Running Title: Structure of rice Os3BGlu6 β-glucosidase

Corresponding Author:
James R. Ketudat Cairns
Institute of Science, Schools of Biochemistry and Chemistry
Suanaree University of Technology
111 University Avenue, Muang District
Nakhon Ratchasima 30000
THAILAND
Phone: +66 44 224304 Fax: +66 44 224185
Email: cairns@sut.ac.th, jrkcairns@yahoo.com

Research Area:
Biochemical Processes and Macromolecular Structures
Structural and enzymatic characterization of Os3BGlu6, a rice \( \beta \)-glucosidase hydrolyzing hydrophobic glycosides and (1→3)- and (1→2)-linked disaccharides

Supriya Seshadri\(^1\), Takashi Akiyama\(^2\), Rodjana Opassiri\(^1\), Buabarn Kuaprasert\(^3\), and James Ketudat Cairns\(^1\)

\(^1\)Schools of Biochemistry and Chemistry, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand; \(^2\)National Agricultural Research Center for the Hokkaido Region, Sapporo, Japan; \(^3\)Synchrotron Light Research Institute, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand
Foot Notes:

1 Financial support for this work was provided by the Synchrotron Light Research Institute of Thailand. Data collection was conducted at the National Synchrotron Radiation Research Center, a national user facility supported by the National Research Council of Taiwan, ROC. The Synchrotron Radiation Protein Crystallography Facility is supported by the National Research Program for Genomic Medicine.
ABSTRACT

Glycoside hydrolase family 1 (GH1) β-glucosidases play roles in many processes in plants, such as chemical defense, alkaloid metabolism, hydrolysis of cell wall-derived oligosaccharides, phytohormone regulation, and lignification. However, the functions of most of the 34 GH1 gene products in rice (*Oryza sativa* L.) are unknown. Os3BGlu6, a rice β-glucosidase representing a previously uncharacterized phylogenetic cluster of GH1, was produced in recombinant *Escherichia coli*. Os3BGlu6 hydrolyzed *p*-nitrophenyl (*p*NP) β-D-fucoside (*k*₅₀/*K*₅₀ = 67 mM⁻¹s⁻¹), *p*NP-β-D-glucoside (*k*₅₀/*K*₅₀ = 6.2 mM⁻¹s⁻¹), and *p*NP-β-D-galactoside (*k*₅₀/*K*₅₀ = 1.6 mM⁻¹s⁻¹) efficiently, but had little activity toward other *p*NP-glycosides. It also had high activity toward *n*-octyl-β-D-glucoside and β-(1→3)- and β-(1→2)-linked disaccharides, and was able to hydrolyze apigenin β-glucoside and several other natural glycosides. Crystal structures of Os3BGlu6 and its complexes with a covalent intermediate, 2-deoxy-2-fluoroglucoside, and a nonhydrolyzable substrate analogue, *n*-octyl-β-D-thioglucopyranoside, were solved at 1.83, 1.81 and 1.80 Å resolution, respectively. The position of the covalently trapped 2-F-glucosyl residue in the enzyme was similar to that in a 2-F-glucosyl intermediate complex of Os3BGlu7 (rice BGlu1). The side chain of a Met251 in the mouth of the active site appeared to block the binding of extended β-(1→4)-linked oligosaccharides and interact with the hydrophobic aglycone of *n*-octyl-β-D-thioglucopyranoside. This correlates with the preference of Os3BGlu6 for short oligosaccharides and hydrophobic glycosides.
INTRODUCTION

β-Glucosidases (3.2.1.21) have a wide range of functions in plants, including acting in cell wall remodeling, lignification, chemical defense, plant-microbe interactions, phytohormone activation, activation of metabolic intermediates, and release of volatiles from their glycosides (Esen, 1993). They fulfill these roles by hydrolyzing the glycosidic bond at the nonreducing terminal glucosyl residue of a glycoside or an oligosaccharide, thereby releasing glucose and an aglycone or a shortened carbohydrate. The aglycone released from the glycoside may be a monolignol, a toxic compound or a compound that further reacts to release a toxic component, an active phytohormone, a reactive metabolic intermediate, or a volatile scent compound (Dharmawardhama et al., 1995; Morant et al., 2008; Lee et al., 2006; Brzobohatý et al., 1993; Barleben et al., 2007; Reuveni et al., 1999). Indeed, the wide range of glucosides of undocumented functions found in plants suggests many β-glucosidase functions may remain to be discovered.

Plant β-glucosidases fall into related families that have been classified as glycosyl hydrolase (GH) families GH1, GH3 and GH5 (Henrissat, 1991, Coutinho and Henrissat, 1998, 1999). Of these, GH1 has been most thoroughly documented and shown to comprise a gene family encoding 40 putative functional GH in Arabidopsis (Arabidopsis thaliana) and 34 in rice (Oryza sativa), in addition to a few pseudogenes (Xu et al., 2004; Opassiri et al., 2006). In addition to β-glucosidases, plant GH1 members include β-mannosidases (Mo and Bewley, 2002), β-thioglucosidases (Burmeister et al., 1997), and disaccharidases, such as primeverosidase (Mizutani et al., 2002), as well as hydroxyisourate hydrolase, which hydrolyzes the internal bond in a purine ring rather than a glycosidic bond (Raychaudhuri and Tipton, 2002). The specificity for the glycone in GH1 enzymes varies. Some enzymes are quite specific for β-D-glucosides or β-D-mannosides, while many accept either β-D-glucosides or β-D-fucosides, and some also hydrolyze β-D-galactosides, β-D-xylosides and α-L-arabinoside (Esen, 1993). However, most GH1 enzymes are thought to hydrolyze glucosides in the plant, and it is the aglycone specificity that determines the functions of most GH1 enzymes.

Aglycone specificity of GH1 β-glucosidases ranges from rather broad to absolutely specific for one substrate and is not obvious from sequence similarity. For
instance, maize (Zea mays) ZmGlu1 β-glucosidase hydrolyzes a range of glycosides, including its natural substrate, 2-O-β-D-glucopyranosyl-4-dihydroxy-1, 4-benzoxazin-3-one (DIMBOAGlc), but not dhurrin, whereas sorghum (Sorghum bicolor) Dhr1, which is 72% identical to ZmGlu1, only hydrolyzes its natural cyanogenic substrate dhurrin (Verdoucq et al., 2003). Similarly, despite sharing over 80% amino acid sequence identity, the legume isoflavonoid β-glucosidases dalcochinase from Dalbergia cochinchinensis and Dnbglu2 from Dalbergia nigrescens hydrolyze each others’ natural substrate very poorly (Chuankhayan et al., 2007). Thus, small differences in the amino acid sequence surrounding the active site may be expected to account for significant differences in substrate specificity.

GH1 is classified in GH Clan A, which consists of GH families whose members have a (β/α)8 barrel structure with the catalytic acid/base on strand 4 of the β-barrel and the catalytic nucleophile on strand 7 (Jenkins et al., 1995; Henrissat et al., 1995). As such, all GH1 enzymes have similar overall structures, but it has been noted that four variable loops at the C-terminal end of the β-barrel strands, designated A, B, C and D, account for much of the difference in the active site architectures (Sanz-Aparicio et al., 1998). The similar structures with great diversity in substrate specificity make plant GH1 enzymes an ideal model system to investigate the structural basis of substrate specificity. To date, seven plant β-glucosidase structures have been reported, including three closely related chloroplastic enzymes from maize (Czjzek et al., 2000, 2001), sorghum (Verdoucq et al., 2004) and wheat (Triticum aestivum, Sue et al., 2006), the cytoplasmic strictosidine β-glucosidase from Rauvolfia serpentine (Barleben et al., 2007), and the secreted enzymes white clover (Trifolium repens) cyanogenic β-glucosidase (Barrett et al., 1995), white mustard (Sinapsis alba) myrosinase (thioglucosidase, Burmeister et al., 1997), and rice Os3BGlu7 (BGlu1, Chuenchor et al., 2008). These enzymes hydrolyze substrates with a range of structures, but they cannot account for the full range of β-glucosidase substrates available in plants, and determining the structural differences that bring about substrate specificity differences in even closely related GH1 enzymes has proven tricky (Verdoucq et al., 2003, 2004; Sue et al., 2006; Chuenchor et al., 2008).

Amino acid sequence based phylogenetic analysis of GH1 enzymes encoded by the rice genome showed that there are eight clusters containing both rice and
Arabidopsis proteins that are more closely related to each other than they are to enzymes from the same plants outside the clusters (Fig. 1, Opassiri et al., 2006). In addition, there are a cluster of sixteen putative β-glucosidases and a cluster of myrosinases in Arabidopsis without any closely related rice counterparts. Comparison to characterized GH1 enzymes from other plants reveals other clusters of related enzymes not found in rice or Arabidopsis, including the chloroplastic enzymes, from which the maize, sorghum and wheat structures are derived, and the cytoplasmic metabolic enzymes, from with the strictosidine hydrolase structure is derived (Fig. 1). Therefore, although the known structures provide good tools for molecular modeling of plant enzymes, most rice and Arabidopsis GH1 enzymes lack a close correspondence in sequence and functional evolution to these structures, suggesting the variable loops that determine the active site may be different. It would, therefore, be useful to know the structures and substrate specificities of representative members of each of the 8 clusters seen in rice and Arabidopsis. To begin to acquire this information, we have expressed Os3BGlu6, a member of cluster At/Os 1 in Fig. 1, characterized its substrate specificity and determined its structure alone and in complex with a glycosyl intermediate and a nonhydrolyzable substrate analogue.

