Crystal structure of a monocotyledon (maize ZMGlu1) β-glucosidase and a model of its complex with *p*-nitrophenyl β-D-thioglucoside

Mirjam CZJZEK*[1], Muzaffer CICEK†, Véronique ZAMBONI*, Wim P. BURMEISTER‡, David R. BEVAN§, Bernard HENRISSAT* and Asim ESEN§[1]

*Architecture et Fonction des Macromolecules Biologiques–AFMB-UMR 6098, CNRS and Universités d'Aix-Marseille I et II, 31 Chemin Joseph Aiguier, F13402 Marseille Cedex 20, France, †Department of Biology, Virginia Polytechnic Institute & State University, Blacksburg, VA 24061, U.S.A., ‡European Synchrotron Radiation Facility, BP 220, F-38043 Grenoble Cedex, France, and §Department of Biochemistry, Virginia Polytechnic Institute & State University, Blacksburg, VA 24061, U.S.A.

The maize β-glucosidase isoenzymes ZMGlu1 and ZMGlu2 hydrolyse the abundant natural substrate DIMBOAGlc (2-*O*-β-D-glucopyranosyl-4-hydroxy-7-methoxy-1,4-benzoxazin-3-one), whose aglycone DIMBOA (2,4-hydroxy-7-methoxy-1,4-benzoxazin-3-one) is the major defence chemical protecting seedlings and young plant parts against herbivores and other pests. The two isoenzymes hydrolyse DIMBOAGlc with similar kinetics but differ from each other and their sorghum homologues with respect to specificity towards other substrates. To gain insights into the mechanism of substrate (i.e. aglycone) specificity between the two maize isoenzymes and their sorghum homologues, ZMGlu1 was produced in *Escherichia coli*, purified, crystallized and its structure solved at 2.5 Å resolution by X-ray crystallography. In addition, the complex of ZMGlu1 with the non-hydrolysable inhibitor *p*-nitrophenyl β-D-thioglucoside was crystallized and, based on the partial electron density, a model for the inhibitor molecule within the active site is proposed. The inhibitor is located in a slot-like active site where its aromatic aglycone is held by stacking interactions with Trp-378. Whereas some of the atoms on the non-reducing end of the glucose moiety can be modelled on the basis of the electron density, most of the inhibitor atoms are highly disordered. This is attributed to the requirement of the enzyme to accommodate two different species, namely the substrate in its ground state and in its distorted conformation, for catalysis.

Key words: family GH-1, glycoside hydrolase, inhibitor, X-ray crystallography.

INTRODUCTION

The carbohydrates are one of the most diverse groups of organic molecules in the biosphere. A reducing hexasaccharide alone is estimated to form more than 10^{12} oligosaccharide stereoisomers [1], although these isomers may not all be encountered in Nature. Carbohydrates occur as simple sugars (e.g. ribose, glucose, galactose, etc.), oligosaccharides (e.g. sucrose, cellobiose, lactose, etc.) and polymers (cellulose, starch, glycogen, etc.), and as conjugates (e.g. many O-, N- and S-glycosides) to other substances, which are generally referred to as aglycones. Such tremendous diversity in stereochemistry, type of assembly, chain length and conjugation to a large variety of non-carbohydrate aglycones is put to use in a multitude of functions that living organisms require for their structure, metabolism, defence and communication. The diversity of carbohydrates in the biosphere is accompanied by a large number of enzymes that are involved in carbohydrate synthesis and degradation. Of these enzymes, glycoside hydrolases catalyse the selective hydrolysis of glycosidic bonds in oligosaccharides, polysaccharides and their conjugates. They occur universally in all living organisms from taxa of the Archaea to humans. A nomenclature system classifying these enzymes into 82 families based on amino acid sequence similarities has been proposed [2,3] and is now widely used. A continuously updated Internet server giving access to these families is available at http://afmb.cnrs-mrs.fr/~pedro/CAZY/db.html.

β-Glycosidases constitute a major group among glycoside hydrolases. They have been the focus of much research recently because of their key roles in a variety of fundamental biological (e.g. growth and development, chemical defence, host–parasite interactions, cellulolysis, lignification, glycosylceramide and vitamin B_6 metabolism, signalling, etc.) and biotechnological (biomass conversion, food detoxification and beverage quality enhancement) processes. For example, plant β-glycosidases have been implicated in: (i) defence against pests [4,5]; (ii) phytohormone activation [6]; (iii) lignification [7]; and (iv) cell-wall degradation in the endosperm during germination [8]. They belong to families 1 and 3 of the glycoside hydrolases and hydrolyse either O-linked β-glycosidic bonds, as is done by β-glucosidases (β-D-glucoside glucohydrolase, EC 3.2.1.21), or S-linked β-glycosidic bonds, as is done by β-thioglucosidases (myrosinase, or β-D-thioglucoside glucohydrolase, EC 3.2.3.1). Thus specific physiological functions and the multiplicity of β-glycosidases in an organism depend on the nature and diversity of the aglycone (aryl or alkyl) or glycone (Glc, Gal, Fuc, Xyl) moiety of their substrates.

In maize (*Zea mays*), β-glucosidase [2-*O*-β-D-glucopyranosyl-4-hydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOAGlc)-hydrolase] occurs in two different forms (isoenzymes, referred to as ZMGlu hereafter), ZMGlu1 and ZMGlu2. The cDNAs corresponding to both isoenzymes have been cloned and sequenced; the deduced protein products of these cDNAs share 90% sequence identity [9]. The catalytically active form of maize β-

Abbreviations used: ZMGlu, isoenzyme of *Zea mays* β-glucosidase; DIMBOAGlc, 2-*O*-β-D-glucopyranosyl-4-hydroxy-7-methoxy-1,4-benzoxazin-3-one; *p*NPTGlc, *p*-nitrophenyl β-D-thioglucoside; TRCBGlu, *Trifolium repens* cyanogenic β-glucosidase; SAMyr, *Sinapis alba* myrosinase; BPBGlu, *Bacillus polymyxa* β-glucosidase; LLBGlu, *Lactococcus lactis* phospho-β-galactosidase; SSBGlu, *Sulfolobus solfataricus* β-glucosidase; TABGlu, *Thermosphaera aggregans* β-glucosidase.

[1] To whom correspondence should be addressed (e-mail czjzek@afmb.cnrs-mrs.fr and aevatan@vt.edu).

