

Thailand Research Fund Contract Number BRG5080007
Structure Function Relationships in Plant Beta-Glucosidases
Final Report

Grant Period: 1 September, 2007- 31 October, 2010.

Project Title: Structure Function Relationships in Plant Beta-Glucosidases II

(Thai): ความสัมพันธ์ระหว่างโครงสร้างและหน้าที่ของเอนไซม์เบต้ากลูโคซิเดสในพืช II

Head of Project: Dr. James R. Ketudat-Cairns

Institution: Institute of Science, Schools of Biochemistry and Chemistry,
Suranaree University of Technology
Nakhon Ratchasima 30000, THAILAND

Phone: 044 224304

Fax: 044 224185

Email: cairns@sut.ac.th, jrkcairns@yahoo.com

Coinvestigator: Dr. Rodjana Opasiri Institute of Science,
Suranaree University of Technology
Nakhon Ratchasima 30000

Advisor: Prof. Dr. M.R. Jisnuson Svasti Centre of Excellence in Protein Structure
and Function & Dept. of Biochemistry
Faculty of Science
Mahidol University
Phayathai, Bangkok 10400

Abstract

Beta-glucosidases and exoglucanases play many important roles in plants including cell wall recycling, which points to their potential use in biomass conversion. In order to make the best use of these enzymes, we need to understand the way they bind to oligosaccharides, as well as the determinants of substrate specificity, which drive them to release a β -linked glucose or similar monosaccharides from oligosaccharides. In this study, we addressed these issues by recombinant expression, functional and structural studies of glycoside hydrolase family 1 (GH1) and 3 (GH3) enzymes. For the GH1 enzymes, we first addressed the structural basis for glucooligosaccharide binding and hydrolysis by solving the X-ray crystal structures of rice Os3BGlu7, its covalent intermediate with 2-fluoroglucoside, and its semiactive catalytic acid/base mutant, E176Q, alone and in complexes with 2-fluoroglucoside, cellotetraose, cellopentaose, and laminaribiose. The cellotetraose and cellopentaose structures showed that, although the nonreducing glucosyl residue is completely bounded by direct bonds to the protein, all the other residues are primarily bound by water-mediated hydrogen bonds and stacking on aromatic residues. Since the Os3BGlu7 E176Q mutant remained active on *p*-nitrophenyl- β -D-glucoside in the presence of small nucleophiles, the energies of binding of the different glucosyl residues could also be estimated from the competitive inhibition constants of the cellooligosaccharides. Mutations of the residues along the active-site cleft also verified their involvement in oligosaccharide binding.

In order to address the determinants of substrate specificity between closely related β -glucosidases and β -mannosidases, the GH1 phylogenetic cluster including Os3BGlu7 β -glucosidase and barley HvBII β -mannosidase was studied by comparison of these enzymes and expression of two more rice isoenzymes from this cluster. Os3BGlu8, which is most similar to Os3BGlu7 was found to prefer β -glucosides, while Os7BGlu26, which is most similar to HvBII was found to prefer β -mannosides, as might be expected. No mutation of a single residue near the active site was able to convert a β -glucosidase to a β -mannosidase or a β -mannosidase to a β -glucosidase, although some shifted the ratio of activities in that direction. Since mannosides and glucosides are thought to transverse different shaped transition states and undergo different conformational pathways to do so, with binding of glucoside, thioglucoside, mannoside, and thiomannoside to Os3BGlu7 and HvBII were compared by inhibition studies, saturation transfer difference NMR, and QM/MM calculations. It was found that the thioglycosides behaved differently from the O-linked glycosides, making them poor models for such interactions, and all glycosides could to bind to the active sites in multiple conformations. Solving of the structure of Os7BGlu26 β -mannosidase and Os3BGlu7 β -glucosidase in complexes with mannosides is underway to gain more insight into this issue.

GH3 serves as another model of β -glucosidase activity, acting on both oligosaccharides and polysaccharides and barley ExoI has been a structural model for this family, but was not previously expressed in recombinant systems. Our development of a recombinant expression system, and purification and crystallization methods for the recombinant ExoI enzyme, rHvExoI, allowed us to investigate the effects of mutating each active site amino acid residue at both the functional and structural level. So far, we showed that mutation of nearly every amino acid surrounding the active site gave negligible to low levels of activity, but two, E220A and W434A had enough activity to characterize and learn about their specific functions.

Together, these studies helped train seven graduate students and produced seven international publications, in addition to basic knowledge for future applications.

บทคัดย่อ

ปีตากลูโคซิเดสและเอกโซกลูแคนเนสมีบทบาทสำคัญในพืชอันได้แก่กระบวนการสลายและสร้างผนังเซลล์ ซึ่งชี้ให้เห็นถึงการโอกาสที่จะใช้เอนไซม์ในการแปลงพลังงานชีวมวล การที่จะนำเอนไซม์เหล่านี้มาใช้ให้ดีที่สุด ต้องเข้าใจวิธีการจับของเอนไซม์กับ oligosaccharides รวมทั้งตัวกำหนดความจำเพาะต่อการย่อยสลายสเตรต ที่กระตุ้นให้เอนไซม์ปลดปล่อยปีตากลูโคสหรือน้ำตาลโมเลกุลเดี่ยวที่คล้ายกันจาก oligosaccharides ในการวิจัยนี้จึงได้ทำการผลิตโปรตีนสายผสม ศึกษาหน้าที่และโครงสร้างของเอนไซม์ตระกูล glycoside hydrolase family 1 (GH1) และ 3 (GH3) สำหรับเอนไซม์ตระกูล GH1 เริ่มจากการศึกษาโครงสร้างพื้นฐาน ในการจับและย่อยสลาย glucooligosaccharide ด้วยการวิเคราะห์โครงสร้างของผลึกเอนไซม์ของเอนไซม์ Os3BGlu7 ของข้าว โครงสร้างของเอนไซม์ในสภาวะเปลี่ยนขณะที่จับกับ 2-fluoroglucoside ด้วยพันธะโควาเลนต์ และโครงสร้างของเอนไซม์กลายพันธุ์ตรงตำแหน่งหมู่เร่ง acid/base คือ E176Q ซึ่งยังสามารถเร่งปฏิกิริยาได้บ้าง ทั้งในสภาพเอนไซม์อิสระและที่จับกับ 2-fluoroglucoside cellotetraose cellopentaose และ laminaribiose จากโครงสร้างของ cellotetraose และ cellopentaose แสดงให้เห็นว่าแม้ nonreducing glucosyl residue จะจับอย่างสมบูรณ์ด้วยการสร้างพันธะโดยตรงกับโปรตีน แต่หน่วยกลูโคสอื่นๆ จับกับกรดอะมิโนชนิด aromatic ด้วยพันธะไฮโดรเจนโดยมีน้ำเป็นตัวกลาง เนื่องจากเอนไซม์กลายพันธุ์ Os3BGlu7 E176Q ยังคงย่อย *p*-nitrophenyl- β -D-glucoside ในสภาวะที่มีโมเลกุลขนาดเล็กที่มีประจุลบ ทำให้สามารถคำนวณค่าพลังงานยึดเหนี่ยวของโมเลกุลกลูโคสแต่ละหน่วยกับเอนไซม์ จากค่าคงที่การยับยั้งแบบแข่งขันของ celooligosaccharides การเปลี่ยนกรดอะมิโนตรงของตำแหน่งเร่งปฏิกิริยาของเอนไซม์ได้พิสูจน์ว่ากรดอะมิโนเหล่านี้มีส่วนร่วมในการจับ oligosaccharide อีกด้วย

เพื่อที่จะทราบถึงตัวกำหนดความจำเพาะต่อการย่อยสลายสเตรตระหว่างปีตากลูโคซิเดสและปีตาแมนโนซิเดสที่มีความสัมพันธ์ใกล้ชิดกัน จึงได้นำเอนไซม์ตระกูล GH1 ที่อยู่บนแผนภูมิความสัมพันธ์ในเชิงวิวัฒนาการใกล้เคียงกัน มาศึกษาเปรียบเทียบ ได้แก่ ปีตากลูโคซิเดส Os3BGlu7 และปีตาแมนโนซิเดส rHvBII ของบาร์เลย์ และได้ผลิตเอนไซม์ข้าวที่อยู่ในกลุ่มนี้อีกสองชนิดคือ Os3BGlu8 และ Os7BGlu26 พบว่า Os3BGlu8 ที่มีความคล้ายคลึงกับ Os3BGlu7 มากที่สุดสามารถย่อย β -glucosides ได้ดี ในขณะที่ Os7BGlu26 ที่มีความคล้ายคลึงกับ HvExoI มากที่สุดพบว่าสามารถย่อย β -mannosides ได้ดี เป็นไปตามที่คาดหมายไว้ อย่างไรก็ตามการเปลี่ยนแปลงกรดอะมิโนเพียงหนึ่งตำแหน่งที่อยู่ใกล้กับตำแหน่งเร่งปฏิกิริยาไม่สามารถเปลี่ยนปีตากลูโคซิเดสให้เป็นปีตาแมนโนซิเดสหรือเปลี่ยนปีตาแมนโนซิเดสให้เป็นปีตากลูโคซิเดสแม้ว่าจะมีการเปลี่ยนแปลงอัตราส่วนของการทำงานบ้าง จากความคิดที่ว่าสภาพเปลี่ยนของ mannosides และ glucosides มีรูปร่างที่ไม่เหมือนกันและมีวิธีการจัดเปลี่ยนโครงสร้างที่แตกต่างกันด้วย จึงได้มีการเปรียบเทียบการจับของ glucoside thioglucoside mannoside และ thiomannoside กับ Os3BGlu7 และ HvBII โดยการศึกษาการยับยั้ง ความแตกต่างของ saturation transfer ด้วย NMR และการคำนวณ QM/MM พบว่า thioglycosides แสดงลักษณะที่แตกต่างจาก O-linked glycosides จึงได้แบบจำลองปฏิสัมพันธ์ที่ดี ส่วน glycosides ทุกชนิดสามารถจับที่ตำแหน่งเร่งปฏิกิริยาด้วยโครงสร้างที่หลากหลาย การวิเคราะห์โครงสร้างของแมนโนซิเดส Os7BGlu26 และกลูโคซิเดส Os3BGlu7 ที่จับเชิงซ้อนกับ mannosides ยังคงอยู่ระหว่างการดำเนินงานเพื่อทำให้เกิดความเข้าใจมากยิ่งขึ้น