RESULTS

Cloning, expression and purification of Os3BGlu6:

The cDNA encoding the Os3BGlu6 protein was isolated from a cold-treated rice library, and the sequence determined (Genbank accession number AY129294). This full length cDNA contained an open reading frame of 1563 nucleotides encoding 521 amino acids, which were predicted to include an N-terminal secretory signal peptide of 38 residues by SignalP (Bendtsen et al., 2004). The Os3BGlu6 protein was expressed in *Escherichia coli* as a fusion protein with N-terminal thioredoxin and His<sub>6</sub> tags, from which it was released by cleavage of a TEV protease site that was placed 4 amino acid residues before precursor residue 38. This allowed a soluble and active Os3BGlu6 β-glucosidase of 55 kDa to be conveniently purified to >95% purity, based on SDS PAGE (Suppl. Fig. 1).
The activity of the expressed Os3BGlBu6 was highest at pH 4 to 5, dropped by 50% at pH 2.5 and 6.0, and was negligible from pH 7.0 upward. The enzyme had highest activity over the temperature range of 40°C to 55°C, but was most stable at 20°C to 30°C and began to lose activity after 40 min at 40°C or higher. Therefore, standard assays were conducted at 30°C, which is physiologically relevant for rice.

**Substrate specificity of Os3BGlBu6**

The ability of OsBGlu6 to hydrolyze p-nitrophenyl (pNP) glycosides was tested to assess its glycone specificity (Table I). Among the pNP glycosides, Os3BGlBu6 preferred pNP-β-D-fucopyranoside (pNPFuc) with a $k_{cat}/K_m$ of 67 mM$^{-1}$s$^{-1}$, due to its relatively low $K_m$ of 0.50 mM, followed by pNP-β-D-glucoside (pNPGLc) and pNP-β-D-galactoside (pNPGal) with $k_{cat}/K_m$ values of 6.2 and 1.6 mM$^{-1}$s$^{-1}$, respectively. Hydrolysis of α-D-glucoside was negligible, but α-L-arabinoside was hydrolyzed at a low rate.

Hydrolysis of synthetic and natural glycosides and oligosaccharides were tested to assess its aglycone specificity (Table II). Os3BGlBu6 hydrolyzed n-octyl-β-D-glucoside with the highest catalytic efficiency ($k_{cat}/K_m = 2.7$ mM$^{-1}$s$^{-1}$), which was approximately 3-fold higher than that for n-heptyl-β-D-glucoside (0.85 mM$^{-1}$s$^{-1}$). Among the disaccharides, Os3BGlBu6 hydrolyzed the β-(1→3)-linked laminaribiose best, followed by β-(1→2)-linked sophorose, β-(1→6)-gentiobiose and β-(1→4)-linked cellobiose, in order of decreasing catalytic efficiency ($k_{cat}/K_m$). Longer β-(1→3)- and β-(1→4)-linked oligosaccharides were hydrolyzed at decreasing rates. Os3BGlBu6 also hydrolyzed natural glycosides, including apigenin-7-glucoside, glycitin, diadzin, genistin, esculin, arbutin, coumaryl alcohol β-D-glucoside, coniferin, and salicin, although at low rates. The cyanogenic glycosides, amygdalin, linamarin, and prunacin and several other glycosides (listed at the bottom of Table II) were not detectably hydrolyzed after reaction overnight.

**Overall structure of Os3BGlBu6**

To understand the structural basis for the substrate specificity of Os3BGlBu6, its 3D structures alone and in complexes with 2-deoxy-2-fluoroglucoside (a slowly hydrolyzed covalent intermediate) and n-octyl-β-D-thioglucopyranoside (a nonhydrolyzable substrate analogue) were determined by protein crystallography. The
data processing and refinement parameters are provided in Table III. The overall structure is a \((\beta/\alpha)_8\) barrel, similar to those of other known structures of GH1 enzymes, with the catalytic acid/base, Glu178, and nucleophile, Glu394, located on \(\beta\)-strands four and seven of the \(\beta\)-barrel (Fig. 2A). The active site forms a slot-like cleft in the side of the protein, the bottom of which is negatively charged (Fig. 2B). A single protein molecule was present in the asymmetric unit, and a lack of close crystal contacts indicated that Os3BGlu6 is monomeric.

N-terminal sequencing of the first eight amino acid residues of the protein recovered from crystals gave a sequence consistent with that predicted from the TEV cleavage site, with 4 residues from the cloning vector and primers, Ser1, Phe2, Thr3, and Met4, followed by the first 4 residues included from the protein sequence, Ala5, Gln6, Gln7, and Ser8. However, these residues, together with Gly9 and Gly10, had no detectable electron density, likely due to high flexibility. The C-terminus was well defined, except that the side chains of the last two residues, Lys487 and Thr488, had poor density. In addition, poor density was observed for the side chains of residues 334 to 337 in loop C in all three structures, although the Ca backbone was clearly visible. This loop is apparently flexible, since it is exposed at the outer surface of the protein and the position of this loop was slightly different in the case of the Os3BGlu6/\(n\)-octyl-\(\beta\)-D-thioglucopyranoside. No density was observed for 6 of the atoms of the surface residues in Lys114, Glu296, Glu383, and Lys432 in the native and Os3BGlu6/G2F structures, so these atoms were deleted from the structures. Due to the high resolution of the model, multiple conformations were observed for the flexible side chains of residues Arg233, Thr234, Asn332, and Glu451 in native Os3BGlu6 and Asn241, Ser231, Arg293, Asp361, and Glu433 in the Os3BGlu6/G2F complex. One disulfide bond was seen linking Cys197 and Cys205 in loop B, a feature which is conserved in plant GH1 enzymes. Two \(cis\)-peptide bonds were found between Ala193 and Pro194 and between Trp444 and Ser445, which is also common in GH1.

The active site pocket of the native protein structure had extra electron density, which could not be fit to glycerol or a component of the precipitant, but was best fit by a molecule of Tris in two possible conformations (Fig 2 C) with an average B-factor of 28.81 \(\text{Å}^2\). Tris was a component of the protein storage buffer and was
confirmed to bind by its competitive inhibition of Os3BGlu6 with a $K_i$ of 5.1± 0.2 mM. Six glycerol molecules from the cryoprotectant were also seen on the surface of the protein in the native structure. Four of these were also seen in the Os3BGlu6/n-octyl-β-D-thioglucopyranoside complex and two in the Os3BGlu6/G2F complex.

Among known GH1 structures, Os3BGlu6 was most similar to the plant secreted β-glucosidases, white clover cyanogenic β-glucosidase (1CBG, Barrett et al., 1995) and rice Os3BGlu7 β-glucosidase (2RGL, Chuenchor et al., 2008), with the root mean squared deviation (RMSD) of their Cα to those of Os3BGlu6 being 0.81 Å for both structures, over 467 and 444 Cα atoms, respectively. For other plant enzymes, the RMSD ranged from 0.83 Å over 445 Cα atoms for white mustard myrosinase (1E4M, Burmeister et al., 2000) to 0.96 Å over 451 atoms for maize ZmGlu1 β-glucosidase (1E1E, Czjzek et al., 2001), which was the same as its RMSD over 423 atoms with human cytosolic β–glucosidase (2E9M, Hayashi et al., 2007). Not surprisingly, the bacterial structures have higher RMSD with Os3BGlu6, such as 1.13 Å over 397 Cα atoms for Bacillus polymxa β-glucosidase (1BGA, Sanz-Aparicio et al., 1998) and 1.31 Å over 354 atoms for Lactococcus lactis 6-phospho-β-galactosidase (1PBG, Wiesmann et al., 1995). Although all the above mentioned structures are similar, many differences reside in the loop regions around their active sites where the glycone and aglycone binding sites are located.