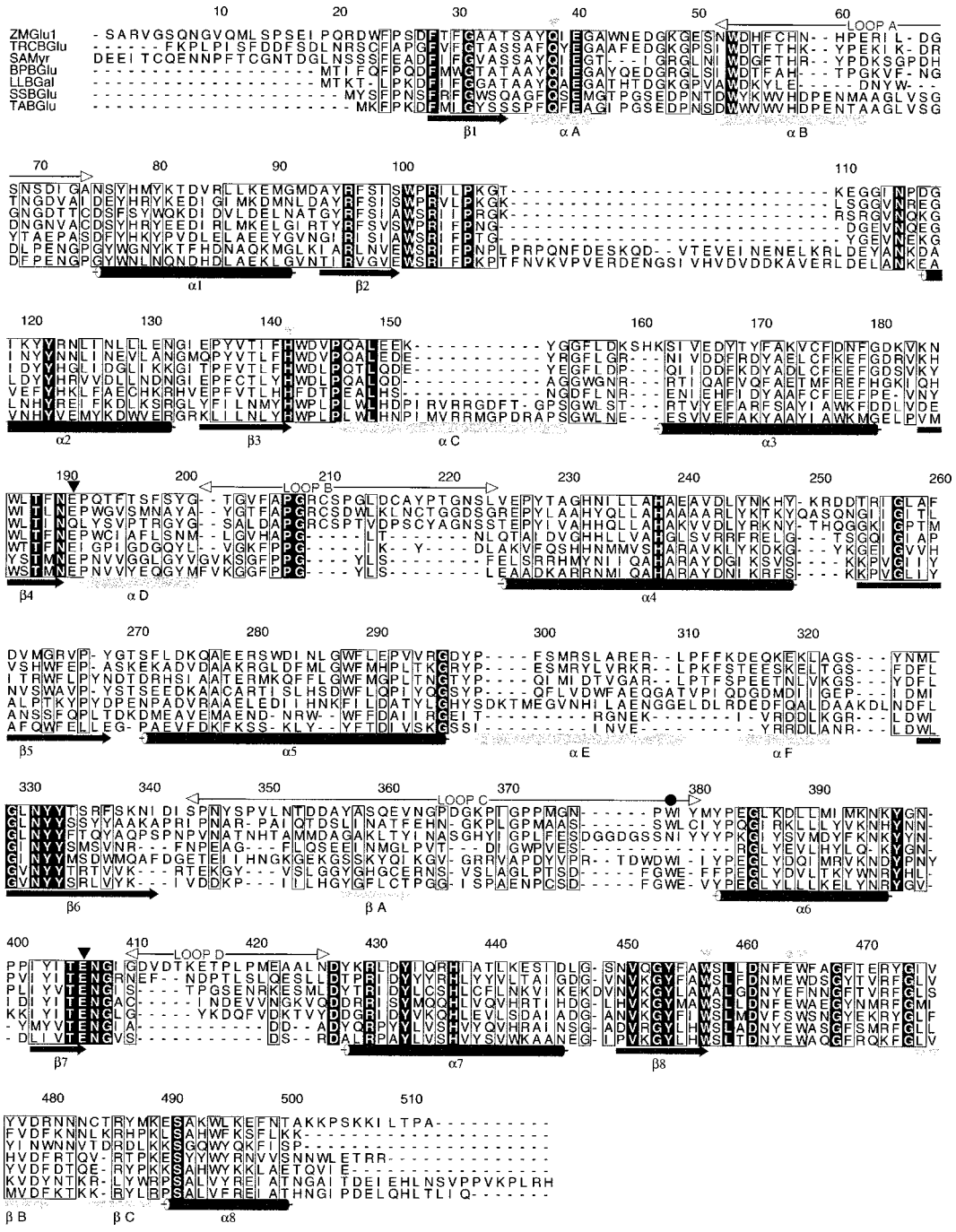

Figure 1 Sequence alignment of the seven β-glycosidases from family 1 with known structures

The residue numbering corresponds to the sequence of ZMGlu1. The secondary-structure elements of the $(\beta/\alpha)_8$ barrel structure are coloured in dark grey. The supplementary secondary-structure elements are coloured in light grey. The four loops A–D, which surround the active site, are delimited above the aligned sequences. The two catalytic glutamates are indicated by black arrowheads. The grey arrowheads show highly conserved residues involved in glucose recognition and the black circle shows Trp-378, which is postulated to be a key residue in the aglycone-binding pocket. ZMGlu1, maize (*Zea mays*) β-glucosidase Glu1; TRCBGlu, white clover (*Trifolium repens*) linamarase; SAMyr, white mustard (*Sinapis alba*) myrosinase; BPBGlu, *Bacillus polymyxa* β-glucosidase; LLBGal, *Lactobacillus lactis* 6-phospho-β-galactosidase; SSBGlu, *Sulfolobus solfataricus* β-glucosidase; TABGlu, *Thermosphaera aggregans* β-glucosidase. Black backgrounds denote the sites of perfect sequence identity among family 1 β-glucosidases in all three domains of living organisms, and the boxes denote the sites of high sequence similarity. The Figure was produced with ALSCRIPT [45].

glucosidase is a 120 kD homodimer [10]. The primary structures of maize β-glucosidases contain the peptide motifs TFNEP and ITENG (residues 188–192 and 404–408, respectively, Figure 1), which are highly conserved and make up the catalytic site in all family 1 β-glycosidases [11–14]. Furthermore, the three-dimensional structures of six family 1 β-glycosidases [Protein Data Bank, PDB, accession code 1cbg, white clover (*Trifolium repens*) linamarase, named TRCBGlu in Figure 1; 1myr, white

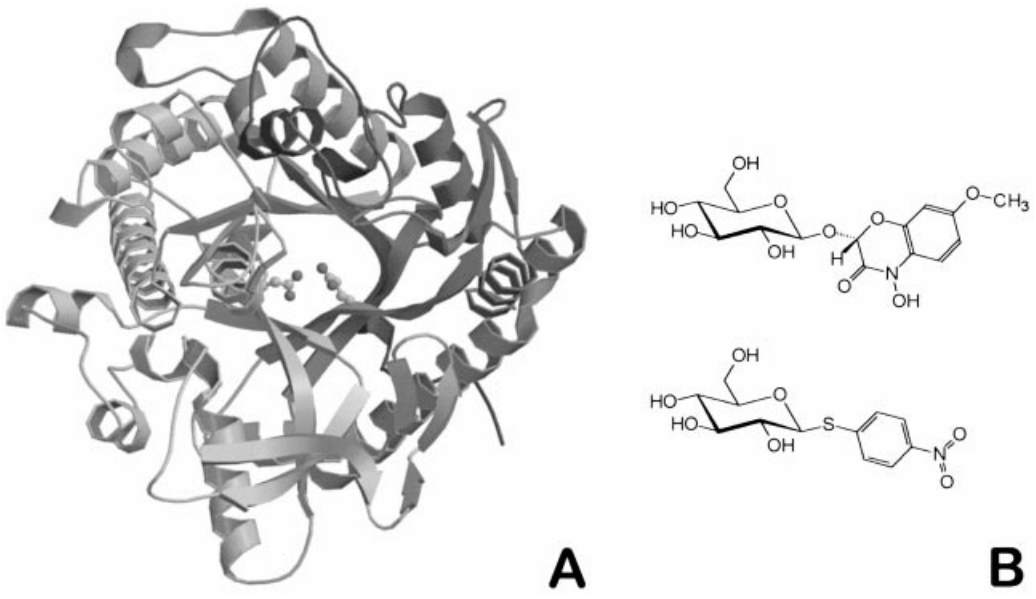

Figure 2 Structure of the overall fold of ZMGlu1 and of ligand molecules

(**A**) Ribbon diagram of the structure of the maize β-glucosidase isoenzyme ZMGlu1 showing the catalytic glutamates Glu-191 and Glu-406. The Figure was produced with MOLSCRIPT [46] and Raster3D [47]. (**B**) Structures of the natural substrate DIMBOAGlc (top) and the competitive inhibitor *p*NPTGlc (bottom).

mustard (*Sinapis alba*) myrosinase, SAMyr; 1bga, *Bacillus polymyxa* β-glucosidase, BPBGlu; 1pbg, *Lactococcus lactis* phospho-β-galactosidase, LLBGal; 1gow, *Sulfolobus solfataricus* β-glucosidase, SSBGlu; and 1qvb, *Thermosphaera aggregans* β-glucosidase, TABGlu] have recently been solved [11–16]. All six enzymes have essentially the same basic $(\beta/\alpha)_8$ barrel fold structure even though they share only 17–44 % sequence identity (Figure 1). All family 1 β-glycosidases are 'retaining' in that the anomeric configuration of the glycone (e.g. glucose) is the same in the product (β-D-glucose) as it was in the substrate (a β-D-glucoside). Substrate hydrolysis involves two steps and requires participation of two glutamic acids (Es). In the glycosylation step, the nucleophilic E in the motif I/VTENG attacks at the anomeric carbon (C-1) of the substrate and forms a covalent glycosyl–enzyme intermediate with concomitant release of the aglycone after protonation of the glycosidic oxygen by the acid catalyst E in the motif T(F/L/M)NEP [17,18]. In the deglycosylation step, the second catalytic E, now an anion and a base catalyst, removes a proton from water, and the resulting -OH group performs a nucleophilic attack on the covalent bond between the glycone and the enzyme, releasing the glycone and regenerating the nucleophilic E. In all structures except myrosinase, a β-S-glucosidase, the residues of the TFNEP and I/VTENG motifs are involved in glycone binding and catalysis within the active site [19]. The two catalytic glutamic acids (i.e. the nucleophile and the acid/base catalyst) are positioned within the active site at expected distances [$\approx$ 5.5 Å (0.55 nm); Figure 2A]. In myrosinases, the motif that corresponds to TFNEP of β-O-glucosidases is TINQL, in which the acid/base catalyst glutamic acid is replaced by a glutamine. There is no need for protonic assistance for aglycone departure in myrosinase, as the aglycone of glucosinolates is already an excellent leaving group. Indeed, the glycosyl-myrosinase intermediate has been trapped by using the mechanism-based inhibitor 2-deoxy-2-fluoro-glucotropaeolin, and the intermediate was shown to have a water molecule hydrogen-bonded to glutamine [13].