GH3 เป็นอีกหนึ่งแบบจำลองของการทำงานของปีตากลูโคซิเดสที่ย่อยทั้ง oligosaccharides และ polysaccharides ได้มีการใช้เอนไซม์ ExoI ของบาร์เลย์เป็นโครงสร้างแม่แบบสำหรับเอนไซม์ในตระกูลนี้ อย่างไรก็ตามเอนไซม์นี้ไม่ได้มีการผลิตมาจากกระบวนโปรตีนสายผสมมาก่อน งานวิจัยนี้จึงทำการพัฒนาระบบการผลิตโปรตีนสายผสม วิธีการแยกโปรตีนให้บริสุทธิ์และตกผลึกสำหรับเอนไซม์สายผสม ExoI (rHvExoI) ทำให้สามารถศึกษาผลจากการเปลี่ยนชนิดของกรดอะมิโนแต่ละตัวในตำแหน่งเร่งปฏิกิริยา ต่อการทำงานและระดับการทำงานของเอนไซม์ ถึงขั้นตอนนี้พบว่าการเปลี่ยนชนิดของกรดอะมิโนเกือบทุกตัวที่ล้อมรอบตำแหน่งเร่งปฏิกิริยานั้นทำให้การทำงานของเอนไซม์มีระดับต่ำลง แต่การทำงานของ E220A และ W434A มีมากพอที่จะศึกษาคุณสมบัติและเรียนรู้เกี่ยวกับหน้าที่ที่แน่นอนของกรดอะมิโนสองชนิดนี้ได้

Executive Summary

Importance: β -Glucosidases and exoglucanases are important enzymes with many functions in plants and other organisms, and also play an important role in lignocellulosic biomass conversion. Understanding the mechanisms by which they bind to oligosaccharides and select their substrates is critical to identifying the functions of proteins identified from genomic sequencing and engineering enzymes for better application.

Objectives: The objectives of this work included:

- 1) To determine the structure of rice Os3BGlu7 alone and in complex with substrates and mechanism based inhibitors, in order to understand how it interacts with sugars.
- 2) To elucidate the mechanism by which closely related family GH1 glycoside hydrolases hydrolyze both β -D-glucosides and β -D-mannosides with different preferences.
- 3) To produce the glycoside hydrolase family GH3 barley β -glucan exohydrolase in a recombinant system to facilitate structure-function studies of active site amino acids.

Methods and Results: The structure of rice Os3BGlu7 β -glucosidase and its complexes were determined by X-ray crystallography. The structures of Os3BGlu7 and its covalent intermediate with 2-fluoroglucoside, a mechanism-based inhibitor, were determined to tell the general shape and sugar conformation in the covalent intermediate. The structures of the Os3BGlu7 acid/base mutant E176Q, which can hydrolyze substrates with easily deprotonated aglycons, like *p*-nitrophenyl- β -D-glucoside (*p*NPGlc), but not oligosaccharides, and its complexes with cellotetraose, cellopentaose and laminaribiose showed that the nonreducing glucosyl residue (glycon) is bound by direct hydrogen bonds, which distort its shape to between a 1S_3 skew boat and a 4E envelope transition-state-like shape. In contrast, other glucosyl residues in the glucooligosaccharides remain in relaxed chairs and are bound by water-mediated hydrogen bonds and interactions with aromatic rings. Comparison of rice Os3BGlu7 and Os3BGlu8 β -glucosidases with barley HvBII and rice Os7BGlu26 β -mannosidases showed that these closely related enzymes have the same amino acids interacting with the glycon, despite their different glycon sugar preferences. Inhibition, saturation transfer difference NMR, and quantum mechanics/molecular mechanics studies showed that thioglucoside and thiomannoside are shaped differently from the O-glycosides and their binding to the active site cannot be directly correlated to those of the O-glycosides, although the S- and O-glucosides seem to bind in similar shapes. The structure of the Os7BGlu26 β -mannosidase was solved and appeared to have a very similar active site to Os3BGlu7 β -glucosidase, although it may be slightly narrower. Further studies with gluco- and manno- configured substrates, mechanism-based inhibitors and transition-state analogues will be necessary to understand the mechanistic differences between these enzymes. The relationship seems to be complex, as no single mutation of a Os3BGlu7 to an HvBII residue or an HvBII to a Os3BGlu7 residue could change the β -glucosidase to the β -mannosidase or the β -mannosidase to a β -glucosidase. The HvExoI β -glucan exohydrolase, which serves as a structural model for plant GH3 enzymes, could be efficiently produced as a secreted enzyme in *Pichia pastoris* and efficiently purified and characterized. The recombinant enzyme and its deglycosylated form had properties very similar to the native protein from barley seedlings, and released glucose from a variety of oligo- and poly- saccharides. Mutagenesis of the active site indicated most residues are critical to the activity and crystallization of the protein facilitates structure function studies.

Output and Benefits: This work provided much information on the structural basis of oligosaccharide hydrolysis by GH1 and GH3 enzymes for future engineering and contributed to 7 international publications and to training 7 graduate students.

BRG5080007

1. Project Name

Thai name: ความสัมพันธ์ระหว่างโครงสร้างและหน้าที่ของเอนไซม์เบต้ากลูโคซิเดสในพืช II

English name: Structure and Function Relationships in Plant β -Glucosidases II

2. Project Participants

2.1.1 Name: James R. Ketudat Cairns

2.1.2 Degree: Ph.D.

2.1.3 Position: Associate Professor of Biochemistry

2.1.4 Work address: Institute of Science, Suranaree University of Technology,
111 University Avenue, Muang District, Nakhon Ratchasima 30000

2.1.5 Duties and responsibilities in project: Head of project, planning and oversight of project

2.2.1 Name: Rodjana Opasiri

2.2.2 Degree: Ph.D

2.2.3 Position: Instructor of Biochemistry

2.2.4 Work address: Institute of Science, Suranaree University of Technology,
111 University Avenue, Muang District, Nakhon Ratchasima 30000

2.2.5 Position and responsibilities: Coinvestigator, protein expression, mutagenesis, enzymology

2.3.1 Prof. Dr. M.R. Jisnuson Svasti

2.3.2 Degree: Ph.D.

2.3.3 Position: Professor

2.3.4 Center for Protein Structure and Function, Dept.
Biochemistry, Faculty of Science, Mahidol University, Rama 6 Road,
Phayathai, Bangkok 10400

2.3.5 Duties and responsibilities: Advisor to project

3. Area of research: Protein Structure and Function (plant enzymes)

4. Problem and Its Importance

The basic problem to be addressed in this line of experiments is the relationship between protein structure and function, which will be explored using plant glycosyl hydrolase family 1 and 3 exo- β -glucanase-like enzymes as a model system. With the sequencing of the genomes of many organisms, including humans, other animals, plants and microorganisms, the genetic plan for living organisms has been revealed. However, the way this information is applied to create all the functions of living organisms depends on how the proteins encoded by this DNA code fold into their structures and interact with other molecules. Discerning the physical interactions involved in these processes is much more complex than determining the DNA sequence or the genetic code by which nucleotide sequence is converted to proteins. Therefore, it requires in-depth study of many proteins to understand these physical relationships. The understanding gained through such study will allow the engineering and design of proteins for specific applications, as well as helping to reveal how living organisms work at the molecular level.

The glycosyl hydrolases, which hydrolyze glycosyl linkages between sugars and other moieties are of particular interest due to their abundance and diversity in living organisms, their high catalytic efficiency, which allows them to increase reaction rates by as much as 10^{17} fold (Rye and Withers, 2000), and their potential for application. Glycosyl hydrolase family 1 and 3 β -glucosidases in plants are of particular interest due to their functional diversity, despite their high amino acid sequence and structural similarity. Therefore, it may be implied that small changes in the amino acid sequence effect changes in enzymatic properties, so that the changes in physical interactions involved may be investigated by protein engineering techniques. It is also of interest to determine what similar interactions may allow the binding of oligosaccharides by enzymes with different structures. Such information might be applied to engineer divergent enzymes that are involved in carbohydrate binding and hydrolysis. In addition, the enzymes may be applied to biomass conversion for fuel production, synthesis of important oligosaccharides and glycosides by transglycosylation, and treatment of plant materials for greater nutritional value, so the engineering may have direct benefits.