3D structures of Os3BGlu6 complexes with G2F and n-octyl-β-D-thioglucopyranoside

To investigate the structural basis for substrate binding and hydrolysis, the structures of Os3BGlu6 in a covalent intermediate with G2F and noncovalently bound to n-octyl-β-D-thioglucopyranoside were solved at 1.81 and 1.80 Å resolution, respectively. The crystals of the covalent G2F intermediate were isomorphous with the native structure, but the crystal unit cell expanded with n-octyl-β-D-thioglucopyranoside, as previously seen in crystals with n-octyl-β-D-glucopyranoside (McPherson et al., 1986). An increase of 9.5 Å was seen in the c side of the unit cell, due to a displacement of the protein molecules related by symmetry along the z-axis by 4.5 to 5 Å in their water-mediated interactions, but the only significant change in the monomer structure was the conformation of loop C, which is not involved in these interactions.
In the intermediate complex, the density for the 2-deoxy-2-fluoroglucosyl residue in a relaxed $^4C_1$ chair conformation covalently bound to the catalytic nucleophile was clearly evident in the glycone-binding subsite, surrounded by the universally conserved plant GH1 residues Gln31, His132, Asn177, Glu178, Glu394, Glu451 and Trp452 (Fig. 3 A and B). The residues Glu451, Gln31, Tyr321, and Trp452 formed direct hydrogen bonds with G2F. The Oe1 and Oe2 of Glu451 were linked to O4 and O6 of the glucose ring at 2.60 and 2.65 Å, which held Glu451 in a single conformation, compared to the two conformations of Glu451 seen in the native structure. The Gln31 Oe1 and Ne2 were hydrogen bonded to the glucosyl O3 and O4 at 2.51 and 2.92 Å, respectively. In addition, the hydroxyl group of Tyr321 was hydrogen bonded to O5 (2.93 Å) and Trp452 Ne1 to O3 (2.84 Å). A hydrophobic stacking interaction between the sugar and the indole ring of Trp444 was observed, as seen in other GH family 1 glycosidases. Superimposition of the G2F complex against the native enzyme (Fig. 3 A) showed that the hydroxyl group of Tyr321, which was 2.75 Å from Oe1 of the catalytic nucleophile Glu394 in the native structure, twisted away in the G2F complex to form a hydrogen bond to O5 of the sugar (2.93 Å). The nucleophile side chain also moved in producing the covalent intermediate, so that Glu394 Oe1 and Oe2, which were positioned at 4.59 Å and 3.88 Å from Oe2 of Glu178 in the native structure, were distanced at 3.99 and 3.25 Å from it in the G2F complex. In addition, the indole ring of Trp452 in the G2F complex was shifted slightly inward compared to the native structure to allow its nitrogen to hydrogen bond with O3 of the sugar ring at 2.84 Å.

In the complex of Os3BGlu6 with the nonhydrolysable substrate analogue $n$-octyl-$\beta$-D-thioglucopyranoside, the inhibitor was located deep in the active site slot (Fig. 2 B, 3C and 4 A). The density for the sugar was well defined (Fig 3C). For the aglycone moiety, the density was strongest for C1', C4' and C8', and was lower for C2', C3', C5', C6' and C7', likely due to the flexibility of the carbon chain, which appeared to be able to bind in multiple positions. Although both the relaxed $^4C_1$ chair and less energetically favorable $^1S_3$ skew boat conformations fit the density well, the distance of Oe1 of the acid-base catalyst Glu178 to the sulfur in the $^4C_1$ chair was 2.13 Å, which was unacceptably close. In contrast, this distance was 3.07 Å in the $^1S_3$ conformation putting the sulfur within distance to accept a proton from the catalytic acid. The glycone residue was in a similar position to that of G2F in its complex and
formed similar hydrogen bonding interactions, except that interactions of the O2 hydroxyl with His132, Asn177 and Glu394 were also evident (Fig. 3 B and C). The aglycone was flanked by hydrophobic residues and a few polar residues. The n-octyl chain was bound by hydrophobic contacts from Tyr321 and Trp366 on one side, and Val250 and Met251, which also form the entrance to the active site, on the other side of the chain. The polar residues Thr322, Thr181, Asp249 His320, Asn319, and Gln273 also lay on either side of the aglycone.

DISCUSSION

**Os3BGlu6 substrate glycone specificity**

Os3BGlu6 showed a strong preference for β-D-fucoside, followed by β-D-glucoside, and then by β-D-galactoside, as was seen for several other plant β-D-glucosidases (Babcock and Esen, 1994; Srisomsap et al., 1996; Opassiri et al., 2004, 2006). Like maize β-glucosidase (Babcock and Esen, 1994) Os3BGlu6 shows a 10-fold higher $k_{cat}/K_m$ for $p$NP-Fuc than for $p$NP-Glc. In maize and rice Os3BGlu7 β-glucosidases, the preference for $p$NP-Fuc over $p$NP-Glc is due to a higher $k_{cat}$ for $p$NP-Fuc, (Opassiri et al., 2004), whereas for Os3BGlu6, the difference is due to a 10-fold lower $K_m$ for $p$NP-Fuc than $p$NP-Glc, similar to *Dalbergia cochinchinensis* β-glucosidase, which had $K_m$ values of 0.54 mM for $p$NP-Fuc and 5.4 mM for $p$NP-Glc (Srisomsap et al., 1996). In Os3BGlu6, this appears to be partly due to a higher tolerance for the equatorial position of the 4’ hydroxyl group in Os3BGlu6, since it had a $K_m$ for $p$NP-Gal similar to that for $p$NP-Glc, whereas Os3BGlu7 had a $K_m$ over 10-fold higher for $p$NP-Gal than $p$NP-Glc. In contrast, Os3BGlu6 had a much lower efficiency ($k_{cat}/K_m$ of 0.06 mM$^{-1}$s$^{-1}$) for hydrolysis of $p$NP-β-D-mannopyranoside than did Os3BGlu7 ($k_{cat}/K_m$ of 1.01 mM$^{-1}$s$^{-1}$), indicating the equatorial position of the 2’ hydroxyl group is more critical to Os3BGlu6. Among natural substrates, Os3BGlu6 would be expected to hydrolyze β-D-glucosides, since β-D-fucosides are generally not found in plants.

In reported GH1 complex structures, the glutamate corresponding to Glu451 hydrogen bonds to both the C4 and C6 hydroxyls of the sugar residue in the -1 site (Burmeister et al., 1997; Sanz-Aparacio et al., 1998; Czjzek et al., 2000), as seen in
the Os3BGLu6 complex with G2F. The flexibility of Glu451 in the native structure may help explain the relatively low $K_m$ for $p$NPFuc, since it may more readily bind to the axial 4-hydroxyl and the loss of entropy that occurs upon its hydrogen bonding to the 4- and 6- hydroxyls may decrease its energetic contribution to binding. In this case, the absence of hydrogen bonding to the 6-hydroxyl in $p$NPFuc would have less effect than in Os3BGLu7, in which the corresponding Glu440 residue maintains a similar position in the presence and absence of glucoside (Chuenchor et al., 2008).

**Os3BGLu6 Aglycone Specificity**

The hydrolysis of glucooligosaccharides and glucosides followed a rather unique pattern in Os3BGLu6, although the $K_m$ values of over 3 mM, suggest the substrates tested would only be hydrolyzed efficiently at high concentrations. The hydrolysis of $\beta$-(1→3)- and $\beta$-(1→2)-linked disaccharides in preference to longer oligosaccharides reflects the relatively short binding cleft, as will be discussed below in comparison to Os3BGLu7. The hydrolysis of the $n$-octyl-$\beta$-D-glucoside with high efficiency suggests that Os3BGLu6 may act on hydrophobic glycosides in the plant, rather than or in addition to disaccharides. Among the natural glycosides tested, the flavonoid glycoside apigenin-7-O-$\beta$-D-glucoside was hydrolyzed most rapidly, followed by 7-O-$\beta$-linked isoflavonoids, which suggests that a flavonoid-7-O-glucoside could serve as a hydrophobic substrate in the plant. The apoplastic or vacuolar location of action suggested by its low pH optimum and secretory signal sequence is consistent with its action on either disaccharides or flavonoids or other hydrophobic glucosides, which may be found in these compartments.

Among plant GH1 enzymes, the aglycone specificities of the chloroplastic $\beta$-glucosidases from maize and sorghum have been most extensively studied (Czjzek et al., 2000, 2001; Verducq et al., 2003, 2004). The crystal structures of the inactive maize ZmGlu1 E191D mutant with DIMBOAGlc (PDB: 1E56, Czjzek et al., 2000) and Os3BGLu6 with $n$-octyl-$\beta$-D-thioglucopyranoside were compared (Fig 4 A and B). The residues that directly interact with the aglycone in ZmGlu1 include Trp378 on one side and three phenylalanines, Phe198, Phe205 and Phe466, on the opposite side of the active-site slot (Czjzek et al., 2000). Although the tryptophan was conserved in Os3BGLu6 (Trp366), the opposing amino acid residues were replaced by Gln185, Gln192 and Ala453 in Os3BGLu6, and the aglycone of $n$-octyl-$\beta$-D-
thioglucopyranoside is located on the opposite side of the active site, where it makes hydrophobic contacts with other residues. Among the residues noted to indirectly affect aglycone binding in ZmGlu1, Ala467 is conserved (Ala454 in Os3BGlu6), while Tyr473, which orients the plane of Trp378 for stacking interaction with the DIMBOA substrate, is replaced by Phe460, which has no hydroxyl to hydrogen bond to the indole ring of Trp366 in Os3BGlu6. Thus, Trp366 is twisted away by 1.13 Å for its interaction with the n-octyl chain of Os3BGlu6. These differences in aglycone binding may in turn affect the position and orientation of the glycone, as noted by Verdoucq et al. (2004).

**Comparison of substrate binding between Os3BGlu6 and Os3BGlu7**

Rice Os3BGlu6 and Os3BGlu7 have different catalytic specificities that are reflected in their active site structures. While Os3BGlu6 hydrolyzed n-heptyl and n-octyl β-D-glucosides and sophorose more efficiently ($k_{cat}/K_m$ of 0.85, 2.7, and 0.95 s$^{-1}$mM$^{-1}$ for Os3BGlu6 and 0.62, 1.25 and 0.42 s$^{-1}$mM$^{-1}$ for Os3BGlu7, respectively), Os3BGlu7 more efficiently hydrolyzed laminaribiose ($k_{cat}/K_m$ of 15.7 s$^{-1}$mM$^{-1}$ vs. 1.7 s$^{-1}$mM$^{-1}$ for Os3BGlu6) and cellobiose ($k_{cat}/K_m$ of 0.05 vs. 0.0085 s$^{-1}$mM$^{-1}$ for Os3BGlu6). More significantly, while Os3BGlu7 hydrolyzed longer β-(1→4)-linked oligosaccharides with increasing efficiency, up to 6 glucosyl residues ($k_{cat}/K_m$ of 153 s$^{-1}$mM$^{-1}$ for cellohexaose, Opasiri et al., 2004), Os3BGlu6 hydrolyzed longer cellooligosaccharides more poorly than cellobiose (Table II).