There is a fundamental aspect of β-glucosidase-catalysed reactions that is poorly understood: what determines substrate specificity, including the site and mechanism of aglycone binding? Significant progress has been made in understanding the mechanism of catalysis and defining the roles of the two catalytic glutamates within the active site that are involved in catalysis [17,18]. However, there is little or no information as to how β-glycosidases recognize and interact with their substrates, specifically the aglycone moiety, which is the basis of tremendous diversity in natural substrates and is responsible for subtle substrate specificity differences among β-glucosidases. None of the six structures published on family 1 enzymes contains either an intact substrate or an aglycone product to address the issue of substrate specificity.

The maize β-glucosidase isoenzymes ZMGlu1 and ZMGlu2 and their sorghum homologues SBDhr1 and SBDhr2 provide an ideal model system to address questions related to substrate (i.e. aglycone) specificity because they represent extremes in substrate specificity in spite of high sequence identity among them (see above). ZMGlu1 and ZMGlu2 hydrolyse a broad spectrum of artificial and natural substrates in addition to their natural substrate DIMBOAGlc (Figure 2B, top structure). However, ZMGlu2 hydrolyses certain artificial substrates (e.g. nitrophenyl glucosides) about five to six times less efficiently than ZMGlu1 and it does not hydrolyse 6-bromo-2-naphthyl β-D-glucoside, which is readily hydrolysed by ZMGlu1. Similarly, SBDhr1 hydrolyses only its natural substrate dhurrin, whereas SBDhr2 hydrolyses certain artificial substrates in addition to the natural substrate dhurrin [20,21]. Studies with reciprocal ZMGlu1/SBDhr1 chimaeric enzymes indicated that the aglycone (i.e. substrate) specificity-determining sites are different in maize β-glucosidase isoenzyme ZMGlu1 and sorghum β-glucosidase isoenzyme SBDhr1 [22]. These data show that specificity for dhurrin hydrolysis resides in a C-terminal-region octapeptide (^{462}SSGYTERF^{469}) of SBDhr1 where SBDhr1 and ZMGlu1 sequences differ from each other by four amino acid

substitutions, while specificity for DIMBOAGlc hydrolysis is not within the ZMGlu1-analogous sequence of the aforementioned octapeptide nor within the extreme 47-amino-acid-long C-terminal domain of ZMGlu1.

Questions about the mechanism and the site of substrate (i.e. aglycone) specificity and affinity could be directly addressed if it were possible to crystallize the enzyme–aglycone, inactive enzyme–substrate or enzyme–unhydrolysed competitive inhibitor complexes and identify the residues that are interacting with the aglycone in the three-dimensional structure. As a first step to this end, we purified and crystallized the maize β-glucosidase isoenzyme ZMGlu1, which represents the first monocotyledon β-glycosidase to be crystallized. In this paper, we present the three-dimensional structure of ZMGlu1. Furthermore we modelled the complex of ZMGlu1 with the commercially available inhibitor *p*-nitrophenyl β-D-thioglucoside (*p*NPTGlc). We compare the three-dimensional data with those available from other members of family 1 β-glycosidases mentioned above. In addition, we relate the features of the ZMGlu1 tertiary structure to the biochemical and physical properties of the enzyme and discuss their biological implications.

MATERIALS AND METHODS

Expression of maize β-glucosidase ZMGlu1

The region of the Glu1 (GenBank accession no. U25157) cDNA coding for the mature ZMGlu1 was amplified by PCR using oligonucleotide primers (β-glu-111, sense, 5′-ACTACAGCT-AGCGCAAGAGTAGGCAGCCAAAAT-3′; and β-glu-113, antisense, 5′-CTATCTCGAGTTAAGCTGGCGTAAGAAT-CTTC-3′) and cloned into the expression plasmid pET-21a (Novagene). The cloned cDNA was expressed in *Escherichia coli* pLys S cells (F$^-$*ompT hsdSP* $r_B^-m_B^-$ *gal dcm*) under the control of the T7 RNA polymerase promoter as described by Cicek and Esen [23]. The recombinant enzyme so produced was 513 amino acids long and did not have any fusion-protein partner or affinity tags. It was thus identical with the native protein produced by the maize plant (512 amino acids long) except for an add-on alanine residue at the N-terminus.

Purification of ZMGlu1

Recombinant ZMGlu1 enzyme was precipitated from crude cell extracts by 35–65 %-satd. $(NH_4)_2SO_4$. The precipitate was dissolved in 50 mM sodium acetate buffer, pH 5, and centrifuged at 18000 ***g*** for 30 min. The supernatant was adjusted to a final concentration of 0.5 M $(NH_4)_2SO_4$ and centrifuged at 18000 ***g*** for 30 min. Then the supernatant was applied to a ToyoPearl-butyl 650 M hydrophobic interaction chromatography column (TasoHaas, Montgomeryville, PA, U.S.A.; 1.6 × 14 cm). The column was washed to baseline absorbance with 0.5 M $(NH_4)_2SO_4$ in buffer and eluted with five bed volumes of a reverse salt gradient of 0.5–0.1 M $(NH_4)_2SO_4$ in 50 mM sodium acetate buffer, pH 5.0. The fractions collected were assayed for β-glucosidase activity using the artificial substrate *p*-nitrophenyl β-D-glucoside (*p*NPGlc) or the natural substrate DIMBOAGlc. The fractions with activity were pooled based on their purity as judged by SDS/PAGE. The pooled fractions were adjusted to 0.5 M $(NH_4)_2SO_4$ and were re-chromatographed on a ToyoPearl-phenyl 650 M column as described above. Again, the fractions with β-glucosidase activity were identified by SDS/PAGE and pooled. The enzyme was precipitated from the pooled sample by adding $(NH_4)_2SO_4$ to 80 % final concentration. The precipitate was dissolved (11 mg/ml) in 20 mM Hepes, pH 7, and used in the crystallization trials.

Table 1 Unit cell parameters and data-collection statistics

$R_{sym} = \Sigma(I - \langle I \rangle)/\Sigma(I)$.

	ZMGlu1	ZMGlu1–*p*NPTGlc complex
Temperature (K)	100	100
Source	Synchrotron	Rotating anode
Space group	$P2_1$	$P2_1$
Cell a, b, c (Å)	62.88, 118.34, 77.10	59.64, 118.15, 80.88
β (°)	90.3	94.3
Resolution range (Å)	27–2.42	39–2.6
No. of observations	153615	137688
No. of unique reflections	42762	34732
Completeness of data (%):		
all data (outer shell)	99.9 (99.9)	99.8 (97.5)
$I/\sigma(I)$ (outer shell)	6.4 (2.3)	8.8 (1.8)
R_{sym} for all data (outer shell)	0.093 (0.276)	0.094 (0.263)

Crystallization and data collection

All crystallization trials were performed by the vapour diffusion method (hanging drops). Crystals of ZMGlu1 were obtained after mixing 2 μl of enzyme solution ($\approx$ 11 mg/ml, 20 mM Hepes, pH 7) with 2 μl of reservoir [0.1 M sodium citrate, pH 5.6/0.2 M ammonium acetate containing 27 % poly(ethylene glycol) (PEG) 4000]. Crystals of the ZMGlu1–*p*NPTGlc complex were obtained in co-crystallization experiments under the same conditions as those mentioned above except for the inclusion of 1.5 % 1,2,3-heptanetriol as an additive in the crystallization drop and 5 mM *p*NPTGlc (K_i of *p*NPTGlc with ZMGlu1 = 0.15 mM; M. Cicek and A. Esen, unpublished work). Micro-seeding was performed to obtain crystals suitable for the structural analysis. In both cases, two crystal forms ($P2_1$ and $P2_12_12_1$) were observed; however, only the monoclinic form was easily reproducible and led to crystals of sufficient size for X-ray diffraction experiments. Some variations of the unit-cell parameters were observed within one crystal form and are summarized in Table 1. All crystals were cryo-cooled in a nitrogen stream at 100 K prior to data collection. For this, 5 % and 10 % glycerol were added to the mother-liquor and the crystals were bathed progressively, first in the 5 % glycerol solution and then in the 10 % glycerol solution, and then flash-frozen in the nitrogen stream. The native ZMGlu1 data set was collected at the European Synchrotron Radiation Facility (Grenoble, France), beamline ID14 EH3, whereas the co-crystallized ZMGlu1–*p*NPTGlc complex data set was collected on a RIGAKU rotating anode equipped with a 300 mm MARresearch imaging plate detector. All the diffraction data were processed with DENZO [24] and scaled using SCALA [25].