In the first 3 year project on this theme, we concentrated on characterization of rice BGlu1 and isoflavonoid β -glucosidases from *Dalbergia* species. In the current project, we focused on cereal exoglucanases and related enzymes from the cluster within glycosyl hydrolase family 1 that are related to rice BGlu1 and those in the β -glucan exohydrolase subfamily of glycosyl hydrolase family 3, since these enzymes are of relevance to recycling of cell wall components for biomass conversion and energy production. In the GH1 BGlu1 cluster, the enzymes are split between β -mannosidases, which hydrolyze β -D-mannosides faster than β -D-glucosides, and β -glucosidases, which hydrolyze β -D-glucosides faster than β -D-mannosides. Linus Pauling (1948) surmised that enzymes catalyze reactions by binding the transition state. Since β -D-glucosides are thought to be hydrolyzed via half-chair-like transition states, while β -D-mannosides are thought to go through a boat-like transition state (Davies et al., 2003), this opens the question as to whether one enzyme can bind two different shaped transition states or whether closely related enzymes bind different transition states to hydrolyze the same reactions. Understanding this issue may

help with engineering of sugar specificity, so we started to collect data on the plant β -mannosidase/ β -glucosidase cluster in this work.

5. Objectives

The purpose of this project is to determine the structure to function relationships in plant exo- β -glycanases by achieving the following objectives:

- 5.1. Determination of the structures of rice BGlu1 (Os3BGlu7) β -glucosidase mutant in complex with its oligosaccharide substrates and analysis of substrate binding interactions.
- 5.2. Structural determination of a barley or rice β -mannosidase/ β -glucosidase and comparison to rice BGlu1 β -glucosidase.
- 5.3. Investigation of the basis for catalytic properties and differential sugar binding and hydrolysis by BGlu1 and closely related β -glycosidases, which form a phylogenetic cluster with both β -glucosidases and β -mannosidases by site-directed mutagenesis.
- 5.4. Recombinant expression of barley Exo-I exo- β -glucanase and characterization of its properties. Mutagenesis of residues around its active site to determine their importance to its activity.
- 5.5. Crystallization of recombinant Exo-I and its mutants.
- 5.6. Cloning of rice Glycosyl Hydrolase family 3 exoglucanases for recombinant expression.

6. Activities and Methods:

6.1. Structural studies of rice BGlu1 β -glucosidase and its binding with oligosaccharides.

The initial structure solution of rice BGlu1 (Os3BGlu7) β -glucosidase was solved in a previous project (Chuenchor et al., 2006), but the final refinement and computational docking with cellotriose was completed in this project. The BGlu1 was produced as a thioredoxin fusion protein in *E. coli* strain Origami(DE3), as previously described (Opassiri et al., 2003) and purified by immobilized metal affinity chromatography (IMAC), enterokinase digest to cleave off the thioredoxin and His6 tags, a second round of IMAC to remove the tag and gel filtration chromatography on a Superdex S200 column (GE Healthcare) in 150 mM NaCl, 20 mM Tris-HCl, pH 8.0, as described elsewhere (Chuenchor et al., 2006, 2008). The protein was then concentrated in a centricon to approximately 10 mg/mL and used for crystallization.

Crystallization was done by optimizing the conditions previously identified and the apoprotein crystal for data collection was obtained by mixing 3 μ L 11.5 mg/mL protein with 1 μ L precipitant containing 23% polyethylene glycol monomethylester (PEG MME) 5000, 0.17 M $(\text{NH}_4)_2\text{SO}_4$, 0.1 M MES, pH 6.7, in hanging-drop vapor diffusion over the reservoir of the precipitant solution. For the complex with 2-deoxy-2-fluoroglucoside (G2F), the protein was mixed in a molar ratio of 1:5 with 2,4-dinitrophenyl β -D-2-deoxy-2-fluoroglucopyranoside, then 2 mL of protein was mixed with 1 mL of precipitant containing 23% PEG MME 5000, 0.20 M $(\text{NH}_4)_2\text{SO}_4$, 0.1 M MES, pH 6.7, in hanging drop vapor diffusion over the precipitant solution. After 2 hours of equilibration, microseeds of a previous incubation were streaked in with a cat whisker. The crystal drops were incubated at 15°C until crystals of adequate size were produced.

The crystals were cryoprotected in a solution of 18% glycerol in 118% precipitant and flash vitrified in liquid nitrogen. X-rays were diffracted by the crystals at beamline B13 of the National Synchrotron Radiation Research Center (NSRRC), Hsinchu, Taiwan, and data was indexed, integrated and scaled with the HKL2000 software (Otwinowski and Minor, 1997) available at the NSRRC. The initial structure was solved with Amore (Navaza, 1994), as previously reported (Chuenchor et al., 2006) and was refined with REFMAC 5.0 (Murshudov et al., 1999), as part of the CCP4 suite. Structures were checked for stereochemical reasonability with PROCHECK (Laskowski et al., 1993). For later structures of BGlu1 with G2F and mutant structures with oligosaccharides, the initial BGlu1 structure without heteroatoms was simply rigid body refined into the data to solve the initial structure, and refinement proceeded essentially as described for the wild type apoenzyme.

After solution of the first two structures, cellotriose was docked into the active site with the AutoDock program (Morris et al., 1998). The structure from the G2F complex with the heteroatoms removed and the cellotriose ligand were prepared with the Sybyl 7.1 program (Tripos, St. Louis, MO, USA) with blocking of the termini and addition of polar hydrogens. Charges were assigned by the Kolman united atom method (Weiner et al., 1984) and the energies were minimized. The shape of the nonreducing end glucosyl residue of the cellotriose was tested as 4C_1 relaxed chair and 1S_3 skew boat, but the skew boat was the one that gave the reasonable docking. Autodock 4.0 was run 100 times with the Lamarckian Genetic Algorithm (Morris et al., 1998) and results were clustered with a cutoff 2.0 Å root mean squared deviation (RMSD).

Mutations of BGlu1 were made to generate catalytically disabled enzymes for producing complexes with native substrates. Site directed mutagenesis was done by either the overlap extension method (Wurch et al., 1998) or the QuikChange Mutagenesis kit (Stratagene, La Jolla, CA, USA), including E176A, E176D, E176Q, E386D and E386Q. These were tested for activity by hydrolysis of *p*-nitrophenol β -D-glucopyranoside (*p*NPGLc), as previously described (Opassiri et al., 2003). Further mutations were made in the oligosaccharide binding cleft by QuikChange mutagenesis, including S334A, Y341A and Y341L.

The E176 mutants were characterized and the E176Q mutant used to determine the binding affinities of oligosaccharides by the inhibition of *p*NPGLc hydrolysis. The pH dependence of the wildtype BGlu1 and the E176 mutants was determined with *p*NPGLc as a substrate in universal pH buffer (0.2 M boric acid, 0.05 M citric acid mixed with 0.05 M tri-sodium phosphate to achieve the pH values). Relative activities of wildtype BGlu1 and BGlu1 E176 mutants were determined in 50 mM buffer: sodium acetate, MES, phosphate or universal buffer, and 50 mM nucleophiles after adjustment of the reaction solution (without enzyme and substrate) to pH 5.0. Kinetic parameters (K_m and apparent k_{cat}) of wildtype BGlu1 and BGlu1 E176 mutants in the presence of 50 mM sodium azide were calculated from nonlinear regression of Michaelis-Menten curves with Grafit 5.0 (Erithicus Software, Surrey, UK). The competitive inhibition constants (K_i) were determined from the plots of the apparent K_m/V_{max} vs. the inhibitor (oligosaccharide) concentration in the presence of 50 mM sodium azide, and were taken to be the apparent dissociation constants (K_D) of the inhibitors. The association constants were calculated as $K_A = 1/K_D$; and Gibbs free energies of binding were calculated as $\Delta G_{binding} = -RT \ln K_A$, where R is the gas constant ($8.314 \text{ J K}^{-1} \text{ mol}^{-1}$) and T is the absolute temperature (303K). Subsite affinities were determined for binding of oligosaccharides to binding cleft mutants as previously described (Opassiri et al., 2004).

6.2. Comparison of the catalytic properties of barley rHvBII, its mutants and closely related rice isoenzymes.

The barley HvBII and rice Os3BGlu8 and Os7BGlu26 cDNA were cloned for expression as recently described (Kuntothom et al., 2009). Briefly, for barley, total mRNA was extracted from 3-day-old germinated barley cv. Clipper seeds with TRIzol® reagent and reverse-transcribed with polyT₁₇ primer and Superscript II reverse transcriptase, according to the manufacturer's instructions (Invitrogen, Carlsbad, CA, USA). The cDNA encoding the full-length barley rHvBII was PCR amplified from the first-strand cDNA pool with *Pfu* DNA polymerase (Promega, Madison, WI, USA) and primers designed based on the BGQ60 cDNA sequence (GenBank accession number L41869; Leah et al., 1995). The cDNA encoding the mature barley β -D-glucosidase, designated rHvBII, was amplified with primers to add an *Nco*I site to the 5' end at the beginning of the code for the mature protein and an *Xho*I site at the *Nco*I site, cloned into pBluescript SK II (Stratagene, La Jolla, CA, USA) at the *Eco*R V site and sequenced (GenBank accession number EU807965). The cDNA encoding rHvBII was subcloned into the pET32a expression vector (Novagen, Madison, WI, USA) at the *Nco*I and *Xho*I sites.