Comparison of the Os3BGlu6 structure with that of Os3BGlu7 docked with the β-(1→4)-linked substrate cellotriose (Chuenchor et al., 2008) revealed differences in the shape of the active site and the residues lining the aglycone binding region (Fig. 4 A and C). While Os3BGlu7 has a deep and relatively straight binding cleft, that of Os3BGlu6 is shorter and narrower. The distance from the deepest part of the active site slot to the surface entrance is 14 Å, which is shorter by 4 Å from that of Os3BGlu7 (18 Å), consistent with the preference of Os3BGlu6 for shorter oligosaccharides. The Os3BGlu6 glycone binding pocket is 7.5 Å deep and 9 Å wide, and is surrounded by residues conserved with Os3BGlu7 and other GH1 enzymes. The broadest region of the aglycone binding site is 7.6 Å, while the narrowest part of aglycone binding site, i.e. at the entrance to the active site is 5.3 Å, at the point where Met251 extends into what would be the +2 subsite in Os3BGlu7. Met251, which
appears to form hydrophobic interactions with the \( n \)-octyl chain of \( n \)-octyl-\( \beta \)-D-thioglucopyranoside, replaces Asn245 of Os3BGlu7, which appears to interact with cellotriose, based on the 40-fold increase in \( K_m \) in the Os3BGlul7 N245V mutation (Cheunchor et al., 2008). Aside from the Met251, many differences were observed in the residues involved in the aglycone site, including Trp133, Thr181, Ile184, Gln185, Gln192, and Ala453 in Os3BGlul6 in place of Tyr131 Ile179, Leu182, Leu183, Asn190, and Leu442 in Os3BGlul7, respectively. These differences may help explain the fact that Os3BGlul6 prefers hydrophobic \( \beta \)-glucosides and short oligosaccharides, while Os3BGlul7 prefers longer \( \beta \)-(1→4)-linked oligosaccharides.

**CONCLUSION**

As plant GH1 enzymes exhibit a wide variety of specificities and release products with a wide range of physiological functions, the structural basis for this functional diversity is a significant problem. The enzymatic characterization of Os3BGlul6, along with the structures of Os3BGlul6, its covalent intermediate with G2F, and its noncovalent complex with the substrate analogue \( n \)-octyl-\( \beta \)-D-thioglucopyranoside extend the understanding of this problem to a new cluster of GH1 in plants. Small differences in the active site amino acids appear to explain the difference between rice Os3BGlul7, which prefers long, straight glucooligosaccharides, and Os3BGlul6, which prefers shorter hydrophobic glycosides. By extension, the small differences in the active sites of plant GH1 enzymes and their impact on the substrate specificity can help explain the expansion of this family to fulfill a range of functional roles in the plant.

**MATERIALS AND METHODS**

**Cloning and PCR amplification of cDNA encoding mature Os3BGlul6**

To produce a cDNA library, 7-day-old rice cultivar Yukihikari seedlings were chilled at 5°C for a further 4 days, and then the RNA was extracted by the method of Bachem et al. (1996). Poly-A RNA was isolated from the total RNA with magnetic poly-T beads (Dynal). The poly-A RNA was reverse transcribed and used to produce a library with the \( \lambda \)ZAP-cDNA/Gigapack cloning kit (Stratagene). This library was differentially screened with cDNA reverse transcribed from cold stressed seedling and control seedling RNA, which had been labeled by random priming with \( \alpha \)-\([\text{32P}]\)dCTP.
with the Rediprime II random prime labeling system (GE Healthcare). The GH1
cDNA encoding Os3BGlu6 was selected as a differentially expressed clone,
sequenced thoroughly and submitted to Genbank as accession number AY129294.

The cDNA encoding the predicted mature Os3BGlu6 rice \( \beta \)-glucosidase, was
amplified by PCR with the specific primers SapporoMstrtF:
CACCATGGGCAGCAGAGCGGAG and SapporoStopR:
GGAGTTCAGGTCTTCAGGAG. Amplification was performed with 30 cycles of
94°C 1 min, 60°C 45 s and 72°C 1 min 30 s, with \( Pfu \) DNA polymerase (Promega).
The PCR product was cloned into pENTR™/TEV/D-TOPO®, according to the
supplier’s directions (Invitrogen). The cDNA was subcloned into the
pET32a(+)/DEST expression vector (Opassiri et al., 2006) by LR Clonase™
recombination (Invitrogen). The recombinant pET32a(+)/DEST/Os3BGlut6
expression clone was verified by automated DNA sequencing with the T7 promoter
and terminator and specific internal primers (F427:
CTTGACAGGCAGATAGTGGA, R600: GCAGTAGAGGTGGAGCAG, F969:
GCACAACACACCAACAC and R946: CGTGTAGTAGGTGTGTAGTG).

**Expression of rice \( \beta \)-glucosidase Os3BGlut6 in *Escherichia coli* and purification**

Os3BGlut6 was expressed in *E. coli* Origami(DE3) as a fusion protein with N-
terminal thioredoxin and His\(_6\) tags. Expression was induced at 20°C and purified by
immobilized metal (Co\(^{2+}\)) affinity chromatography (IMAC, on Talon resin, Clontech)
as previously described for rice OsBGlu7 (Chuenchor et al., 2006). The active elution
fractions containing \( p \)NPGlc hydrolysis activity were pooled and concentrated and the
buffer changed to 20 mM Tris-HCl, pH 8.0, in a 30 kDa cutoff ultrafiltration
membrane (Amicon YM-30, Millipore). The N-terminal thioredoxin, His\(_6\), and S tags
were then excised with tobacco etch virus (TEV) protease, produced as described by
Kapust et al. (2001), at a ratio of 1:50 (w/w) TEV protease to fusion protein for 12 to
14 h at 4°C. The cleaved N-terminal fusion tag and TEV protease were removed by
adsorption to Co\(^{2+}\) IMAC resin. The flow-through and 5 mM and 10 mM imidazole
wash fractions containing activity were pooled, the buffer was changed to 20 mM
Tris-HCl, pH 8.0, and the pool concentrated by ultrafiltration (Amicon YM-10,
Millipore). Protein concentrations were determined by the Bradford protein assay
(Bio-Rad) with bovine serum albumin as the standard.
Os3BGlul6 pH and temperature optimum & stability studies

The optimum pH was determined by incubating 1 μg enzyme in a reaction volume of 140 μL containing 100 mM universal buffer (citric acid-disodium hydrogen phosphate), pH 2-11, at 0.5 pH unit increments with 1 mM pNPGlc for 10 min at 30°C. The reaction was stopped with 100 μl of 2 M sodium carbonate. The amount of p-nitrophenol (pNP) released was determined from its absorbance at 405 nm. The enzyme’s pH stability was studied by incubating 25 μg enzyme in 20 μL universal buffer at the same pH range for time periods of 10 min, 1, 2, 6, 12 and 24 h at 30°C. The enzyme was diluted 400 fold in 100 mM buffer at the optimum pH, and the activity was determined as described above. The temperature optimum was determined by incubating 1 μg enzyme in 100 mM sodium acetate, pH 5.0, over a temperature range of 5 to 100°C in increments of 5 degrees for 10 min and then pNPGlc was added to the reaction to 1 mM final concentration and incubated for 10 min. Temperature stability was determined by preincubating 1 μg enzyme over a temperature range of 5 to 70°C. At 10 min increments from 0 to 60 min, aliquots of the samples were placed on ice and the enzyme activity at 30°C was determined as described above.

Substrate specificity and enzyme kinetics

Os3BGlul6 was tested for hydrolysis of glycosides of pNP and commercially available natural and synthetic substrates (listed in Tables I and II) by measuring release of pNP or glucose from the substrates and by thin layer chromatography (TLC). Activity was assayed in triplicate in 100 mM sodium acetate, pH 5.0, at 30°C. For TLC, the reactions were stopped after overnight digest by boiling 5 min, and products from the reactions were spotted on silica gel F254 plates (Merck) and chromatographed with solvents of ethyl acetate:acetic acid:water (2:1:1, v/v) for the oligosaccharides and ethyl acetate:acetic acid:methanol:water (15:2:1:2, v/v) for natural glycosides. The products were visualized under UV light at 254 nm, and then by coating the plates with 10% (v/v) sulfuric acid in methanol and heating at 120°C until the spots were visible. For oligosaccharides and glucosides other than pNPGlc that were detectably hydrolyzed on TLC, glucose release was measured by the glucose-oxidase/peroxidase coupled assay, as previously described (Opassiri et al., 2003).