Structure determination and refinement

The structure of ZMGlu1 was determined by the method of molecular replacement using the program AMoRe [26] and the white clover (*T. repens*) cyanogenic β-glucosidase (PDB accession code 1cbg) as the search model. A solution with two orientations was obtained, each orientation corresponding to one of the two molecules in the asymmetric unit related by a non-crystallographic two-fold symmetry. The two molecules in the asymmetric unit corresponded to the biological dimeric unit. The two-body solution gave a correlation coefficient of 35 % and an *R* factor of 45 % in the range of resolution 10.0–3.8 Å. The (F_o–F_c) electron density in the active site of the complex crystals did not reveal the exact positions of all atoms of the inhibitor molecule. A computer model of the inhibitor molecule in two conformations was therefore produced and superimposed with the partial electron

Table 2 Refinement statistics

r.m.s., root mean square.

	ZMGlu1	ZMGlu1–*p*NPTGlc complex
Resolution range (Å)	27.0–2.5	39.0–2.6
Sigma cut-off	0.1	0.1
R_{cryst} factor (%) (no. of reflections)	24.0 (38334)	19.53 (32112)
R_{free} factor (%) (5% of reflections)	28.3 (1900)	25.48 (1335)
No. of non-H protein atoms	7937	7936
No. of hetero-atoms	–	42
No. of solvent atoms	272	359
r.m.s. deviation from ideal geometry		
Bond lengths (Å)	0.009	0.007
Bond angles (°)	1.35	1.33
Dihedrals (°)	23.1	22.6
Impropers (°)	0.9	0.86
Average *B* factors (Å^2) for all atoms	22.6	18.2
Main-chain atoms, chain A/chain B	22.4/22.3	17.7/17.3
Side-chain atoms, chain A/chain B	23.0/23.1	18.3/17.9
Water molecules, molecule A/molecule B	21.2/19.7	19.7/21.0
Ligand	–	40.1
Luzzati co-ordinate error (Å)	0.44	0.3
Ramachandran outliers	2	1
Residues in most favourable regions (%)	81.7	84.8

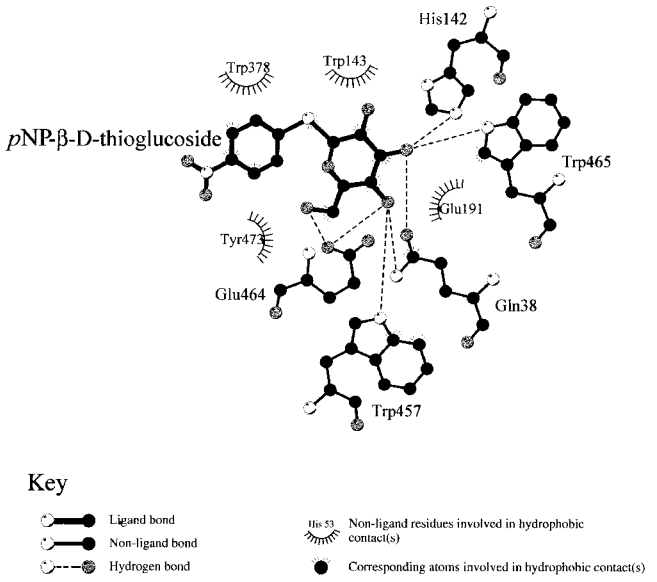

Figure 3 Inhibitor–enzyme interactions

Schematic representation showing all amino acid residues interacting with the *p*NPTGlc molecule in the active site of ZMGlu1. The Figure was produced with LIGPLOT [48].

density in order to account for a maximum number of atoms. Both structures were refined with the program CNS [27], using the maximum-likelihood simulated annealing protocol and restraining the non-crystallographic symmetry. To include the reflections in the total resolution range, a bulk-solvent model as implemented in CNS was applied. The data were finally submitted to a TLS group tensor refinement, as installed in REFMAC version 5.0 [25], in order to take into account anisotropic overall movements of the dimeric molecule. The stereochemistry of the final models was analysed with PROCHECK [28]. Table 2 summarizes the R_{cryst} and R_{free} factors as well as the refinement parameters. Uncomplexed ZMGlu1 crystals were rarely single crystals. This lack of crystal quality is reflected by highly anisotropic TLS group tensors and the overall average *B* factor of 22.64 Å^2, which explains the fact that the model displays high *R* and R_{free} factors at the end of refinement. The co-ordinates and structure factors have been deposited with the PDB (accession codes 1e1e and 1e1f).

RESULTS

Protein fold and overall structure

Both crystal forms $P2_1$ and $P2_12_12_1$ contain the biological dimer in the asymmetric unit. The ZMGlu1 monomer folds as a classical $(\beta/\alpha)_8$ barrel fold (Figure 2A). The connections between β-strands and α-helices within each β/α repeat are elaborated further by the occurrence of short secondary-structural elements (light-grey β-strands and α-helices in Figure 1), which have also been observed in the other family 1 β-glycosidase structures [11–16]. The inserts αE and αF fold against the exterior surface of the β/α barrel and are involved in the formation of the dimer interface, which also contains the loop between β5 and α5 such as the N-terminal part of helix α5 (Figure 1). After residue 286, helix α5 has a 45° kink, which is present in all family 1 enzyme structures and has also been described in the β-glucosidase Ssβ-Gly (SSBGlu) from *S. solfataricus* [15].

In both structures (ZMGlu1 and ZMGlu1–*p*NPTGlc complex), the extreme N-terminal 12 and C-terminal 11 residues were not defined in the electron-density maps, suggesting that these regions are disordered. The N-terminal residues 7–12 are visible only in the native ZMGlu1 B-molecule due to particular crystal packing contacts. The results of the structural refinement are summarized in Table 2.

The maize β-glucosidase tertiary structure, which is the first from monocotyledonous plants and represents a non-glycosylated, plastid-targeted enzyme, shows remarkably high similarity with the six other family 1 enzymes from three different domains (i.e. Eubacteria, Archaea and Eukarya) of living

Table 3 Structural comparison of the glycoside hydrolase family 1 members

The root mean square (r.m.s.) values between two structures were calculated by a least-squares fit after superimposing the structurally equivalent Cα atoms of the two molecules compared. The names are the same as those defined in Figure 1 and the Introduction section.