The *Os7BGlu26* cDNA was amplified by RT-PCR from RNA isolated from 7-day-old rice (*O. sativa* L., cv. KDML105) seedlings with (dT)₂₀ primer and the Thermoscript RT-PCR system (Eppendorf, Hamburg, Germany). The full-length *Os7BGlu26* cDNA was amplified with primers designed from the predicted start and stop codons and *Pfu* DNA polymerase. A cDNA encoding the predicted mature *Os7BGlu26* isoenzyme was amplified from the initial PCR product with primers designed from the sequence encoding the mature protein N-terminus with an additional CACC on the 5' end for directional cloning and the stop codon and C-terminal region, cloned into a Gateway entry clone with the pENTR™-D-TOPO® Cloning Kit (Invitrogen), and sequenced. The cDNA encoding the mature *Os7BGlu26* (GenBank accession number EU835514) was recombined into the pET32a/DEST expression vector (Opassiri et al., 2006) with LR clonase (Invitrogen).

A plasmid constructs containing a full-length cDNA encoding the *Os3BGlu8* rice β -D-glucosidase isoenzymes (AK120790, clone ID 215509) was acquired from the Rice Genome Resource Center, Tsukuba, Japan (<http://www.rgrc.dna.affrc.go.jp/>). The cDNA encoding the mature *Os1BGlu1* isoenzyme was amplified from the clone 204872 plasmid with *Pfu* DNA polymerase and primers that added an *Nco*I site to the start of the predicted mature protein N-terminus encoding sequence and an *Xho*I site to the stop codon encoding sequence. The cDNA encoding the mature *Os3BGlu8* protein was cloned, sequenced and subcloned into pET32a, as described for rHvBII.

Protein expression was carried out after transformation of the pET32a expression constructs into *E. coli* Origami(DE3) cells (Novagen) and selection on Luria-Bertani (LB) agar containing 50 μ g/ml ampicillin, 15 μ g/ml kanamycin, and 12.5 μ g/ml tetracycline. The selected clones were grown and induced for expression as previously described for BGlu1 (Opassiri et al., 2003). The optimum expression conditions were determined by varying the duration of expression from 3 to 18 h and the concentration of IPTG from 0.2 to 0.6 mM, at temperatures of 20°C and 25°C.

The cells were collected by centrifugation after overnight expression, frozen at -70°C and lysed to release total soluble proteins, as previously described (Opassiri et al., 2003). The heterologously expressed enzymes were first purified by IMAC resin (GE Healthcare, Buckinghamshire, United Kingdom) charged with Co²⁺ (rHvBII and *Os3BGlu8*) or with Ni²⁺ (*Os7BGlu26*), as with *Os3BGlu7* (Opassiri et al., 2003; Chuenchor et al., 2006). The IMAC column was washed twice with 10 column volumes

(CV) of equilibration/wash buffer (100 mM NaCl, 50 mM Tris-HCl, pH 7.2) to remove unbound protein. Resin-bound fusion proteins were eluted with 2 CV of elution buffer (150 mM imidazole, 50 mM Tris-HCl, pH 7.2). The eluted fractions were analysed for hydrolysis of 4NPGlc before overnight dialysis in 20 mM Tris-HCl, pH 7.2, at 4°C.

The Os3BGlu8 and Os7BGlu26 proteins were then purified by Q-Sepharose chromatography (GE Healthcare) in 50 mM sodium acetate, pH 5.0, with elution by a gradient from 0-0.5 M NaCl. The Os3BGlu8 and Os7BGlu26 β -D-glucosidases eluted at 0.18 M, and 0.22 M NaCl, respectively. The NaCl concentration of the Os3BGlu8 β -D-glucosidase solution was adjusted to 3 M with solid NaCl and the enzyme was loaded on a Phenyl Sepharose (GE Healthcare) column, which was pre-equilibrated with 50 mM sodium acetate, pH 5.0, containing 3 M NaCl. The protein was eluted with a decreasing gradient of 3-0 M NaCl over 50 ml at a flow rate of 1 ml/min. The Os3BGlu8 β -D-glucosidase eluted at 0.22 M NaCl. The fractions that were active toward *p*NPGlc were pooled and dialysed in 20 mM Tris-HCl, pH 7.2, typically for 16 h at 4°C. Os7BGlu26 was further purified by size-exclusion chromatography on a Superdex S200 column (10 x 300 mm, GE Healthcare), in 50 mM Tris-HCl, pH 8.0, containing 150 mM NaCl, at a flow rate of 0.5 ml/min.

The NH₂-terminal thioredoxin-His tag of the rHvBII proteins were removed by digestion with 2 μ g of enterokinase (New England Biolabs, Beverly, MA, USA) per 1 mg of total protein according to the supplier's instructions. The released fusion tag was removed by adsorption to a second IMAC column. The unbound and wash fractions from IMAC with hydrolytic activity toward *p*NPGlc were pooled and dialyzed for 16 h against 50 mM sodium acetate, pH 5.0. The dialyzed rHvBII was purified by cation-exchange chromatography on CM-Sepharose (Takara, Tokyo, Japan) with elution by a 0-0.5 M NaCl gradient in 50 mM sodium acetate, pH 5.0, at a flow rate of 1 ml/min. rHvBII eluted at 0.15 mM NaCl. Because Os3BGlu8 and Os7BGlu26 appeared to be more stable as fusion proteins, they were not digested with enterokinase, and the fusion proteins were investigated, as was described for BGlu1 (Opassiri et al., 2004).

Tryptic maps were made by MALDI TOF/TOF mass spectrometry (MS) of expressed rHvBII and Os3BGlu8, and confirmed that the purified proteins corresponded to the expected enzymes.

Enzyme activities were assayed for 10-30 min, depending on the enzyme, in reaction mixtures with 1 mM *p*NPGlc in 50 mM sodium acetate, pH 5.0, at 30°C. Reactions were terminated with two volumes of 2 M Na₂CO₃ and absorbance at 405 nm was measured. The optimum temperature was determined over the range of 0 to 90°C in 10°C increments, by incubating 4 pmol of enzyme at the varying temperatures for 15 min, after which *p*NPGlc was added to 1 mM concentration and incubation proceeded for a further 30 min. The optimum pH was determined by incubating 4 pmol of an enzyme with 1 mM *p*NPGlc in McIlvaine buffers, with pH ranging from 3.0 to 12.0, for 30 min at 30°C. The reactions were terminated and rates measured as specified above. The substrate specificities of Os3BGlu8, Os7BGlu26 and rHvBII were evaluated by incubating 4 pmol of enzyme in 100 μ l reaction mixtures with 3 mM substrates at 30°C for 1 to 16 h in 50 mM sodium acetate, at the optimum pH. Hydrolysis of *p*NP glycosides was detected as described above for *p*NPGlc. For cellooligosaccharides and laminari-oligosaccharides, the reactions were terminated by heating at 95°C for 5 min. The amount of glucose released was determined by peroxidase/glucose oxidase coupled reactions (PGO assay, Sigma Aldrich, St. Louis, MO, USA) as previously described (Opassiri et al., 2003). Hydrolysis of mannoligosaccharides and natural glucosides was detected by TLC. Kinetic parameters were determined in a manner similar to that described for BGlu1 above.

Mutations of rHvBII were made to try to determine the differences that cause it to have higher β -D-mannosidase activity than β -D-glucosidase activity, while BGlu1 has higher β -D-glucosidase activity. The mutations were made according to the QuikChange kit, as described for BGlu1. The mutations made included V184I, A187L, L246V, V250N, and Q336S, which changed rHvBII residues to BGlu1 residues, and Y136W, F346P, and E391G, which changed residues common to BGlu1 and HvBII. The mutant proteins were purified as described for HvBII above and their hydrolysis of *p*NPGLc and *p*-nitrophenol β -D-mannoside (*p*NPMan) characterized, as described above.

6.3 Crystallization and preliminary structure analysis of Os7BGlu26 β -mannosidase.

In order to optimize expression of Os7BGlu26 β -mannosidase/ β -glucosidase for crystal structure determination, the pET32a/Os7BGlu26 expression vector described above was transformed into Origami B(DE3) and Rosetta-gami(DE3) strains of *E. coli* and expression tested by inducing at 20 °C with concentrations of IPTG from 0 to 0.5 mM. Since Rosetta-gami(DE3) cells induced with 0.3 mM IPTG at 20°C for 16-18 h gave the best expression levels, this condition was used for production of protein. The protein was extracted and purified by IMAC, followed by enterokinase cleavage and a second IMAC, as described for Os3BGlu7 above. The purified protein was screened for crystallization with the Hampton Research HR134 and Emerald Biosystems Wizard crystallization kits. The condition of 0.5 M Na/K tartarate, 0.1 M HEPES, pH 7.5, was found to give clustered crystals and optimized by varying the Na/K tartarate concentration and pH to give an optimal condition of 0.34 M Na/K tartarate, 0.1 M HEPES, pH 7.25, for single crystal production.

Crystals of Os7BGlu26 were screened for diffraction at the NSRRC beamline 13B with 1.0 Å X-rays, and preliminary datasets were collected for Os7BGlu26 and Os7BGlu26 crystals soaked with 2,4-dinitrophenyl-deoxy-2-fluoromannoside (a kind gift from Prof. Stephen Withers, University of British Columbia, Canada). Datasets were processed with HKL2000, as described above, and the preliminary structure solved by molecular replacement with the Os3BGlu7 (2RGL) structure in the MolRep program of the CCP4 Suite (Murshudov et al., 1999; Computational Computing Project 4, 1994).