17
Kinetic parameters were determined for substrates of interest, based on the initial relative activity assays. The initial velocities ($V_0$) were determined at time points and enzyme concentrations where the reaction rate was linear and the absorbance value was in the range of 0.1 to 1.0, and used to calculate the kinetic constants. Substrate concentrations over a range of approximately 1/5 to 5 times the apparent $K_m$ were included. The $k_{cat}$, $K_m$, and $k_{cat}/K_m$ were calculated by nonlinear regression of Michaelis-Menten plots with Grafit 5.0 (Erithacus Software). Competitive $K_i$ values were determined by incubating 1 μg enzyme with eight different concentrations of the inhibitors (0 to 30 mM) in 100 mM sodium acetate buffer, pH 5.0, in the presence of 0 to 20 mM pNPGlc substrate under the reaction conditions described above. Lineweaver-Burk and Dixon plots were used to calculate the inhibition constants.

**Screening and optimization for Os3BGlul6 crystallization**

Purified Os3BGlul6 in 20 mM Tris-HCl, pH 8.0, was filtered through an Ultrafree MC 0.22 μm filter (Millipore) and used to screen for crystallization conditions by the microbatch under oil method at 15°C with precipitants from the Crystals Screen High Throughput HR2-130 and HR2-134 kits (Hampton Research). Clusters of crystals were seen in the condition with 0.1 M Bis-Tris, pH 6.5, 20% polyethylene glycol monomethyl ether (PEG MME) 5000 (in the HR2-134 kit) after 200 days. Optimization with microseeding yielded diffraction quality native crystals of Os3BGlul6 in 21% PEG5000 MME, 0.1 M Bis-Tris, pH 6.5. The crystals were soaked in the cryoprotectant (24.15% PEG5000 MME, 0.115 M Bis-Tris pH 6.5, and 18% (v/v) glycerol), for 5 to 10 s before flash freezing in liquid nitrogen. To obtain Os3BGlul6 crystals with inhibitors, native crystals were soaked in 5 mM 2-deoxy-2-fluoro-β-D-glucoside (DNPG2F) or 10 mM n-octyl-β-D-thioglucopyranoside overnight in 19% PEG5000 MME, 0.1 M Bis-Tris, pH 6.5. The cryoprotectants contained the same concentrations of inhibitors present in 21.85% PEG5000 MME, 0.115 M Bis-Tris, pH 6.5, and 18% v/v glycerol.

**Data Collection, processing and structure refinement**

The datasets for native Os3BGlul6, the Os3BGlul6/G2F complex and the Os3BGlul6/n-octyl-β-D-thioglucopyranoside complex were collected on the PX 13B beamline at the National Synchrotron Radiation Research Center, Hsinchu, Taiwan with a 1.0 Å
wavelength x-ray beam and an ADSC Quantum 315 CCD detector. The crystals were maintained at 110K with a cold stream of nitrogen (CryoJet, Oxford Instrument, Oxford, UK) throughout diffraction. All datasets were indexed, integrated and scaled and with the HKL-2000 package (Otwinski & Minor, 1997). The crystals belonged to space group P2$_1$2$_1$2$_1$ with the unit cell parameters shown in Table III. The crystal structure for native Os3BGlu6 was solved by molecular replacement with the *Trifolium repens* cyanogenic β-glucosidase structural model (PDB code 1CBG, Barrett et al., 1995) with the MolRep program in the CCP4 suite (Vagin and Teplyakov, 1997). A subset of 5% of the structure factor amplitudes were reserved for the $R_{\text{free}}$ determination. Iterative refinement was done with Refmac 5.0 (Murshudov et al., 1997; CCP4 suite), and the structure was built into the electron density map with the coot graphics program (Emsley & Cowtan, 2004). The overall quality of each model was evaluated with the program PROCHECK (Laskowski et al., 1993). The structures of the Os3BGlu6/G2F and Os3BGlu6/n-octyl-β-D-thioglucopyranoside complexes were solved by molecular replacement with the native structure as the template model. Structural figures were generated in Pymol (Delano Scientific). Structural comparisons between Os3BGlu6 and other known structures from GH1 β-glucosidase were executed with the Superpose program of the CCP4 suite (Krissinel and Henrick, 2004).

**N-terminal amino acid sequencing of Os3BGlu6**

Os3BGlu6 crystal clusters were dissolved in sample buffer and separated on a 12% polyacrylamide gel by the method of Laemmli (1970). The gel was electroblotted onto a Biotrace™ polyvinylidene flouride (PVDF) membrane (Pall Corporation) and stained with Coomassie Brilliant Blue (R250). The protein bands were excised and Edman degradation sequencing was performed at the Prince of Songkla University Protein Analysis Center.

**ACKNOWLEDGEMENT**

The authors are grateful to Phimonphan Chuankhayan and Chun-Jung Chen and his laboratory for assistance in crystal screening and the staff of the NSRRC PX 13B beamline for assistance in data collection. Robert C. Robinson is thanked for data processing advice. Prof. Jisnuson Svasti is thanked for advice and support on the project and providing dalcochinin glucoside.
LITERATURE CITED

Babcock GD, Esen A (1994) Substrate specificity of maize β-glucosidase. Plant Sci 101: 31-39

Bachem CW, van der Höven RS, De Bruijin SM, Vreugdenhil D, Zabeau M, Visser RG (1996) Visualization of differential gene expression using novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. Plant J 9: 745-753

Barleben L, Panjikar S, Ruppert M, Koepke J, Stockigt J (2007) Molecular architecture of strictosidine glucosidase: The gateway to the biosynthesis of the monoterpenoid indole alkaloid family. Plant Cell 19: 2886–2897

Barrett T, Suresh CG, Tolley SP, Dodson EJ, Hughes MA (1995) The crystal structure of a cyanogenic β-glucosidase from white clover, a family 1 glycosyl hydrolase. Structure. 3: 951-960

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

Burmeister WP, Cottaz, S, Driguez H, Palmieri S, Henrissat B (1997) The crystal structures of Sinapis alba myrosinase and of a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase. Structure 5: 663-675

Burmeister, WP., Cottaz, S., Rollin P., Vasella, A., Henrissat, B (2000) High resolution X-ray crystallography shows that ascorbate is a cofactor for myrosinase and substitutes for the function of the catalytic base. J Biol Chem 275: 39385-39393

Brzobohatý B, Moore I, Kristoffersen P, Bako L, Campos N, Schell J, Palme K (1993) Release of active cytokinin by a β-glucosidases localized to maize root meristem. Science 262: 1051-1054

Chuankhayan P, Rimlumduan T, Tantanuch W, Mothong N, Kongsaeree PT, Methenukul P, Svasti J, Jensen ON, Ketudat Cairns JR (2007) Functional and structural differences between isoflavonoid β-glycosidases from Dalbergia sp. Arch Biochem Biophys 468: 205-216
Chuenchor W, Pengthaisong S, Yuvaniyama J, Opassiri R, Svasti J, Ketudat-Cairns JR (2006) Purification, crystallization and preliminary X-ray analysis of rice BGlu1 β-glucosidase with and without 2-deoxy 2-fluoro β-D-glucoside, Acta Crystallogr F 62: 708-801

Chuenchor W, Pengthaisong S, Robinson R.C., Yuvaniyama J, Oonanant W, Bevan D.R., Esen A, Chen C.J., Opassiri R, Svasti J, Ketudat-Cairns JR (2008) Structural insights into rice BGlu1 β-glucosidase oligosaccharide hydrolysis and transglycosylation. J Mol Biol 377: 1200–1215

Czjzek M, Cicek M, Zamboni V, Bevan DR, Henrissat B, Esen A (2000) The mechanism of substrate (aglycone) specificity in β-glucosidase is revealed by crystal structures of mutant maize β-glucosidase-DIMBOA, -DIMBOAGlc, and -dhurrin complexes. Proc Natl Acad Sci, USA 97: 13555-13560

Czjzek M, Cicek M, Zamboni V, Burmeister WP, Bevan DR, Henrissat B, Esen A (2001) Crystal structure of a monocotyledon (maize ZMGlu1) beta-glucosidase and a model of its complex with p-nitrophenyl beta-D-thioglucoside. Biochem J 354: 37–46

Coutinho PM, Henrissat B (1998-2009) CAZY: Carbohydrate Active Enzymes. http://www.cazy.org

Coutinho PM, Henrissat B (1999) Carbohydrate-active enzymes: an integrated database approach. In Davies GJ, Gilbert HJ, Henrissat B, Svensson B eds, Recent Advances in Carbohydrate Engineering, The Royal Society of Chemistry, Cambridge, pp 3-11

Dharmawardhama DP, Ellis BE, Carlson JE (1995) A β-glucosidase from lodgepole pine xylem specific for the lignin precursor coniferin. Plant Physiol 107: 331-339

Emsley P, Cowtan K (2004) Coot--Model-building tools for molecular graphics. Acta Crystallogr D 60: 2126-2132

Esen A (1993) β-Glucosidase overview. In A Esen ed, β-Glucosidases: Biochemistry and Molecular Biology, ACS Symposium Series 533. American Chemical Society, Washington, D.C., pp 1-14
Hayashi Y, Okino N, Kakuta Y, Shikanai T, Tani M, Narimatsu H, Ito M (2007) Klotho-related protein is a novel cytosolic neutral beta-glycosylceramidase. J Biol Chem 282: 30889-30900