	Comparison between ZMGlu1 and ...					
	TRCBGlu	SAMyr	BPBGlu	SSBGlu	TABGlu	LLBGal
r.m.s. value (Å)	0.96	1.05	1.15	1.26	1.31	1.36
No. of Cα atoms superimposed	468	462	417	392	391	411
Number of residues	490	501	448	489	482	468
PDB code	1cbg	2myr	1bgg	1gow	1qvb	1pbg

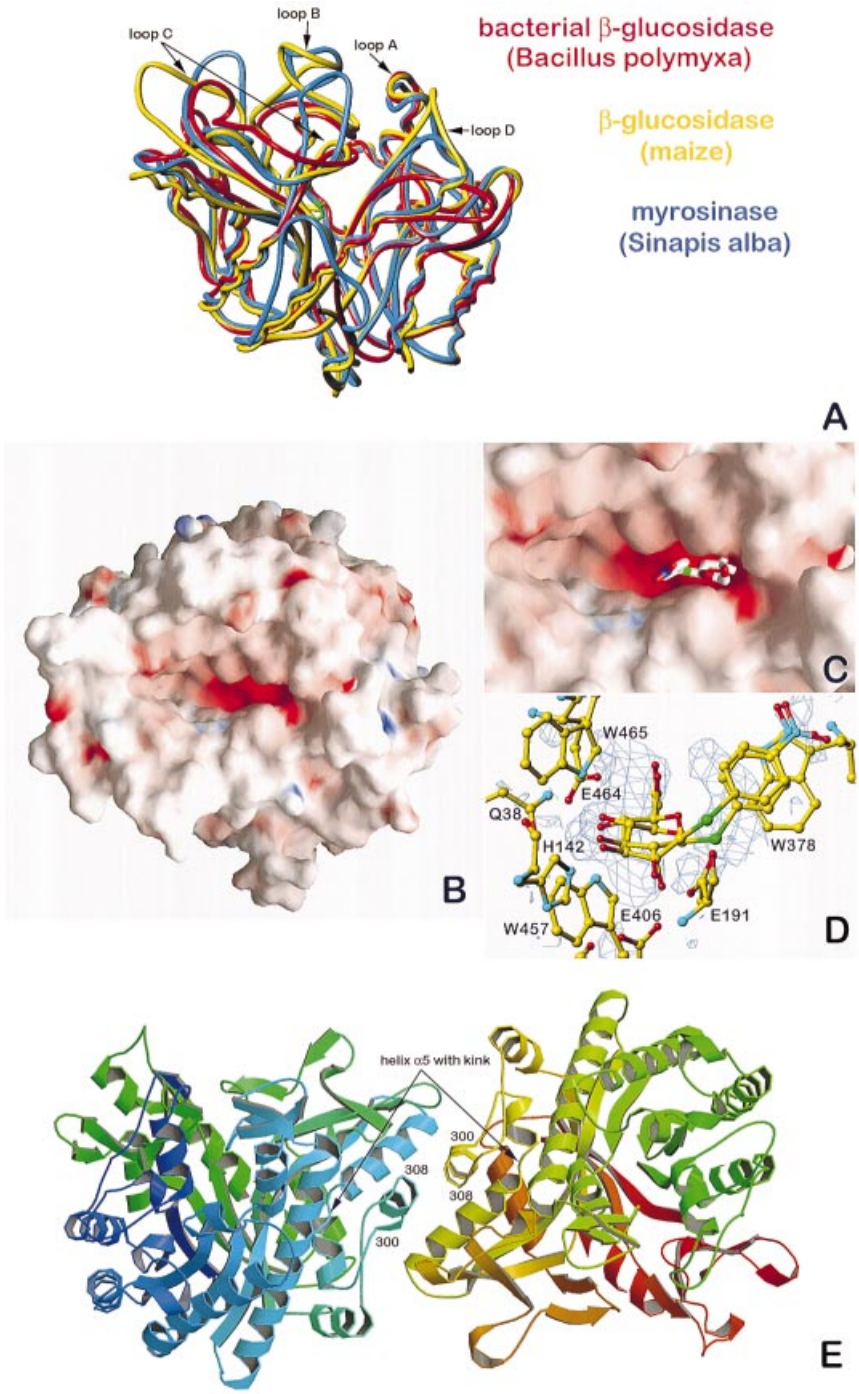

Figure 4 Structural representations of the model of the enzyme–inhibitor complex

(**A**) Superimposition of the Cα traces of ZMGlu1 (yellow), SAMyr (blue; PDB 2myr) and BPBGlu (red; PDB 1bgg). Arrows indicate the four loops (A–D) surrounding the active-site slot. The Figure was produced with TURBO [49]. (**B**) Electrostatic surface representation of ZMGlu1 showing positively charged regions in blue, negatively charged regions in red and neutral regions in white. (**C**) Close-up view of the electrostatic surface showing the slot-like active site with bound *p*NPTGlc, represented as sticks, inside. Panels (**B**) and (**C**) were produced with GRASP [50]. (**D**) Close-up view of the active-site residues with the experimental electron density (F_o-F_c map, contoured at 2σ) for the *p*NPTGlc molecule in the active site. The two superimposed models of the inhibitor, in the $^4C_1$ chair and the $^1S_3$ skew-boat conformations respectively, are equally compatible with the observed electron density. The experimental electron density probably also accounts for other intermediate conformations. The Figure was produced with TURBO [49]. (**E**) The ZMGlu1 oligomeric (dimeric) structure showing the ribbon diagram representation of the respective monomer positions and the positions of the interacting helices that are at the dimer interface. Panel (**E**) was produced with MOLSCRIPT [46] and Raster3D [47].

organisms whose three-dimensional structures have recently been determined. These include enzymes from plants [12,13], eubacteria [11,14,29] and hyperthermophilic archaea [15,16]. ZMGlu1 shares sequence identities ranging from about 17% (for the archaea *T. aggregans* and *S. solfataricus*) to 44% (for *T. repens*) with other family 1 glycoside hydrolases. The extent of

similarity among different family 1 enzyme structures is also evident from the root mean square deviations measured after superimposition (Table 3).

Active-site and substrate binding

The structure of ZMGlu1 has also been determined after co-crystallizing the enzyme with *p*NPTGlc. The difference electron density with the uncomplexed structure, unfortunately, does not reveal the exact positions of all atoms of the inhibitor molecule. The structures of unhydrolysed enzyme–substrate and enzyme–inhibitor complexes of retaining cellulases and of a chitobiase have demonstrated recently that the sugar ring at the site of cleavage adopts the energetically less favourable skew-boat ($^1S_3$) conformation prior to the cleavage of the glycosidic bond [30,31]. We therefore produced computer models of both $^4C_1$ chair and $^1S_3$ skew-boat conformations and have superimposed them on to the partial electron-density map (see Figure 4D). It appears that the atoms at the non-reducing end, such as O-6, C-6, C-5, C-4, O-4, O-3 and O-2 of the glucose moiety, superimpose well with the partial electron density. However, the entire aglycone moiety looks disordered. It is unlikely that the *p*NPTGlc has been hydrolysed with subsequent loss of the aglycone moiety from the active site, since no detectable hydrolysis has been observed in activity assays with this inhibitor. Furthermore, the crystals or crystallization solutions did not turn yellow, indicative of free *p*-nitro-thiophenolate. Thus it is likely that the aglycone moiety is in the active site in the crystals, though the observed disorder of the electron density may also result from partial occupation and/or superimposition with glycerol or water molecules bound in the active site.

The amino acids that are surrounding the inhibitor molecule in the active site cleft of Glu1 are indicated in Figure 3. Four extended and solvent-exposed loops (loops A–D in Figures 1 and 4A) form a cleft-like gate to the active site of ZMGlu1 (Figure 4B) and each of the above-mentioned family 1 enzymes.

Oligomeric structure

The crystallographic data showed that ZMGlu1 is a dimer (Figure 4E) composed of monomers whose calculated molecular mass is 58.4 kDa and that estimated by SDS/PAGE is 60 kDa. This is in agreement both with molecular-mass estimations of about 120 kDa on the native enzyme by gel filtration and electrophoresis and with heterodimers of intermediate electrophoretic mobility observed in zymograms of the F_1 hybrids from maize inbred lines whose ZMGlu1 allozymes differ in electrophoretic mobility [32]. In the dimer (Figure 4E), each monomer is of globular shape and has overall dimensions of about 52 Å (long) × 51 Å (wide) × 47 Å (high), where the height is the distance from the top to the bottom of the barrel. Likewise, the dimer is 103 Å (long) × 51 Å (wide) × 47 Å (high). The two molecules in the dimer are related by a two-fold non-crystallographic symmetry axis. The positioning of the monomers in the dimer is side by side, but slightly tilted and asymmetrical such that the active-site entrance of each monomer is on opposite sides with respect to the longitudinal axis of the dimer. The surface area that is accessible to solvent (1.4 Å probe radius) in monomer A of each dimer is about 17681.7 Å^2 whereas that which is accessible to solvent in monomer B is 17903.8 Å^2. The surface area buried on the dimer interface is 2005.2 Å^2 ($\approx$ 1000 Å^2/monomer), which corresponds to 6 % of the monomer's solvent-exposed surface, a value typical for buried surfaces in dimers [33].

