | 22.0 |
| RMS bonds (Å) | 0.019 | 0.007 | 0.017 |
| RMS angles (°) | 1.75 | 1.39 | 1.60 |

¹ Values in parentheses are for the outer shell.
² R_sym = ∑(I_{hk0})−<I>/∑(I_{hk0}).
³ 772 AA from original sequence plus 1 AA from His tag.
⁴ 5% of reflections excluded from refinement.

Table 3. Data collection and refinement statistics for the native YicI datasets.
Figure 1.
Figure 2.
Figure 3.
Figure 4.
Figure 5a.
Figure 5b.
Figure 6.
Figure 7a.
Figure 7b.
Figure 8a.
Figure 8b.
SUPPLEMENTARY MATERIALS AND METHODS

Labeling and Proteolysis for Electrospray Mass Spectrometry

A stock solution of the enzyme (20 µL, 8.8 mg/ml) was incubated with ax- or eq-5FαXylF (20 µL, 10 mM) at 37 ºC for 30 min. The sample was diluted with 0.05 M phosphate buffer (pH 2.0, 90 µl) and incubated with pepsin (15 µL, 1 mg/ml) for 15 min at room temperature. The sample was then rapidly frozen and analyzed immediately upon thawing. A control sample was prepared according to the same procedure, except that no inactivator was added.

Mass Spectrometry

Mass spectra were recorded using an ABI MDS-SCIEX API QSTAR Pulsar i mass spectrometer (Sciex, Thornhill, ON). Peptides were separated on a reverse phase C18 column using an Ultimate Capillary HPLC system (LC Packings, Amsterdam, Netherlands) interfaced with the mass spectrometer. A post-column splitter was used in all experiments, splitting off 85% of the sample into a fraction collector and sending 15% into the mass spectrometer. Spectra were obtained in either the single-quadrupole scan mode (LC/MS) or the tandem MS daughter ion scan mode (MS/MS).

In LC/MS experiments, proteolytic digests of the protein were loaded onto a C18 column (300 µm × 150 mm) and eluted with a gradient of 2 to 40% eluting solvent (0.1% formic...
acid and 85% acetonitrile in water). The mass analyzer was scanned over a mass-to-charge ratio range of 300 – 2400 amu, with a step size of 0.1 amu and a scan time of 1 second. The ion source potential was set at 5 kV; the orifice energy was 50 V. After the LC/MS experiment, total ion chromatograms of the labeled and unlabeled enzyme digests were compared to find the fraction containing the labeled peptide fragments. Samples of the labeled peptide were collected from the post-column flow splitter and lyophilized. The concentrated sample was then sequenced via tandem MS fragmentation analysis.

To determine amino acid sequences of peptides derived from YicI, the mass spectrometer was operated in an IDA (information dependent acquisition) MS/MS mode, where the precursor ion is selected “on the fly” from the previous scan. An m/z ratio (902 in this case) for an ion that had been selected for fragmentation was placed in a list. The peptides including m/z 902 (doubly charged) previously fractionated by HPLC were introduced into the mass spectrometer via a nanospray ion source (Protana, Staermosegaardvej, Denmark). Following mass selection in the first quadrupole (Q1), the peptide of interest was fragmented by collision with nitrogen gas in the second quadrupole (Q2) and the resulting product ions were analyzed in the TOF mass analyzer. The following settings were used: TOF scan range of m/z 100 – 1820 amu, step size of 0.1 amu, and the scan time of 1 second, Q2 potential of –42 V and source voltage of 1000 V.