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

Henrissat B, Callebaut I, Fabrega S, Lehn P, Mormon JP, Davies G (1995) Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. Proc Natl Acad Sci USA 92: 7090–7094

Husebye H, Arzt S, Burmeister, WP, Haertel, FV, Brandt A, Rossiter JT, Bones AM (2005) Crystal structure at 1.1 Å resolution of an insect myrosinase from Brevicoryne brassicae shows its close relationship to beta-glucosidases. Insect Biochem Mol Biol 35: 1311-1320

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

Kapust RB, Tözsér J, Fox JD, Anderson, DE, Cherry S, Copeland TD, Waugh DS (2001) Tobacco etch virus protease: Mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Prot Eng 14: 993-1000

Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D 60: 2256-2268

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) Procheck: A program to check the stereochemical quality of protein structures. J Appl Crystallogr 26: 283-291

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 277: 680–685
Lee KH, Piao HL, Kim HY, Choi SM, Jiang F, Hartung W, Hwang I, Kwak JM, Lee IJ (2006) Activation of glucosidase via stress-induced polymerization rapidly increases active pools of abscisic acid. Cell 126: 1109–1120

McPherson A, Koszelak S, Axelrod H, Day J, Williams R, Robinson L, McGrath M, Cascio D (1986) An experiment regarding crystallization of soluble proteins in the presence of β-octyl-glucoside. J Biol Chem 261: 1969-1975

Mo B, Bewley JD (2002) β-Mannosidase (EC 3.2.1.25) activity during and following germination of tomato (Lycopersicon esculentum Mill.) seeds, Purification, cloning and characterization. Planta 215: 141-152

Mizutani FM, Nakanishi H, Ema J, Ma S-J, Noguchi N, Ochiai M-I, Mizutani F, Nakao M, Sakata K (2002) Cloning of β-primeverosidase from tea leaves, a key enzyme in tea aroma. Plant Physiol 130: 2164–2176

Morant AV, Jørgensen K, Jørgensen C, Paquette SM, Sánchéz-Perez R, Møller BL, Bak S (2008) β-Glucosidases as detonators of plant chemical defense. Phytochemistry 69: 1795-1813

Murdushov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum likelihood method. Acta Crystallogr D 53: 240-255

Nijikken Y, Tsukada T, Igarashi K., Samejima M, Wakagi T, Shoun H, Fushinobu S (2007) Crystal structure of intracellular family 1 β-glucosidase BGL1A from the basidiomycete Phanerochaete chrysosporium. FEBS Lett 581: 1514-1520

Opasiri R, Ketudat Cairns JR, Akiyama T, Wara-Aswapati O, Svasti J, Esen A (2003) Characterization of a rice β-glucosidases highly expressed in flower and germinating shoot. Plant Sci 165: 627-638.

Opasiri R, Hua Y, Wara-Aswapati O, Akiyama T, Svasti J, Esen A, Ketudat Cairns JR (2004) β-Glucosidase, exo-β-glucanase and pyridoxine transglucosylase activities of rice BGlul. Biochem J 379: 125-131
Opassiri R, Pomthong B, Onkoksoong T, Akiyama T, Esen A, Ketudat Cairns JR (2006) Analysis of rice glycosyl hydrolase family 1 and expression of Os4bglu12 β-glucosidase. BMC Plant Biol 6: 33

Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. In: CW Carter and RM Sweet, Eds., Meth Enzymol, Vol 276. Academic Press, New York, pp 307–326

Page RD (1996) TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12: 357–358

Raychaudhuri A, Tipton PA (2002) Cloning and expression of the gene for soybean hydroxyisourate hydrolase. Localization and implications for function and mechanism. Plant Physiol 130: 2061–2068

Reuveni M, Sagi Z, Evnor D, Hetzroni A (1999) β-Glucosidase activity is involved in scent production in Narcissus flowers. Plant Sci 147: 19-24

Sanz-Aparicio J, Hermoso JA, Martínez-Ripoll M, Lequerica JL, Polaina J (1998) Crystal structure of beta-glucosidase A from Bacillus polymyxa: insights into the catalytic activity in family 1 glycosyl hydrolases. J Mol Biol 275: 491-502

Srisomsap C, Svasti J, Surarit R, Champattanachai V, Sawangareetrakul P, Boonpuan K, Suphasitanont P, Chokchaichamnankit D (1996) Isolation and characterization of an enzyme with β-glucosidase and β-fucosidase activities from Dalbergia cochinchinensis Pierre. J Biochem 119: 585-590

Sue M, Yamazaki K, Yajima S, Nomura T, Matsukawa T, Iwamura H, Miyamoto T (2006) Molecular and structural characterization of hexameric β-D-glucosidases in wheat and rye. Plant Physiol 141: 1237-1247

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25: 4878-4882

Vagin A, Teplyakov A (1997) MOLREP: an automated program for molecular replacement. J Appl Cryst 30: 1022-1025
Verdoucq L, Czjzek M, Moriniere J, Bevan DR, Esen A (2003) Mutational and structural analysis of aglycone specificity in maize and sorghum β-glucosidases. J Biol Chem 278: 25055- 25062

Verdoucq L, Morinière J, Bevan DR, Esen A, Vasella A, Henrissat B, Czjzek M (2004) Structural determinants of substrate specificity in family 1 β-glucosidases: Novel insights from the crystal structure of sorghum dhurrinase-1, a plant β-glucosidase with strict specificity, in complex with its natural substrate. J Biol Chem 279: 31796-31803

Wiesmann C, Beste G, Hengstenberg W, Schulz GE (1995) The three-dimensional structure of 6-phospho-beta-galactosidase from Lactococcus lactis. Structure 3: 961-8

Xu Z, Escamilla-Treviño LL, Zeng L, Lalgondar M, Bevan DR, Winkel BSJ, Mohamed A, Cheng C, Shih M, Poulton JE, Esen A (2004) Functional genomic analysis of Arabidopsis thaliana glycoside hydrolase family 1. Plant Mol Biol 55: 343–367
**Figure legends**

**Figure 1.** Simplified phylogenetic tree of the amino acid sequences of eukaryotic GH1 proteins with known structures and those of rice and Arabidopsis GH1 gene products. The protein sequences of the eukaryotic proteins with known structures are marked with 4-letter PDB codes for one of their structures, including *Trifolium repens* cyanogenic β-glucosidase (1CBG, Barrett et al., 1995), *Sinapsis alba* myrosinase (1MYR, Burmeister et al., 1997), *Zea mays* ZmGlu1 β-glucosidase (1E1F, Czjzek et al., 2000), *Sorghum bicolor* Dhr1 dhurrinase (1V02, Verdoucq et al., 2004), *Triticum aestivum* β-glucosidase (2DGA, Sue et al., 2006), *Rauvolfia serpentina* strictosidine β-glucosidase (2JF6, Barleben et al., 2007), and *Oryza sativa* Os3BGlul7 (BGlu1) β-glucosidase (2RGL, Cheunchor et al., 2008) from plants, along with *Brevicoryne brassicae* myrosinase (1WCG, Husebye et al., 2005), *Homo sapiens* cytoplasmic (Klotho) β-glucosidase (2E9M, Hayashi et al., 2007), and *Phanerochaete chrysosporium* (2E3Z, Nijekkin et al., 2007), while those encoded in the *Arabidopsis thaliana* and *Oryza sativa* genomes are labeled with the systematic names given by Xu et al. (2004) and Opasiri et al. (2006), respectively. One or two example proteins from each plant are given for each of the eight clusters of genes shared by Arabidopsis (At) and rice (Os), and the Arabidopsis-specific clusters At I (β-glucosidases) and At II (myrosinases), with the number of Arabidopsis or rice enzymes in each cluster given in parentheses. These sequences were aligned with all the At and Os sequences in Clustalx (Thompson et al., 1997), the alignment manually edited, all but representative sequences were removed, and the tree was calculated by neighbor joining method and bootstrapped with 1000 trials, then drawn with TreeView (Page, 1996). The grass plastid β-glucosidases, which are not represented in Arabidopsis and rice, are marked in the group marked “Plastid.” Percent bootstrap reproducibility values are shown on internal branches where they are greater than 60%. Except those marked by asterisks, all external branches represent groups with 100% bootstrap reproducibility. *To avoid excess complexity, those groups of sequences marked with an asterisk are not monophyletic and represent more branches within the designated cluster than are shown. For a complete phylogenetic analysis of Arabidopsis and rice GH1 proteins, see Opasiri et al. (2006).
**Figure 2.** Overall and active site structure of Os3BGl6. A, Overall ribbon diagram of native Os3BGl6. The catalytic residues Glu394 and Glu178 and one molecule of Tris in two conformations are shown in ball and stick representation. The α-helices are colored purple, β-strands green and loops cyan. B, Surface view of the Os3BGl6 structure showing the active site cleft with the ligand $n$-octyl-$\beta$-D-thioglucopyranoside. The surface is colored by electrostatic potential, with positively charged, negatively charged and neutral regions colored blue, red and white respectively. C, Close-up of the electron density of Tris in the active site in stereo view. The Fo-Fc omit electron density map for the Tris is shown as a mesh contoured at $I/\sigma = 3$. The sidechains of the surrounding amino acids are represented by sticks, and the Tris ligand is represented by balls and sticks. In all frames, oxygen atoms are shown in red, nitrogen in blue, sulfur in dark yellow, protein carbons in yellow, Tris carbons in green or pink and $n$-octyl-$\beta$-D-thioglucopyranoside carbons in green.