6.4. Recombinant expression and characterization of barley Exo I exoglucanase.

RNA was extracted from barley cultivar clipper seedlings and reverse-transcribed, as described above. The cDNA encoding the mature protein of HvExoI (Genbank Accession number AF102868) was amplified with *Pfu* DNA polymerase and the ExoMatF (5'-CACCGACTACGTGCTCTACAAGGA-3' and ExoIstopR (5'-CTAGTACTTCTTCGTCGCGTTGGT-3') primers. After gel-purification, this cDNA was cloned into the pENTR/D-TOPO entry vector according to the manufacturer's instructions (Invitrogen). The cDNA was then transferred to the pET32a/DEST (Opassiri et al., 2006) and pPICZ α BNH₈/DEST (Toonkool et al., 2006) vector by LR Clonase recombination (Invitrogen). An optimized gene with optimal codons for *Pichia pastoris* expression was ordered from GenScript and cloned into the pPICZ α BNH₈ expression vector, between the *Pst* I and *Eco* RI sites. The vectors were sequenced to confirm the correct nucleotide sequences.

The pET32a/DEST vector containing the ExoI cDNA was transformed into Origami(DE3) and Origami B(DE3) *E. coli* strains, and induced as described above. Since only inclusion bodies were produced, attempts were made to refold the protein from inclusion bodies. Folding trials were done by a standard protocol of dissolving the inclusion bodies in 6 M guanidine HCl, 50 mM Tris-HCl, 100 mM NaCl, 10 mM

EDTA and 10 mM DTT, then gradually dialyzing into the similar buffer with decreasing concentrations of guanidine HCl and 1 mM reduced and 0.1 mM oxidized glutathione in place of DTT. Since no soluble protein be detected, the iFOLD protein folding kit 2 (Novagen) was used to screen for folding conditions, according to the manufacturer's specifications. Soluble active protein was screened by assaying for hydrolysis of 4NPGlc.

For expression of the ExoI protein in *P. pastoris*, the pPICZ α -based plasmids were linearized with *Pme* I and used to transform *P. pastoris* strains Y11430 and SMD11680H by electroporation. The transformed yeast were selected on zeocin and single colonies were picked for growth in liquid culture. A single transformed colony was inoculated into 500 mL BMGY medium (Invitrogen) containing 100 lg/mL zeocin and grown at 28 °C with shaking until the culture reached an OD600 of 2–3. Cells were harvested by centrifugation and resuspended in the BMMY medium (Invitrogen) to an OD600 of 1. Expression of rHvExoI was induced with 1% (v/v) methanol for 4 days at 20 °C. Cells were removed by centrifugation and the media used for screening for enzyme activity or protein purification.

Protein was purified from the culture media by ion exchange chromatography and IMAC. First, fresh phenylmethyl sulfonyl fluoride (PMSF) was added to the media to 1 mM and the pH was adjusted to 4.7 with acetic acid. The solution was loaded on an SP-Sepharose cation exchange column and washed with 50 mM acetate, pH 4.7, then eluted with a linear gradient of 0 to 2 M NaCl in the same buffer. The fractions with most of the activity were concentrated in a centrifugal concentrator and the buffer changed to 300 mM NaCl in 50 mM sodium phosphate, pH 7.8. These were then mixed with cobalt-bound IMAC resin and the resin was washed in a column and eluted with a gradient of imidazole from 0 to 0.5 M in the same buffer. Active fractions were changed to 20 mM sodium acetate, pH 5, buffer and concentrated in a centrifugal concentrator (10 kDa MWCO).

A 100 μ g aliquot of the ExoI purified from *P. pastoris* media was deglycosylated with 500 U endoglycosidase F (New England BioLabs) at 4 °C for 3 days. The IMAC purification and concentration was repeated with this sample to remove the enterokinase and cleaved sugar. The substrates hydrolyzed, pH and temperature optima and enzyme kinetics for the glycosylated and deglycosylated proteins were determined similar to those of the other enzymes described above.

To verify the protein identity, 10 mg aliquots of deglycosylated or glycosylated (not deglycosylated) protein was digested with trypsin and the digest was mixed with a-cyano-4-hydroxycinnaminic acid and evaluated by matrix assisted laser-desorption ionization (MALDI) time-of-flight (TOF) mass spectrometry (MS). Similarly, aliquots of undigested proteins were mixed with and submitted to MALDI TOF MS, as well, to determine the molecular weight of the full protein.

A number of mutants of the barley Exo I active site were made by the QuikChange method (Stratagene), as described for other proteins above. Since we would eventually like to compare the structures of the mutants of expressed barley Exo I with the wildtype enzyme to assess the effects of the mutations on the structure and substrate binding to mutants, the protein was screened for crystallization. After the purification steps described above, the N-deglycosylated rHvExoI wild-type and mutant enzymes were eluted from a BioGel-P100 size-exclusion column with 50 mM sodium acetate buffer, pH 5.25 containing 200 mM sodium chloride and 1 mM dithiothreitol, concentrated to approx. 12.5 mg/mL and filtered through a 0.22 μ m filter. Crystallization of the protein was assessed in the standard screening kits in the laboratory, manually by microbatch under oil, and robotically by sitting drop vapor-

diffusion, and by hanging drop under the conditions used to crystallize native barley ExoI (Hrmova et al., 1998) with and without seeding with native barley ExoI microcrystals. For microbatch, the protein was mixed one to one with the precipitants from the Crystal Screen 2, Crystal Screen Lite and Grid Screen Ammonium Sulfate (Hampton Research, Aliso Viejo, CA, USA) kits and incubated at 14 °C. Robotic screening was done at the Bio21 Collaborative Crystallisation Centre (CSIRO, Parkville, Australia) (Newman et al., 2008) at 8 °C and 20 °C, using the PSS_1_Com5 screen formulation from Emerald BioSystems (Bainbridge Island, WA, USA) or the PS gradient-mid range formulation prepared in-house (http://www.csiro.au/c3/Facility/c3_centre_robotic_crystal.htm). For hanging drop vapor diffusion, drops were set up with 1:1 protein under the previously published conditions for native Exo I (Hrmova et al., 1998), which include a precipitant solution of 100 mM Hepes-NaOH buffer pH 7.0, 2.4% (w/v) polyethylene glycol 400, 1.6 M ammonium sulfate. The protein (4 mL 12.5 mg/mL) and precipitant (4 mL) were mixed on 22-mm siliconized circular glass cover slips (Hampton Research). For seeding, a cat whisker was gently touched the surface of a microseed stock of native HvExoI and subsequently the whisker was immersed into the rHvExoI drop, then the slide was fixed with the drop down over the reservoir with vacuum grease.

For the data collection of the initial barley Exo I crystals, single crystals with longest dimensions of 100 to 500 µm were cryo-protected with 20% (v/v) glycerol in precipitant and flash cooled in the cold N₂ stream at the beamline MX1 of the Australian Synchrotron. The X-ray diffraction data sets of wild-type and mutant rHvExoI were collected at 0.5° oscillations through 360°. The data was processed using the HKL 2000 suite of programs (Otwinowski et al., 1997).

7. Results:

7.1. Structural investigation of rice BGlu1 (Os3BGlu7) β-glucosidase and its binding to oligosaccharides.

7.1.1. Structure of rice Os3BGlu7 and docking with cellotriose.

Initially, the structure of rice Os3BGlu7 (BGlu1) was solved by molecular replacement with the white clover cyanogenic beta-glucosidase as the template, as reported in the previous grant (Chuenchor et al., 2006; 2008). In this project, we completed the structure refinement of Os3BGlu7 and its complex with 2-fluoroglucoside (G2F) in an alpha-linkage to the catalytic nucleophile (E386) at 2.2 Å and 1.55 Å resolution, respectively. The residual factors indicated that the models represented the data well, with R of 16.2% and R-free of 20.5% for the apoenzyme structure and R of 17.1% and R-free of 19.7% for the covalent intermediate with G2F. Furthermore, we improved the computational molecular docking of the cellotriose molecule in the active site, and tested the effects of some residues near the active site on oligosaccharide specificity by site-directed mutagenesis.

The structure of Os3BGlu7 is typical for GH1 enzymes, being a (β/α)₈ barrel with the catalytic acid/base at the end of strand 4 of the β-barrel and the nucleotide at the end of strand 7 (Figure 1). Two molecules of protein were found bound by a Zn²⁺ ion chelated between them in the asymmetric unit, however, the protein was seen to be a monomer in solution by gel filtration chromatography and dynamic light scattering. The G2F ring was clearly seen in the complex structure and was found to take a relaxed ⁴C₁ chair conformation. Few changes were seen between the complex and apoenzyme structures, except for the movement of the nucleophile to link to the sugar and slight adjustments of a

few surrounding residues, and these structures had an RMSD of 0.093-0.096 Å upon superimposition.