SUPPLEMENTARY RESULTS AND DISCUSSION

Experimental Identification of the Catalytic Nucleophile of YicI
The accumulation of the covalent glycosyl-enzyme intermediate provided an opportunity to directly identify the amino acid labeled. Fully inactivated enzyme was prepared with both inactivators, separately, along with a control sample containing no inactivator. These were subjected to peptic digestion, followed by LC/ESI MS comparative mapping. A comparison of the masses of all the peptides in the inactivated and control samples revealed that the only significant difference was a peptide fragment corresponding to m/z 601 (triply charged) and 902 (doubly charged) which was detected in the two inactivated samples while no such peptide was detected in the control sample (Supp. Figure I). If these are the labeled peptides of interest, then a peptide of mass ~550 (triply charged) or ~827 (doubly charged) might be expected in the unlabeled sample, this being the mass difference between the peptide of mass 1011 and the 5-fluoro-xylosyl label of mass 51 (triply charged) or 75 (doubly charged). Unfortunately, no such peptide was observed, possibly indicating that the unlabeled peptides are susceptible to further peptic digestion. Differences in proteolytic cleavage as a consequence of the presence of a sugar residue are not rare(1,2). Thus, the labeled fragment was isolated from both inactivated samples by HPLC and sequenced by ESI tandem mass spectrometry (Supp. Figure II). Daughter ion spectra of peptide samples derived from inactivation with axial and equatorial 5-fluoro substituents are identical and only the MS/MS analysis of the peptide labeled by eq-5FoXF is shown. The daughter ion spectrum reveals doubly charged fragments corresponding to m/z 902 and 827 (mass difference of m/z = 75; if singly charged, Δm/z =150) arising from the peptide with and without a label (5-fluoro-β-D-xylosyl moiety). The pattern of the other singly charged fragments readily yielded a sequence of FKTDFGERIPTDQV. Further inspection of the daughter ion spectrum reveals singly
charged fragments of m/z 1528, 1245 and 827, which are consistent with peptides TDFGERIPTDQV, FKTDFGERI and FKTDFG, each bearing the 5-fluoro-β-D-xylosyl moiety. Consensus amino acid residues correspond to Thr 415, Asp 416, Phe 417 and Gly 418 in the sequence (Figure 1 and reference(3)). Only Asp 416 among these four residues is absolutely conserved through all enzymes of GH family 31, and this also corresponds to the invariant residue proposed previously as the catalytic nucleophile in GH family 31 α-glucosidases and α-1,4-glucan lyases(1,4,5). These results strongly suggest that Asp 416 is indeed the catalytic nucleophile of this α-xylosidase.

SUPPLEMENTARY REFERENCES

1. Lee, S. S., He, S., and Withers, S. G. (2001) Biochem. J. 359, 381 - 386
2. Stoll, D., He, S., Withers, S. G., and Warren, R. A. J. (2000) Biochem. J. 351, 833 - 838
3. Chaillou, S., Lokman, B. C., Leer, R. J., Posthuma, C., Postma, P. W., and Pouwels, P. (1998) J. Bacteriol. 180, 2312 - 2320
4. Frandsen, T. P., and Svensson, B. (1998) Plant Mol. Biol. 37, 1 - 13
5. Lee, S. S., Yu, S., and Withers, S. G. (2002) J. Am. Chem. Soc. 124, 4948 - 4949
SUPPLEMENTARY FIGURE LEGENDS

Figure I. Comparative Mapping of Peptic Digests of Control Sample taken at 30.2 minutes of total ion chromatogram (A), Labeled Peptide Samples with ax-5FαXF taken at 30.4 min (B) and eq-5FαXF taken at 30.5 min (C).

Figure II. ESI MS/MS daughter ion spectrum of the peptide fragment m/z 902 (doubly charged) labeled by eq-5FαXF: A. Interpretation; B. MS/MS Spectrum
Supp. Figure I.
A. Sequence

| Sequence       | m/z       |
|----------------|-----------|
| FKTDGERIPTDVQ  | 827 (doubly charged) |
| FKTDGES       | 1094      |
| FKTDGEQ       | 780       |
| FKTDGE        | 492       |
| FKT           | 377       |
| FK            | 276       |
| TEDGERIPTDVQ  | 1377      |
| DFERIPTDVQ    | 1276      |
| FGERIPTDVQ    | 1161      |
| GERIPTDVQ     | 1014      |
| RIPTDVQ       | 828       |
| IPTDVQ        | 672       |
| PTDVQ         | 559       |
| DVQ           | 361       |
| VQ            | 246       |
| FKTDGERIPTDVQ + 5FαXy1 | 902 (doubly charged) |
| FKTDGERIPTDV + 5FαXy1 | 1657      |
| FKTDGESRI + 5FαXy1 | 1245      |
| FKTDGES + 5FαXy1 | 847       |
| TEDGERIPTDVQ + 5FαXy1 | 1528      |

Supp. Figure II.
Mechanistic and structural analysis of a family 31 α-glycosidase and its glycosyl-enzyme intermediate
Andrew L. Lovering, Seung Seo Lee, Young-Wan Kim, Stephen G. Withers and Natalie C. J. Strynadka

J. Biol. Chem. published online October 22, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M410468200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2004/11/09/M410468200.DC1