**Figure 3.** Binding of 2-fluoroglucoside and $n$-octyl-$\beta$-D-thioglucopyranoside in the Os3BGl6 active site. A, Stereoview of the superimposition of active site residues at the -1 subsite of the native Os3BGl6 and Os3BGl6/G2F complex structures. The residues surrounding the -1 subsite are represented by sticks colored with carbons in blue for native Os3BGl6 and in yellow for Os3BGl6/G2F. The G2F moiety bound to the catalytic nucleophile, Glu394, is represented by balls and sticks with carbon in pink. B, Active site of the Os3BGl6/G2F complex showing the electron density omit map of G2F and protein-ligand binding interactions. Hydrogen bonding interactions between G2F and amino acid residues are shown as black dashed lines, while the Fo-Fc omit map of G2F contoured at $I/\sigma = 3.0$ is shown as blue mesh. C, Glycone and aglycone interactions of $n$-octyl-$\beta$-D-thioglucopyranoside with Os3BGl6 residues. Representation of the protein and ligand are as in B, with the Fo-Fc omit map for the ligand contoured at $I/\sigma = 2.5$. D, Two views of the superimposition of the G2F and $n$-octyl-$\beta$-D-thioglucopyranoside ligands, showing the glucose ring is in nearly the same position, but is shaped as a relaxed $^4C_1$ chair in G2F (green carbons) and a $^1S_3$ skew boat in $n$-octyl-$\beta$-D-thioglucopyranoside (pink carbons). In all frames, oxygen is in red, nitrogen in blue, fluorine in cyan, and sulfur in yellow.

**Figure 4.** Comparison of the Os3BGl6/$n$-octyl-$\beta$-D-thioglucopyranoside complex active site surface with those of maize ZmGlu1 and rice Os3BGl7 substrate
complexes. A, Active site the Os3BGlU6/n-octyl-β-d-thioglu copyranoside complex structure. The extension of Met251 to restrict the active site can be seen in the lower left of the picture. B, The aglycone binding pocket of the inactive ZmGlu1 E191D mutant/DIMBOA-Glc (PDB: 1E56) complex structure. The glucosyl moiety is hidden behind the aglycone. C, Active site of Os3BGlu7 with a docked cel lotriose ligand. The smaller Asn245 (in place of Met251 in Os3BGlu6) provides a more open cleft for extended oligosaccharide binding. In all frames, active site amino acid residues are shown as sticks behind the grey transparent surface. The ligands are shown in ball and stick representations. In the online version, protein carbons are shown in green, n-octyl-β-D-thioglu copyranoside carbons in pink, DIMBOA-Glc carbons in blue, cel lotriose carbons in yellow, oxygen in red, nitrogen in dark blue, and sulfur in yellow.

**Supplementary Figure 1**: SDS PAGE analysis of Os3BGlU6 purification. M. Bio-Rad low molecular weight markers; lane 1, crude protein extract of Origami(DE3) cells after Os3BGlU6 expression; lane 2, fusion protein from initial IMAC purification; lane 3, Os3BGlU6 protein after TEV protease digest; lane 4, wash fractions after 2nd IMAC, showing >95% pure Os3BGlU6 protein. Approximately 5 μg of protein was loaded in each lane.
**Table I:** Kinetic parameters of Os3BGlu6 for the hydrolysis of pNP-glycosides, oligosaccharides, alkyl and natural glucosides

| SUBSTRATE                  | nmole pNP/min/mg protein | Percent Relative Activity | $K_m$ (mM) | $k_{cat}$ (s$^{-1}$) | $k_{cat}/K_m$ (mM$^{-1}$s$^{-1}$) |
|----------------------------|--------------------------|--------------------------|------------|----------------------|----------------------------------|
| pNP-β-D-glucoside          | 5844 ± 1.5               | 100                      | 6.3 ± 0.4  | 38.9 ± 0.9           | 6.2                              |
| pNP-β-D-fucoside           | 24,470 ± 1.5             | 415 ± 0.03               | 0.50 ± 0.02| 33.5 ± 0.7           | 67                               |
| pNP-β-D-galactoside        | 1719 ± 1.4               | 29.4 ± 0.02              | 6.06 ± 0.5 | 9.8 ± 0.2            | 1.6                              |
| pNP-α-L-arabinoside        | 51.9 ± 0.3               | 0.88 ± 0.005             | nd$^a$     | nd$^a$               | nd$^a$                           |
| pNP-β-D-cellobioside       | 27.67 ± 0.15             | 0.46 ± 0.003             | nd$^a$     | nd$^a$               | nd$^a$                           |
| pNP-β-D-xylloside          | 6.3 ± 0.5                | 0.10 ± 0.008             | nd$^a$     | nd$^a$               | nd$^a$                           |
| pNP-β-D-mannoside          | 3.8 ± 0.5                | 0.06 ± 0.008             | nd$^a$     | nd$^a$               | nd$^a$                           |
| pNP-β-D-glucuronide        | 2.7 ± 0.3                | 0.04 ± 0.005             | nd$^a$     | nd$^a$               | nd$^a$                           |
| pNP-α-D-glucoside          | 0.43 ± 0.008             | 0.007 ± 0.001            | nd$^a$     | nd$^a$               | nd$^a$                           |

$^a$nd indicates “not determined”
Table II: Hydrolysis of glycosides and oligosaccharides by Os3BGlul6

| SUBSTRATE                        | nmole Glc/min/mg protein \(^a\) | Percent Relative Activity \(^a\) | \(K_m\) (mM) | \(k_{cat}\) (s\(^{-1}\)) | \(k_{cat}/K_m\) (mM\(^{-1}\)s\(^{-1}\)) |
|----------------------------------|---------------------------------|---------------------------------|--------------|-----------------|------------------|
| \(n\)-octyl-\(\beta\)-D-glucoside| 2282±0.9                        | 100±0.04                        | 4.5 ± 0.3    | 12.4 ± 0.4      | 2.7               |
| \(n\)-heptyl-\(\beta\)-D-glucoside| 673.8±0.9                       | 29.5±0.04                       | 5.0 ± 0.3    | 4.25 ± 0.10     | 0.85              |
| **Glucooligosaccharides**        |                                 |                                 |              |                 |                  |
| \(\beta\)-(1→3)-linked           |                                 |                                 |              |                 |                  |
| Laminaribiose,                  | 1272 ± 1.3                      | 55.7±0.05                       | 3.6 ± 0.3    | 6.17 ± 0.19     | 1.7               |
| Laminaritriose                  | 169.6±0.9                       | 7.43±0.04                       | 8.7 ± 0.7    | 1.64 ± 0.06     | 0.18              |
| Laminaritetraose                | 2.0±0.01                        | 0.08±0.0004                     | nd           | nd              | nd                |
| Laminaripentaose                | 1.4±0.01                        | 0.06±0.001                      | nd           | nd              | nd                |
| \(\beta\)-(1→4)-linked           |                                 |                                 |              |                 |                  |
| Cellobiose                      | 15.0±0.5                        | 0.65±0.001                      | 15.3 ± 1.2   | 0.13 ± 0.001    | 0.0085            |
| Cellotriose                     | 13.1±0.6                        | 0.57±0.03                       | nd           | nd              | nd                |
| Cellotetraose                   | 12.9±0.5                        | 0.56±0.02                       | nd           | nd              | nd                |
| Cellopentaose                   | 8.0±0.3                         | 0.35±0.01                       | nd           | nd              | nd                |
| Cellohexaose                    | 6.38±0.3                        | 0.28±0.01                       | nd           | nd              | nd                |
| Sophorose [\(\beta\)-(1→2)-linked] | 728.7±1.3                     | 31.9±0.05                       | 9.8 ± 0.8    | 9.8 ± 0.4       | 0.96              |
| Gentiobiose [\(\beta\)-(1→6)-linked] | 5.25±0.17                    | 0.23±0.007                      | 14.9 ± 1.1   | 0.17± 0.003     | 0.011             |