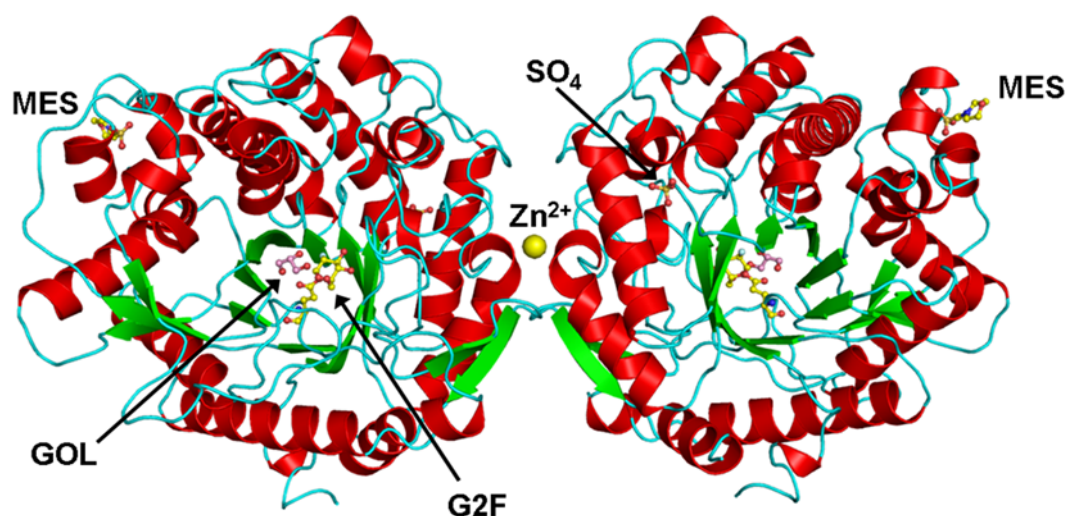


Figure 1. Overall structure of the rice Os3BGlu7 asymmetric unit. The structure of Os3BGlu7 in a covalent complex with 2-fluoroglucoside (G2F) is shown with the other heteroatoms bound to the protein indicated with glycerol (GOL) a MES buffer molecule, sulfate and a single Zn^{2+} ion that was chelated symmetrically between the two protein molecules in the asymmetric unit.

The covalent intermediate structure showed that the G2F ring was forming the typical hydrogen bonds and aromatic stacking interactions with the conserved amino acids in the -1 subsite (Figure 2 a), as previously reported for other similar complexes (Burmeister et al., 1997). Interestingly, a glycerol molecule was also bound in the active site and took a position with the 2-hydroxyl oxygen 2.68 Å from the catalytic acid/base and 3.37 Å from the anomeric carbon, as if it were positioned to donate a proton to the catalytic base and attack the anomeric carbon for a transglycosylation reaction (Figure 2b). The glycerol was held in position by further hydrogen bonds, including possible bonds to both the catalytic acid/base and nucleophile at the glycerol O1 and the glucosyl 6 OH to glycerol O3. Since Os3BGlu7 efficiently transglycosylates to sugars and the glycerol structure is essentially like half a sugar, this structure suggests the mechanism by which this efficient transglycosylation can happen.

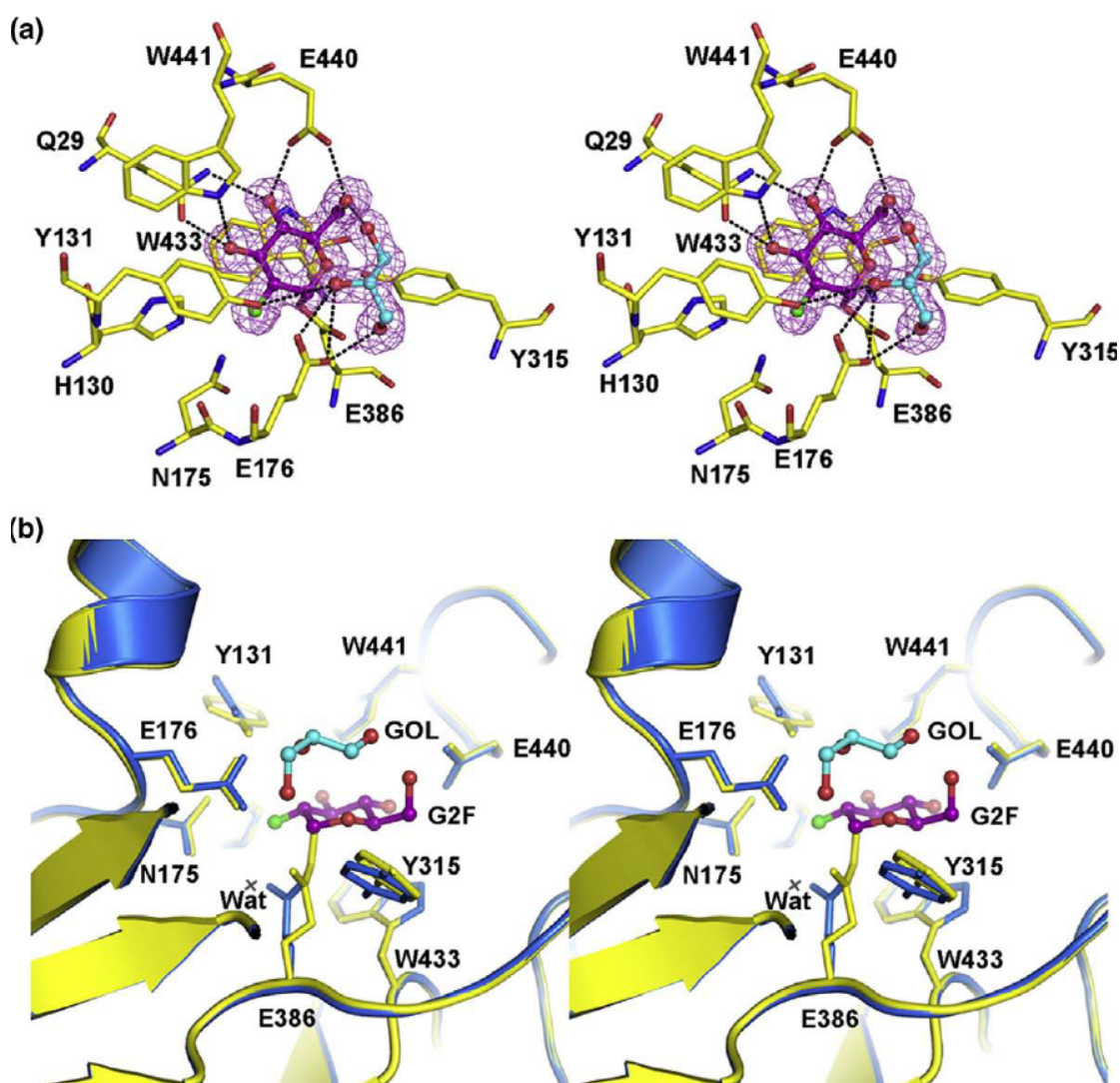


Figure 2. The active site of Os3BGlu7 in the covalent complex with 2-fluoroglucoside. Part a) shows the stereoview of the active site with the |Fo-Fc| omit map of the 2-fluoroglucose and glycerol shown in purple mesh at 3 sigma above background. Groups within the distance to make hydrogen bonds between the ligand and the protein are indicated by dotted lines to represent the possible hydrogen bonds. Part b) is a comparison of the covalent intermediate active site structure in yellow with the apo protein active site in blue. The superimposed active sites show the adjustment of the catalytic nucleophile (E386) and the neighboring tyrosine, Y315, to the covalent binding of the ligand. The stereo view also clearly shows the position of the glycerol 2 OH above the anomeric carbon of the sugar. The figure is from Chuenchor et al., 2008.

In order to get a better idea of the binding of carbohydrates, a cellotriose was docked into the active site of Os3BGlu7 by computational docking. This allowed us to get a general idea where the carbohydrate molecule would bind. In order to get the cellotriose to bind in the correct direction in the active site, the nonreducing terminal glucosyl residue had to be shaped as a 1S_3 skew boat, rather than a relaxed chair. This is in line with previous reports that this shape would be formed upon binding (Davies et al., 2003; Verdoucq et al., 2004). When further structures were determined with the acid/base mutant E176Q with cellotetraose and cellopentaose, it was found that the nonreducing glucosyl residue appeared to be mobile around the C1 to O1 region, and the average shape was

displaced toward a planar transition state in this region, most similar to a 4E envelope, although this is quite similar to the expected 4H_3 half-chair transition state, so that the two may not be easily distinguished. As a result of this distortion and the tetrahedral angle differences around the glycosidic bonds, the second and third glucosyl residues were offset and the positions of the oxygens in the third glucose were off by about 2-3.5 Å in the docked structure, compared to the mutant-complex structure. This meant that the exact hydrogen bonding could not be determined from the docking, although the approximate position suggested some residues to test.

Based on homology modeling based on structures of the previously determined plant GH1 structure active site residues, the molecular docking of cellotriose into Os3BGlu7 and the structures of Os3BGlu7 mutants in complex with cellooligosaccharides, a series of active site mutants were made. The first set consisted of those made from the first two criteria, included the mutations changing +1 to +2 site residues with the corresponding amino acid of barley beta-glucosidase isoenzyme II (HvBII) to see whether the oligosaccharide preferences could be changed to hydrolyze cellobiose better and cellotriose worse, like the barley enzyme. These changes included I179V, N190H and L442R in the +1 subsite and N245V in the +2 subsite. The L442R mutation had little effect, since the oligosaccharide chain runs a different direction out of the active site and is not close to this residue (in both the structures from computational docking and from mutant-substrate complexes, Chuenchor et al., 2008, 2010). However, the I179V increased the k_{cat} and K_m for all substrates, resulting is a slightly improved ratio of k_{cat}/K_m for cellobiose vs. cellotriose (cellobiose $k_{cat}/K_m = 0.06 \text{ mM}^{-1}\text{s}^{-1}$, cellotriose $k_{cat}/K_m = 3.06 \text{ mM}^{-1}\text{s}^{-1}$ ratio = 0.0196) compared to wildtype (cellobiose $k_{cat}/K_m = 0.05 \text{ mM}^{-1}\text{s}^{-1}$, cellotriose $k_{cat}/K_m = 10.7 \text{ mM}^{-1}\text{s}^{-1}$, ratio = 0.00467). The N190H mutation also had an effect, in which it increased the k_{cat} and decreased the K_m for all substrates (4NPGlc, cellobiose and cellotriose), resulting in a slight improvement in the ratio of cellobiose to cellotriose hydrolysis efficiency (cellobiose $k_{cat}/K_m = 0.15 \text{ mM}^{-1}\text{s}^{-1}$, cellotriose $k_{cat}/K_m = 19.1 \text{ mM}^{-1}\text{s}^{-1}$, ratio = 0.00785). The biggest effect was seen for the N245V mutation, which increased the K_m and k_{cat} for both substrates and resulted in a much better ratio of cellobiose to cellotriose hydrolysis efficiency (cellobiose $k_{cat}/K_m = 0.10 \text{ mM}^{-1}\text{s}^{-1}$, cellotriose $k_{cat}/K_m = 0.71 \text{ mM}^{-1}\text{s}^{-1}$, ratio = 0.14). These effects suggested that, while a combination of the residues in the active site may contribute to the cellobiose vs. cellotriose hydrolysis efficiency, no single residue could explain the difference between rice Os3BGlu7 and barley HvBII.