The monomer surface regions that are interacting in the dimer interface and the nature of the interactions holding the monomers together are as follows. First, a short loop formed by Thr-270 and Ser-271 between β-strand 5 and α-helix 5 in one monomer interacts with the connecting loop Lys-396–Pro-400 between α-helix 6 and β-strand 7 of the other monomer. Second, a large loop between α-helix 5 and strand β6, which spans residues Trp-288 to Tyr-326, contains a short α-helix formed by residues that span Phe-300–Glu-308 (two-and-a-half turns). One side of this short α-helix is completely buried in the dimer interface where it interacts with the corresponding identical helix in the other monomer. The ends of the short α-helices in both monomers are surrounded by the β-loops formed by residues spanning Asp-342 to Tyr-357. The interaction between these short α-helices makes the carbonyl group of Ser-304 in each monomer the invariant contact point in the two-fold symmetry relating the two monomers. Finally, there are two identical salt bridges at the dimer interface, one between Arg-295 NH1 of monomer A and Asp-342 OD1 of monomer B and the other (symmetrical) between Asp-342 of monomer A and Arg-295 of monomer B. There are no intermolecular disulphide bonds holding the monomers in the dimer. Thus 80 % of the monomer contacts (closer than 4.0 Å) in the dimer interface involve hydrophobic interactions while 20 % involve ionic and hydrogen bonding.

DISCUSSION

The overall structure of ZMGlu1 compared with the other members of family 1

The structural superimposition data show that the main differences among family 1 β-glycosidases are at the level of the surface loops surrounding the active-site cleft (Figure 4A), and this is also evident at the primary-structure level when sequences are aligned (Figure 1). Furthermore, both the alignment and the structural superimposition data indicate that the tertiary structures of family 1 β-glucosidases are essentially similar, although their primary-structure similarities may be as low as 17 % (e.g. maize ZMGlu1 compared with the β-glycosidases from the archaea *Sulfolobus* and *Thermosphaera*). The alignment data also indicate that only 47 amino acids out of about 500 are strictly conserved in the above-mentioned β-glucosidases, which originate from three different domains of living organisms ranging from archaea to plants. The acid/base catalyst glutamate (Glu-191), located in the motif TFNEP, is at the C-terminal end of strand β4 and the nucleophilic glutamate (Glu-406), located in the motif I/VTENG, is at the C-terminal end of strand β7. This specific location of the catalytic machinery is strictly conserved, not only in family 1 glycoside hydrolases, but also in a large superfamily named clan GH-A, which encompasses several families of retaining β-glycosidases [3,34,35].

It should be emphasized that all seven family 1 glycoside hydrolases with known three-dimensional structures, which now includes ZMGlu1, occur as either dimers (ZMGlu1, TRCBGlu, SAMyr and LLBGal) or tetramers (SSBGlu and TABGlu), while BPBGlu occurs as a tetramer of dimers. This suggests that the catalytically functional unit of these enzymes has a quaternary structure. However, it is not known whether oligomerization is a requirement for activity or stability, or both. What is known for ZMGlu1 is that once the dimeric structure is disrupted by SDS and 2-mercaptoethanol with or without heating, the resulting monomers are not catalytically active based on results from zymograms developed on SDS gels [10]. The present results do not provide any clues as to whether activity loss is due to denaturation of the monomeric structure by the very treatments disrupting the dimeric structure or due to disruption of the dimeric structure alone. Future studies changing the amino acids that are involved in oligomerization (e.g. those at the dimer

interface) by site-directed mutagenesis may provide insights and answers to this question.

The three-dimensional structures reveal that the regions involved in the formation of the oligomers vary widely from one enzyme to the other. A comparison of specific amino acids and polypeptide domains contributing to the dimer interface, thus responsible for holding the dimeric structure together, among the three plant β-glycosidases ZMGlu1, TRCBGlu and SAMyr indicates no similarity or obvious pattern. For example, essentially all of the residues (Glu-412, Leu-415, Thr-417, Asp-421, Tyr-424, Arg-425, Tyr-428/429, Leu-479) that are reported to be at the dimer interface in TRCBGlu are located at the C-terminal region [12], whereas all of the dimeric interface residues (Thr-43, Gly-45, Leu-48, His-56, Arg-57, Arg-106, Asp-150, Glu-151) in SAMyr are located at the N-terminal region [13]. In contrast, the ZMGlu1 dimer interface is formed by residues and peptide spans (see above) that are located in the mid-section (Thr-270–Pro-400) of the polypeptide chain. None of the above-mentioned interface residues is conserved universally among family 1 glycoside hydrolases. These data suggest that dimerization and other higher levels of quaternary associations in at least plant β-glycosidases are recently derived characteristics rather than ancient ones. Apparently, dimerization and other forms of stable inter-monomer associations are important for stability and/or activity in β-glycosidases. However, the specific monomer surface sites and domains that interact to form a quaternary structure are not critical as long as they do form a stable quaternary structure and do not block the active site.

The stability of the ZMGlu1 monomer

The $(\beta/\alpha)_8$ fold is known to be a rather stable structural fold. With respect to the factors that govern thermal stability or resistance to denaturing agents [36], ZMGlu1 has a remarkably high number of prolines, 32, or 6.2 % of the amino acid content, a total of 24 intramolecular ion pairs and 515 hydrogen and electrostatic bonds in a polypeptide backbone of 512 residues. These values have recently been compared for the different family 1 β-glycosidases by Chi et al. [16]. The number of ion pairs in ZMGlu1 compares well with those observed for mesophilic members; however, the number of prolines is clearly in the range observed for thermophilic members. Similarly, the ratio of hydrogen bonds per residue (calculated with HBPLUS [37]) is rather high: 1.006 when compared with other β-glycosidases (BPBGlu, 0.91; TABGlu, 0.944; TRCBGlu, 0.959; LLBGal, 1.0). Only SSBGlu (1.061) and SAMyr (1.028) have even higher ratios of hydrogen bonds per residue. In the cases of SAMyr and ZMGlu1, the increased stability might play a role in view of the function of these plant β-glucosidases, which are involved in pest defence. The typical *in vivo* reaction medium for ZMGlu1 is a rather 'hostile' one, rich in proteases and partly dehydrated, since it includes lysates of disrupted and injured host cells, secretions of insect mouth parts and gut as well as those of phytopathogens. For the enzyme to perform its physiological functions [e.g. production of toxic chemicals such as DIMBOA (2,4-hydroxy-7-methoxy-1,4-benzoxazin-3-one) for defence against pests and release of active phytohormones from their inactive glucoconjugates for wound healing at the site of injury] in such a hostile environment, it must have a compact and stable tertiary structure.

The only disulphide bond found in the crystal structure of ZMGlu1 is intramolecular, involving Cys-210 and Cys-216. In fact, this disulphide bridge stabilizes the loop that shields a cluster of hydrophobic residues in the active site from the solvent. The importance of disulphide bond(s) for enzyme activity was established earlier [38]; it was shown that ZMGlu1 was inactivated by 2-mercaptoethanol in a concentration- and time-dependent manner. Recently, Rotrekl et al. [39] have shown that the replacement of either Cys-210 or Cys-216 in an allozyme of ZMGlu1 with Ala, Ser, Arg or Asp by site-directed mutagenesis resulted in misfolded monomers directed to inclusion bodies in the *E. coli* expression system used. This suggests that the intramonomeric disulphide bond formed between Cys-210 and Cys-216 is essential for the correct folding of the monomer for dimerization. An alignment of 61 plant β-glycosidases available in sequence databases (results not shown) shows that the sites corresponding to Cys-210 and Cys-216 are invariant in 58 out of 61 and 50 out of 61 enzymes, respectively, suggesting that a disulphide bond between these two cysteines is indeed critical to enzyme structure and function.