\(^a\)Natural \(\beta\)-D-glucosides

| Apigenin-7-glucoside            | 15.3±0.2                        | 0.66±0.009                      | nd           | nd              | nd                |
| Glycitin                        | 4.4±0.07                        | 0.19±0.003                      | nd           | nd              | nd                |
| Diadzin                         | 4.0±0.04                        | 0.17±0.001                      | nd           | nd              | nd                |
| Genistin                        | 2.9±0.02                        | 0.13±0.001                      | nd           | nd              | nd                |
| Esculin                         | 2.4±0.01                        | 0.10±0.005                      | nd           | nd              | nd                |
| Arbutin                         | 0.59±0.05                       | 0.02±0.002                      | nd           | nd              | nd                |
| Coumaryl alcohol \(\beta\)-D-glucoside | 0.54±0.01                   | 0.02±0.0005                     | nd           | nd              | nd                |
| Coniferin                       | 0.46±0.007                      | 0.02±0.0003                     | nd           | nd              | nd                |
| Salicin                         | 0.18±0.001                      | (79±0.3)x10\(^{-4}\)            | nd           | nd              | nd                |
aActivities were determined by the glucose oxidase (PGO) assay and percent relative rates are set relative to hydrolysis of \( n\)-octyl-\( \beta\)-D-glucoside as 100% Amounts released from disaccharides were divided by two, since two glucose molecules are released from one reaction. nd- not determined

bThe natural glycosides for which relative activities were determined are those for which hydrolysis was detected on TLC. Amygdalin, dalcochinin-glucoside, gossypin, linamarin, naringin, phlorizin, prunasin, pyridoxine-5'-O-\( \beta\)-D-glucoside, and quercitin 3-\( \beta\)-D-glucoside were not hydrolyzed, as observed on TLC.
Table III: Data collection, processing and refinement parameters

| Dataset | native Os3BGluc6 | Os3BGluc6/n-octyl-β-D-thioglucopyranoside complex | Os3BGluc6/G2F complex |
|---------|------------------|--------------------------------------------------|-----------------------|
| PDB code | 3GNO            | 3GNP                       | 3GNR                  |
| Beamline | BL13B1          | BL13B1                       | BL13B1                |
| Wavelength (Å) | 1.00            | 1.00                        | 1.00                   |
| Space group | P2_12_1         | P2_12_1                    | P2_12_1               |
| Unit-cell parameters (Å) | a = 56.8, b = 90.5, c = 101.7 | a = 57.2, b = 91.1, c = 111.4 | a = 56.7, b = 90.6, c = 101.9 |
| Resolution range (Å) | 30.0 - 1.83 | 30.0 - 1.80 | 30.0 - 1.81 |
| Resolution outer shell (Å) | 1.90 - 1.83 | 1.86 - 1.80 | 1.87 - 1.81 |
| Completeness (%) | 98.9 (96.8)^a | 91.1 (98.0)^a | 92.9 (95.1)^a |
| Average redundancy per shell | 7.0 (6.9) | 6.9 (6.9) | 3.7 (4.0) |
| I/σ(I) | 18.9 (4.9)^a | 30.7 (6.4)^a | 23.9 (8.7)^a |
| R_{merge} (%) | 9.5 (36.6)^a | 5.2 (29.3)^a | 4.5 (18.4)^a |
| R_{factor} (%) | 16.3 | 17.8 | 16.2 |
| R_{free} (%) | 20.4 | 20.7 | 18.5 |
| No. of residues in protein | 1-488 | 1-488 | 1-488 |
| No. waters | 466 | 340 | 452 |
| Mean B-factor | | | |
| Protein | 12.7 | 18.5 | 16.0 |
| Ligand | 28.8 | 41.2 | 14.8 |
| Water | 26.2 | 27.1 | 27.0 |
| r.m.s. bond deviations (Å) | 0.013 | 0.015 | 0.012 |
| r.m.s angle deviations (degrees) | 1.32 | 1.44 | 1.28 |
| Ramachandran plot | | | |
| Residues in most favorable regions (%) | 89.6 | 89.3 | 89.8 |
| Residues in additional allowed regions (%) | 10 | 10 | 9.5 |
| Residues in generously allowed regions (%) | 0.5 | 0.7 | 0.7 |

^a Numbers in parentheses designate to the highest resolution shell.
Figure 1. Simplified phylogenetic tree of the amino acid sequences of eukaryotic GH1 proteins with known structures and those of rice and Arabidopsis GH1 gene products. The protein sequences of the eukaryotic proteins with known structures are marked with 4-letter PDB codes for one of their structures, including *Trifolium repens* cyanogenic β-glucosidase (1CBG, Barrett et al., 1995), *Sinapsis alba* myrosinase (1MYR, Burmeister et al., 1997), *Zea mays* ZmGlu1 β-glucosidase (1E1F, Czjzek et al., 2000), *Sorghum bicolor* Dhr1 dhurrinase (1V02, Verdoucq et al., 2004), *Triticum aestivum* β-glucosidase (2DGA, Sue et al., 2006), *Rauvolfia serpentina* strictosidine β-glucosidase (2JF6, Barleben et al., 2007), and *Oryza sativa* Os3BGlu7 (BGlu1) β-glucosidase (2RGL, Cheunchor et al., 2008) from plants, along with *Brevicoryne brassicae* cytoplasmic (Klotho) β-glucosidase (2E9M, Hayashi et al., 2007), and *Phanerochaete chrysosporium* (2E3Z, Nijekkin et al., 2007), while those encoded in the *Arabidopsis thaliana* and *Oryza sativa* genomes are labeled with the systematic names given by Xu et al. (2004) and Opassiri et al. (2006), respectively. One or two example proteins from each plant are given for each of the eight clusters of genes shared by Arabidopsis (At) and rice (Os), and the Arabidopsis-specific clusters At I (β-glucosidases) and At II (myrosinases), with the number of Arabidopsis or rice enzymes in each cluster given in parentheses. These sequences were aligned with all the At and Os sequences in Clustalx (Thompson et al., 1997), the alignment manually edited, all but representative sequences were removed, and the tree was calculated by neighbor joining method and bootstrapped with 1000 trials, then drawn with TreeView (Page, 1996). The grass plastid β-glucosidases, which are not represented in Arabidopsis and rice, are marked in the group marked “Plastid.” Percent bootstrap reproducibility values are shown on internal branches where they are greater than 60%. Except those marked by asterisks, all external branches represent groups with 100% bootstrap reproducibility. *To avoid excess complexity, those groups of sequences marked with an asterisk are not monophyletic and represent more branches within the designated cluster than are shown. For a complete phylogenetic analysis of Arabidopsis and rice GH1 proteins, see Opassiri et al. (2006).
Figure 2. Overall and active site structure of Os3BGlul6. A, Overall ribbon diagram of native Os3BGlul6. The catalytic residues Glu394 and Glu178 and one molecule of Tris in two conformations are shown in ball and stick representation. The α-helices are colored purple, β-strands green and loops cyan. B, Surface view of the Os3BGlul6 structure showing the active site cleft with the ligand n-octyl-β-D-thioglucopyranoside. The surface is colored by electrostatic potential in the online version, with positively charged, negatively charged and neutral regions colored blue, red and white respectively. C, Close-up of the electron density of Tris in the active site in stereo view. The Fo-Fc omit electron density map for the Tris is shown as a mesh contoured at I/σ = 3. The sidechains of the surrounding amino acids are represented by sticks, and the Tris ligand is represented by balls and sticks. In all frames, oxygen atoms are shown in red, nitrogen in blue, sulfur in dark yellow, protein carbons in yellow, Tris carbons in green or pink and n-octyl-β-D-thioglucopyranoside carbons in green.
Figure 3. Binding of 2-fluoroglucoside and \( n \)-octyl-\( \beta \)-D-thioglucopyranoside in the Os3BGlu6 active site. 
A, Stereoview of the superimposition of active site residues at the -1 subsite of the native Os3BGlu6 and Os3BGlu6/G2F complex structures. The residues surrounding the -1 subsite are represented by sticks colored with carbons in blue for native Os3BGlu6 and in yellow for Os3BGlu6/G2F. The G2F moiety bound to the catalytic nucleophile, Glu394, is represented by balls and sticks with carbon in pink. B, Active site of Os3BGlu6/G2F complex showing the electron density omit map of G2F and protein-ligand binding interactions. Hydrogen bonding interactions between G2F and amino acid residues are shown as black dashed lines, while the Fo-Fc omit map of G2F contoured at \( I/\sigma = 3.0 \) is shown as blue mesh. C, Glycone and aglycone interactions of \( n \)-octyl-\( \beta \)-D-thioglucopyranoside with Os3BGlu6 residues. Representation of the protein and ligand are as in B, with the Fo-Fc omit map for the ligand contoured at \( I/\sigma = 2.5 \). D, Two views of the superimposition of the G2F and \( n \)-octyl-\( \beta \)-D-thioglucopyranoside ligands, showing the glucose ring is in nearly the same position, but is shaped as a relaxed \( 4C_1 \) chair in G2F (green carbons) and a \( 1S_3 \) skew boat in \( n \)-octyl-\( \beta \)-D-thioglucopyranoside (pink carbons). In all frames, oxygen is in red, nitrogen in blue, fluorine in cyan, and sulfur in yellow.
Figure 4. Comparison of the Os3BGlu6/n-octyl-β-D-thioglucopyranoside complex active site surface with those of maize ZmGlu1 and rice Os3BGlu7 substrate complexes. A, Active site of the Os3BGlu6/n-octyl-β-D-thioglucopyranoside complex structure. The extension of Met251 to restrict the active site can be seen in the lower left of the picture. B, The aglycone binding pocket of the inactive ZmGlu1 E191D mutant/DIMBOA-Glc (PDB: 1E56) complex structure. The glucosyl moiety is hidden behind the aglycone. C, Active site of Os3BGlu7 with a docked cellotriose ligand. The smaller Asn245 (in place of Met251 in Os3BGlu6) provides a more open cleft for extended oligosaccharide binding. In all frames, active site amino acid residues are shown as sticks behind the grey transparent surface. The ligands are shown in ball and stick representations. In the online version, protein carbons are shown in green, n-octyl-β-D-thioglucopyranoside carbons in pink and DIMBOAGlc carbons in blue, and cellotriose carbons in yellow, oxygen in red, nitrogen in dark blue, and sulfur in yellow.