7.1.2. Structural investigation of oligosaccharide binding complexes.

To define the binding better, a series of catalytic residue mutants were tested for their interference with hydrolysis and the isosteric acid/base mutant E176Q was chosen for determining structures with oligosaccharide substrates (Chuenchor et al., 2010). This mutant had the rather unique property of efficiently hydrolyzing substrates with good leaving groups that do not require acid-assistance, such as 4NPGlc and 2,4-dinitrophenyl β -D-glucoside, but not those with poor leaving groups, such as oligosaccharides, provided that a small nucleophile was available to replace the catalytic base. The E176Q and E176A mutants could both be rescued in this manner by nucleophiles, while the E176D mutant could not, but E176Q was most efficient at hydrolyzing the substrates under all conditions. In the case of E176Q and E176A, the basic branch of the pH dependence was largely lost, thereby confirming the lack of acid catalyst that needs the lower pH for protonation. Since the E176Q could not detectably hydrolyze oligosaccharides, they could be used as competitive inhibitors of 4NPGlc hydrolysis, allowing direct measurement of their dissociation constants as the competitive inhibition constants (K_i). As shown in Table I,

the K_i indicated that the strongest binding occurred for glucosyl residue 3 in the +2 site, followed by residue 4 in the +1 site, and significant binding could be detected out to the sixth residue of cellohexaose, consistent with previous results from kinetic subsite affinity experiments (Opassiri et al., 2004).

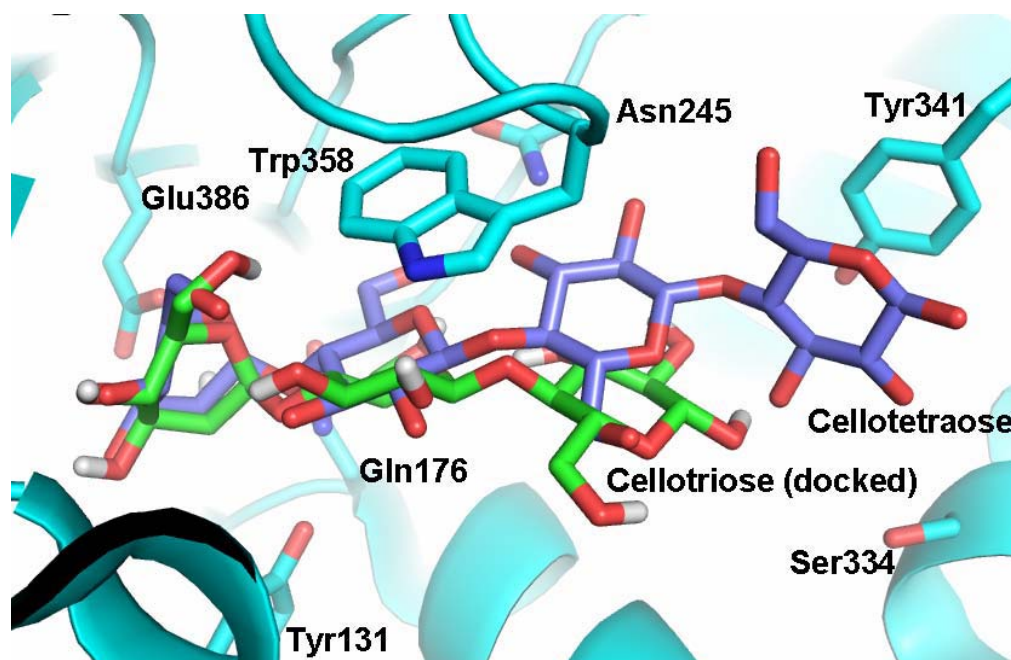


Figure 3. Comparison of the position of a docked cellotriose and a bound cellotetraose in the Os3BGlu7 (BGlu1) active site. The cellotriose was docked into the original structure of Os3BGlu1 (Chuenchor et al., 2008) via Autodock 5.0, whereas the cellotetraose was soaked into crystals of the Os3BGlu7 E176Q mutant (note the Gln 176 in place of Glu 176). Although the cellotriose is in roughly the right position, the difference in 2-3.5 Å in glucosyl residue 3 oxygen positions compared to bound cellotetraose makes it impossible to predict the difference between a direct and a water mediated hydrogen bonding interaction.

Table 1: Inhibition of Os3BGlu7 (BGlu1) hydrolysis of 4NPGlc by oligosaccharides and calculated binding energies. (Chuenchor et al., 2010)

Oligosaccharide	Type of inhibition	Apparent K_i (mM)	Apparent K_A (M^{-1})	Apparent $\Delta G_{\text{binding}}^*$
Cellobiose	Mixed	44 ± 3	23	-7.9**
Cellotriose	Competitive	0.48 ± 0.03	2,100	-19.3
Cellotetraose	Competitive	0.052 ± 0.006	19,200	-24.8
Cellopentaose	Competitive	0.024 ± 0.002	41,700	-26.8
Cellohexaose	Competitive	0.018 ± 0.001	56,000	-27.5
Laminaribiose	Competitive	0.191 ± 0.02	5,200	-21.6

*The apparent K_A and $\Delta G_{\text{binding}}$ are calculated based on the assumption of purely competitive inhibition. As such, they represent maximal values.

**Since cellobiose gave mixed inhibition, the apparent K_A is expected to be an overestimate of the maximal value. The $\Delta G_{\text{binding}}$ is therefore more energetically favorable than is realistic.

The structures of Os3BGlu7 E176Q mutant with the oligosaccharides cellotetraose, cellopentaose and laminaribiose were determined in order to get a more direct idea of the interactions involved in the binding of oligosaccharides in the active site. No crystals could

be produced with cocrystallization with cellobiose and no cellobiose could be seen in crystals soaked with this substrate, suggesting that there is a steric or conformational disturbance in the crystal matrix with this longer substrate. As shown in Table 2, the structures were determined to 1.67 Å resolution for the apo Os3BGlu7 E176Q enzyme, 1.90 Å resolution for its covalent G2F complex, 1.95 Å resolution for the complex with cellotetraose, and 1.80 Å resolution for the structure with cellopentaose. Initially, a 1.37 Å resolution structure was determined for crystals soaked with laminaribiose, but the occupancy was very low and the presence of 2 glycerol molecules in the apo enzyme active site made the exact position of the laminaribiose atoms difficult to interpret. Another dataset with 10 mM laminaribiose, instead of 1 mM, allowed determination of the laminaribiose position, but the resolution was only 2.80 Å. All complexes were refined to give good residual factors and geometric parameters, as shown in Table 3.

Inspection of the oligosaccharide complexes showed that the hydrogen bonds seen in the G2F complex with the wild type Os3BGlu7 complex were maintained in binding of the nonreducing terminal glucosyl residue in the -1 subsite (Chuenchor et al., 2010). No water mediated hydrogen bonds were found at this subsite, but hydrogen bonding interactions at the +1 to +4 sites were predominantly through water, with the only direct hydrogen bonds to Asn 245 in the +2 site (Figure 4). Laminaribiose bound with the nonreducing residue in the same position, but its second residue was flipped approximately 180° compared to the second residue in the cellobiosaccharides, which allowed it to hydrogen bond to Asn 245 with its anomeric hydroxyl in the beta configuration and make maximal contacts between the weakly positively charged protons of the sugar ring and the pi electrons of the Trp 358 indole ring (Figure 4 and Figure 5). These strong interactions at the +1 site could explain why laminaribiose binds much more strongly than cellobiose and even cellotriose. It was also noted that the orientation of the cellobiosaccharides and laminaribiose in the Os3BGlu7 structures was clearly correct for hydrolysis, while the previously determined bacterial structure with cellotetraose had the catalytic acid/base and nucleophile too far away from the glycosidic bond and anomeric carbon to allow hydrolysis in that position (Figure 5 c).

Table 2. Data collection statistics for Os3BGlu7 E176Q and its complexes.