The active site of ZMGlu1

The active site of the maize β-glucosidase isoenzyme ZMGlu1 appears as a slot-like flat pocket (Figure 4B). This topology results from a 45° kink in helix α5. Helix α5 of the $(\beta/\alpha)_8$ barrel structure of clan GH-A enzymes [34,35] very often appears to play an essential role in defining the shape of the active site. For example, in family 5 endocellulases [40–42], where the active site must recognize a polysaccharide chain, a very short helix α5 or the total absence of helix α5 is the main structural determinant generating a long-groove active-site topology. Another example is found in family 2 β-galactosidase from *E. coli*, where the catalytic domain also belongs to clan GH-A. Here, a complete lack of helix α5 provides space for a loop coming from another domain, which contributes to the shape of the active site [43].

There is no detectable conformational change between the enzyme present in the crystals of the native ZMGlu1 and the crystals obtained in presence of the inhibitor *p*NPTGlc. The active site is about 18 Å deep and 23 Å wide. The distance between the two walls forming the slot-like pocket ranges from 11 Å at its widest point to about 7 Å at its narrowest point (Figure 4B). The model of the *p*NPTGlc inhibitor molecule is entirely buried in the active site where it occupies the narrowest half of the slot (Figure 4C). The glucose ring is located at the very bottom of the binding pocket and the *p*-nitrophenyl moiety is found above, with its ring parallel to the walls. The distance between the walls in the aglycone-binding region is very short ($\approx$7 Å), suggesting that a substrate (or inhibitor) molecule cannot directly enter the active site through this constricted point without substantial dilation of the active-site slot via a conformational change. In contrast, the unoccupied region of the slot is much wider ($\approx$ 11 Å) and appears better adapted for the entrance of the substrate (or inhibitor) into the active site. At the bottom of the slot-like active site, a cluster of conserved residues (highlighted by grey arrowheads in Figure 1) forms hydrogen bonds with the hydroxy groups O-3, O-4 and O-6 of the glucose ring of the inhibitor (Figure 3). The position of the glucose in our model of the glucose-binding pocket is comparable with that observed in other family 1 enzymes [13,14,29] and the hydrogen bonds are formed in a highly similar way. The amino acid residues involved in these hydrogen bonds are Gln-38, His-142, Trp-457, Glu-464 and Trp-465, which are responsible for the recognition of the glycone moiety. The first three residues are strictly conserved in all family 1 β-glycosidases with known primary structures, while the last two residues are highly conserved. Glu-464 is replaced by a serine in LLBGal, interacting with the negatively charged 6-phosphate group of the 6-phospho-galactoside [11,29]. Trp-465 is replaced by a phenylalanine in SAMyr. No hydrogen bonds are encountered on the aglycone

side of the inhibitor. However, the aromatic aglycone of the inhibitor is stacked against Trp-378 (black circle in Figure 1), which presumably is a crucial residue for aglycone binding and recognition in ZMGlu1. This is consistent with the fact that all known natural and artificial substrates that are hydrolysed by ZMGlu1 have an aryl group as their aglycone moiety. It is worth noting that Trp-378 is present in other family 1 glycosidases, but structural superimposition shows that it has a different conformation in ZMGlu1, lining the active site in a way that makes stacking interactions with an aromatic aglycone possible. Trp-378 is at the end of loop C, which shows high variation between the different family 1 enzymes, and is preceded by a proline residue (Pro-377) that is present in ZMGlu1 only. Pro-377 could be responsible for the particular conformation of Trp-378.

The *p*NPTGlc inhibitor molecule is highly disordered in the active site of ZMGlu1. Hydrogen bonds are formed by three atoms only, namely O-6, O-4 and O-3, while π-stacking interactions probably maintain the aromatic aglycone moiety. Therefore, the inhibitor molecule does not seem to be tightly constrained by interactions with the protein, which are only present at the extreme end of the sugar ring. The fact that we only observe weak density, which does not result from residual cleavage, leads us to postulate that the active site and the residues interacting with the glucose and the aglycone moieties are arranged in a manner that stabilizes both conformers of the inhibitor molecule. Such a flexibility of the glucose ring is consistent with recent findings on the catalytic mechanism and reaction intermediates of retaining β-glycosidases from other families. The structures of unhydrolysed enzyme–substrate or enzyme–inhibitor complexes of retaining cellulases and of a chitobiase have shown that the sugar ring at the site of cleavage adopts an energetically less favourable conformation ($^1S_3$ skew-boat) prior to the cleavage of the glycosidic bond [30,31]. This conformation, which yields a quasi-axial orientation for the glycosidic bond, is necessary to expose the sugar C-1 for the 'in line' attack by the enzymic nucleophile glutamic acid [30,44]. The disorder at the distal part of the aglycone moiety may be a simple mechanical consequence of the flexibility of the glucose ring. Alternatively, it is also possible that the observed disorder results from the fact that the *p*-nitrophenyl moiety is synthetic and not the natural aglycone recognized by this enzyme; the differences in length and bonding angle of the S-glycosidic bond compared with a natural O-glycosidic bond may also be important. Nevertheless, we believe that the disorder seen in the inhibitor is primarily due to the fact that the active site must be able to recognize at least two species, e.g. the low-energy $^4C_1$ conformation taken by the substrate in solution and the higher-energy $^1S_3$ conformation required prior to catalysis.

The present structural data allow us to define the active site of maize β-glucosidase, and to propose a model of the ZMGlu1–*p*NPTGlc complex. The structural analysis of ZMGlu1 provides a working hypothesis as to how substrate (i.e. aglycone) specificity is tailored within the active site of a family 1 β-glycosidase. The possible interaction of the inhibitor with Trp-378 shows that this residue is an important component of the aglycone-binding pocket. Site-directed mutagenesis should provide more information about the precise role of Trp-378. However, the aglycone (substrate) specificity observed for this type of enzyme is obviously not determined by a single residue. Therefore the next step will be to prepare and solve the structure of a catalytically inactive mutant of ZMGlu1 for co-crystallization with the natural substrate DIMBOAGlc. This would provide a detailed understanding of the mechanism of substrate recognition and binding as well as the basis of the observed substrate specificity differences among family 1 plant β-glucosidases.

This research is supported by the National Science Foundation (NSF) grant MCB-9906698 and the NSF (INT-991088) and CNRS (INT-9147) Cooperative Research Program. We also gratefully acknowledge the European Synchrotron Radiation Facility for generously according us beam time on ID14 EH3.

REFERENCES

1 Laine, R. A. (1994) A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05×10^{12} structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. Glycobiology **4**, 759–767
2 Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. **280**, 309–316
3 Henrissat, B. and Davies, G. (1997) Structural and sequence-based classification of glycoside hydrolases. Curr. Opin. Struct. Biol. **7**, 637–644
4 Niemeyer, H. M. (1988) Hydroxamic acids (4-hydroxy-1,4-benzoxazin-3ones), defense chemicals in the gramineae. Phytochemistry **27**, 3349–3358
5 Poulton, J. E. (1990) Cyanogenesis in plants. Plant Physiol. **94**, 401–405
6 Smith, A. R. and van Staden, J. (1978) Changes in endogenous cytokin levels in kernels of *Zea mays* L. during inhibition and germination. J. Exp. Bot. **29**, 1067–1073
7 Dharmawardhana, D. P., Ellis, B. E. and Carlson, J. E. (1995) A β-glucosidase from lodgepole pine xylem specific for the lignin precursor coniferin. Plant Physiol. **107**, 331–339
8 Leah, R., Kigel, J., Svendsen, I. and Mundy, J. (1995) Biochemical and molecular characterization of a barley seed β-glucosidase. J. Biol. Chem. **270**, 15789–15797
9 Bandaranayake, H. and Esen, A. (1996) Nucleotide sequence of a cDNA corresponding to a second β-glucosidase gene in maize (*Zea mays* L.). Plant Physiol. **110**, 1048
10 Esen, A. and Gungor, G. (1993) Stability and activity of plant and fungal β-glucosidases under denaturing conditions. In β-Glucosidases: Biochemistry and Molecular Biology, ACS Symposium Series vol. 533 (Esen, A, ed.), pp. 214–239, American Chemical Society, Washington DC
11 Wiesmann, C., Beste, G., Hengstenberg, W. and Schulz, G. E. (1995) The three-dimensional structure of 6-phospho-β-galactosidase from *Lactococcus lactis*. Structure **3**, 961–968
12 Barrett, T., Suresh, C. G., Tolley, S. P., Dodson, E. J. and Hughes, M. A. (1995) The crystal structure of a cyanogenic β-glucosidase from white clover, a family 1 glycosyl hydrolase. Structure **3**, 951–960
13 Burmeister, W. P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S. and Henrissat, B. (1997) The crystal structures of *Sinapis alba* myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase. Structure **5**, 663–675
14 Sanz-Aparicio, J., Hermoso, J. A., Martinez-Ripoll, M., Lequerica, J. L. and Polaina, J. (1998) Crystal structure of β-glucosidase A from *Bacillus polymyxa*: insights into the catalytic activity in family 1 glycosyl hydrolases. J. Mol. Biol. **275**, 491–502
15 Aguilar, C. F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M. and Pearl, L. H. (1997) Crystal structure of the β-glycosidase from the hyperthermophilic archeon *Sulfolobus solfataricus*: resilience as a key factor in thermostability. J. Mol. Biol. **271**, 789–802
16 Chi, Y.-I., Martinez-Cruz, L. A., Jancarik, J., Swanson, R. V., Robertson, D. E. and Kim, S.-H. (1999) Crystal structure of the β-glucosidase from hyperthermophile *Thermosphaera aggregans*: insights into its activity and thermostability. FEBS Lett. **445**, 375–383
17 Withers, G. S., Warren, R. A. J., Street, I. P., Rupitz, K., Kempton, J. B. and Aebersold, R. (1990) Unequivocal demonstration of the involvement of a glutamate residue as a nucleophile in the mechanism of a "retaining" glycosidase. J. Am. Chem. Soc. **112**, 5887–5889
18 Wang, Q., Trimbur, D., Graham, R., Warren, R. A. and Withers, S. G. (1995) Identification of the acid/base catalyst in *Agrobacterium faecalis* β-glucosidase by kinetic analysis of mutants. Biochemistry **34**, 14554–14562
19 Davies, G. and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. Structure **3**, 853–859
20 Hösel, W., Tober, I., Eklund, S. H. and Conn, E. E. (1987) Characterisation of β-glucosidase with high specificity for the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L) Moench seedings. Arch. Biochem. Biophys. **252**, 152–162
21 Cicek, M. and Esen, A. (1998) Structure and expression of a dhurrinase (β-glucosidase) from Sorghum. Plant Physiol. **116**, 1469–1478
22 Cicek, M., Blanchard, D. J., Bevan, D. R. and Esen, A. (2000) The aglycone specificity determining sites are different in DIMBOA-glucosidase (maize β-glucosidase) and dhurrinase (Sorghum β-glucosidase). J. Biol. Chem. **275**, 20002–20011
23 Cicek, M. and Esen, A. (1999) Expression of soluble and catalytically active plant (monocot) β-glucosidases in *E. coli*. Biotechnol. Bioeng. **63**, 392–400

24 Otwinowski, Z. and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode. In Macromolecular Crystallography, vol. 276, part A (Carter, C. W. and Sweet, R. M., eds.), pp. 307–326, Academic Press, New York

25 CCP4 (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. **D50**, 760–763

26 Navazza, J. (1994) AMoRe: an automated package for molecular replacement. Acta Crystallogr. **A50**, 157–163

27 Brünger, 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. et al. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. **D54**, 905–921

28 Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. **26**, 283–291

29 Wiesmann, C., Hengstenberg, W. and Schulz, G. E. (1997) Crystal structures and mechanism of 6-phospho-β-galactosidase from *Lactococcus lactis*. J. Mol. Biol. **269**, 851–860

30 Sulzenbacher, G., Schülein, M. and Davies, G. J. (1997) Structure of the endoglucanase I from *Fusarium oxysporum*: native, cellobiose, and 3,4-epoxybutyl β-D-cellobioside-inhibited forms, at 2.3 Å resolution. Biochemistry **36**, 5902–5911

31 Tews, I., Perrakis, A., Oppenheim, A., Dauter, Z., Wilson, K. S. and Vorgias, C. E. (1996) Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease. Nat. Struct. Biol. **3**, 638–648

32 Goodman, M. M. and Stuber, C. W. (1983) Maize. In Isozymes in Plant Genetics and Breeding, part B (Tanksley, S. D. and Orton, T. J., eds.), pp. 1–33, Elsevier, New York

33 Miller, S. (1989) The structure of interfaces between subunits of dimeric and tetrameric proteins. Protein Eng. **3**, 77–83

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

35 Jenkins, J., Lo Leggio, L., Harris, G. and Pickersgill, R. (1995) β-Glucosidase, β-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily with 8-fold β/α architecture and with two conserved glutamates near the carboxy-terminal ends of β-strands four and seven. FEBS Lett. **362**, 281–285

36 Vogt, G., Woell, S. and Argos, P. (1997) Protein thermal stability, hydrogen bonds, and ion pairs. J. Mol. Biol. **269**, 631–643

37 McDonald, I. K. and Thornton, J. M. (1994) Satisfying hydrogen bonding potential in proteins. J. Mol. Biol. **238**, 777–793

38 Esen, A. (1992) Purification and partial characterisation of maize (*Zea mays* L.) β-glucosidase. Plant Physiol. **98**, 174–182

39 Rotrekl, V., Nejedla, E., Kucera, I., Abdallah, F., Palme, K. and Brzobohaty, B. (1999) The role of cysteine residues in structure and enzyme activity of a maize β-glucosidase. Eur. J. Biochem. **266**, 1056–1065

40 Davies, G. J., Dauter, M., Brzozowski, A. M., Bjornvad, M. E., Andersen, K. V. and Schülein, M. (1998) Structure of the *Bacillus agaradherans* family 5 endoglucanase at 1.6 Å and its cellobiose complex at 2.0 Å resolution. Biochemistry **37**, 1926–1932

41 Ducros, V., Czjzek, M., Belaich, A., Gaudin, C., Fierobe, H. P., Belaich, J. P., Davies, G. J. and Haser, R. (1995) Crystal structure of the catalytic domain of a bacterial cellulase belonging to family 5. Structure **3**, 939–949

42 Dominguez, R., Souchon, H., Spinelli, S., Dauter, Z., Wilson, K. S., Chauvaux, S., Béguin, P. and Alzari, P. M. (1995) A common protein fold and similar active site in two distinct families of β-glycanases. Nat. Struct. Biol. **2**, 569–576

43 Jacobsen, R. H., Zhang, X. J., DuBose, R. F. and Matthews, B. W. (1994) Three dimensional structure of β-galactosidase from *E. coli*. Nature (London) **369**, 761–766

44 Davies, G. J., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, A. M., Schülein, M. and Withers, S. G. (1998) Snapshots along an enzymatic reaction coordinate: analysis of a retaining β-glycoside hydrolase. Biochemistry **37**, 11707–11713

45 Barton, G. J. (1993) ALSCRIPT: a tool to format multiple sequence alignments. Protein Eng. **6**, 37–40

46 Kraulis, P. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. **24**, 946–950

47 Merritt, E. A. and Murphy, M. E. P. (1994) Raster3D Version 2.0. A program for photorealistic molecular graphics. Acta Crystallogr. **D50**, 869–873

48 Wallace, A. C., Laskowski, R. A. and Thornton, J. M. (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng. **8**, 127–134

49 Roussel, A. and Cambillau, C. (1991) TURBO-FRODO, Silicon Graphics, Mountain View, CA

50 Nicholls, A., Sharp, K. A. and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins **11**, 281–296

Received 16 June 2000/6 November 2000; accepted 1 December 2000