PDB code:	3F4V	3F5I	3F5J	3F5K	3F5L
Complex:	Apo enzyme	G2F	Cellotetraose	Cellopentaose	Laminaribiose
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Unit cell parameters (Å, °)	a = 79.8 b = 101.0 c = 127.6 α = β = γ = 90°	a = 79.2 b = 100.4 c = 127.4 α = β = γ = 90°	a = 79.7 b = 100.8 c = 127.5 α = β = γ = 90°	a = 79.8 b = 101.2 c = 127.4 α = β = γ = 90°	a = 79.7 b = 101.5 c = 128.3 α = β = γ = 90°
No. of molecules per ASU	2	2	2	2	2
Resolution range (Å)	30-1.65 (1.71-1.65)	30-1.90 (1.97-1.90)	30-1.95 (2.02-1.95)	30-1.80 (1.86-1.80)	30-2.80 (2.90-2.80)
No. of observations	636554	560709	462300	429830	158410
No. of unique observations	124448	79635	73603	94774	24681
Redundancy	5.1 (5.0)	7.0 (6.7)	6.3 (5.8)	4.5 (3.6)	6.1 (6.3)
Completeness (%)	99.9 (99.9)	99.1 (98.9)	97.4 (93.7)	98.9 (93.3)	99.8 (99.8)
<i>I</i> /σ(<i>I</i>)	24.5 (4.5)	23.3 (6.1)	19.2 (6.9)	17.9 (2.7)	16.1 (5.8)
<i>R</i> _{sym} (%) ^a	6.2 (36.5)	8.3 (35.2)	8.3 (29.2)	8.3 (45.6)	11.7 (34.1)

Numbers in parentheses are outer shell parameters.

$$^a R_{\text{sym}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl).$$

Table 3. Refinement statistics for Os3BGlu7 E176Q and its complexes.

PDB code:	3F4V	3F5I	3F5J	3F5K	3F5L
Complex:	Apo enzyme	G2F complex	Cellotetraose	Cellopentaose	Laminaribiose
Resolution range (Å)	30-1.65	26.29-1.90	30-1.95	30-1.80	30-2.80
No. of amino-acid residues	944	944	944	944	942
No. of protein atoms	7618	7618	7618	7618	7602
No. of water molecules	901	668	727	801	71
Refined carbohydrate	None	G2F	cellotetraose	cellopentaose	laminaribiose
No. of carbohydrate atoms	None	22	90	112	46
No. of other hetero atoms	53	53	35	47	35
R_{factor} (%) ^b	17.5	18.6	18.0	17.8	20.7
R_{free} (%) ^c	19.6	18.9	21.4	21.0	24.8
Ramachandran statistics					
Favored region (%)	97.4	97.6	97.6	98.0	95.6
Allowed region (%)	2.6	2.4	2.4	2.0	4.4
Outlier region (%)	0	0	0	0	0
R.m.s.d. from ideality					
Bond distances (Å)	0.011	0.014	0.018	0.015	0.008
Bond angle (°)	1.357	1.461	1.585	1.512	1.057
Mean B factors (Å ²)					
All protein atoms	12.03	14.43	9.15	14.74	21.874
Waters	24.43	27.35	19.10	26.88	31.87
Hetero atoms	19.54	28.44	20.59	34.49	46.97
Carbohydrate atoms	None	9.84	24.29	32.71	43.16
Subsite -1 (A/B)	None	9.67/10.01	20.32/19.58	26.09/23.99	41.20/41.93
Subsite +1 (A/B)	None	None	21.73/21.57	29.75/29.06	44.69/44.59
Subsite +2 (A/B)	None	None	21.94/21.82	26.97/26.94	None
Subsite +3 (A/B)	None	None	32.87/32.90	32.91/32.66	None
Subsite +4 (A/B)	None	None	None	49.01/46.49	None

^b $R_{\text{factor}} = (\sum |F_o| - |F_c|) / \sum |F_o|$. ^c Based on 5% of the unique observations not included in the refinement.

Upon inspection of the structures, it was clear that tyrosine residue Y341 appears to interact with the 4th and 5th sugars in the +3 and +4 subsites for cellooligosaccharide binding, while Ser 334 appeared to be involved in a water-mediated hydrogen bond with the sugar in the +3 subsite. Therefore, these residues were mutated and the affect on oligosaccharide binding and hydrolysis was assessed via their effects on k_{cat}/K_m . As seen in Table 4, the S334A mutant had little effect on hydrolysis except for decreasing the k_{cat}/K_m for cellotetraose by about 25%. In contrast, mutations of Y341 to A or L decreased the k_{cat}/K_m for cellotetraose and cellopentaose, but increased it for cellotriose, by increasing the k_{cat} . This may reflect less binding of cellotriose in nonproductive positions including the +3 and +4 sites.

Table 4 Kinetic parameters of BGlu1 wild type and BGlu1 mutants for hydrolysis of cellotriose, cellotetraose and cellopentaose

Substrate	Kinetic parameters	BGlu1	BGlu1 S334A	BGlu1 Y341A	BGlu1 Y341L
Cellotriose	K_m (mM)	0.63±0.02	0.72±0.03	1.26±0.02	1.18±0.03
	k_{cat} (s ⁻¹)	9.66±0.38	11.09±0.44	36.41±0.23	29.60±1.19
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	15.33±0.4	15.40±0.4	28.90±0.3	25.08±0.5
Cellotetraose	K_m (mM)	0.26±0.00	0.50±0.05	0.89±0.09	0.81±0.09
	k_{cat} (s ⁻¹)	11.19±0.58	15.73±0.71	23.71±0.89	18.65±0.35
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	42.69±2.0	31.5±2.4	26.6±2.6	23.0±2.1
Cellopentaose	K_m (mM)	0.21±0.03	0.21±0.01	0.53±0.03	0.43±0.01
	k_{cat} (s ⁻¹)	14.14±0.92	13.05±0.45	21.91±1.32	14.87±0.41
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	67.33±4.7	62.1±4.5	41.34±0.5	34.58±0.2

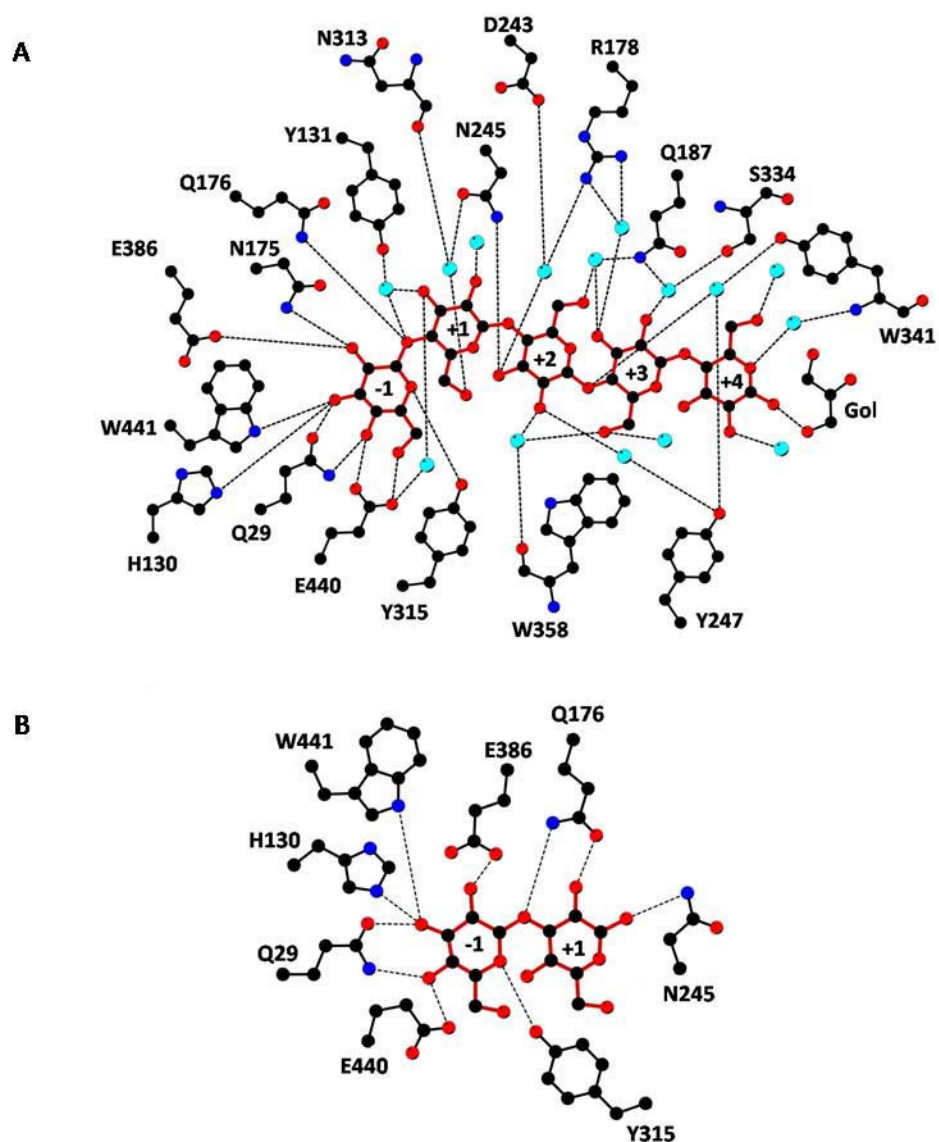


Figure 4. Hydrogen bonding interactions in the active site of Os3BGlu7 (BGlu1) β -glucosidase E176Q mutant bound to oligosaccharides. A shows the interactions formed with cellopentaose in the active site with light colored that are not connected by solid bonds representing water and hydrogen bonds are represented by dashed lines. B shows the interactions with laminaribiose. The figure is from Chuenchor et al., 2010.

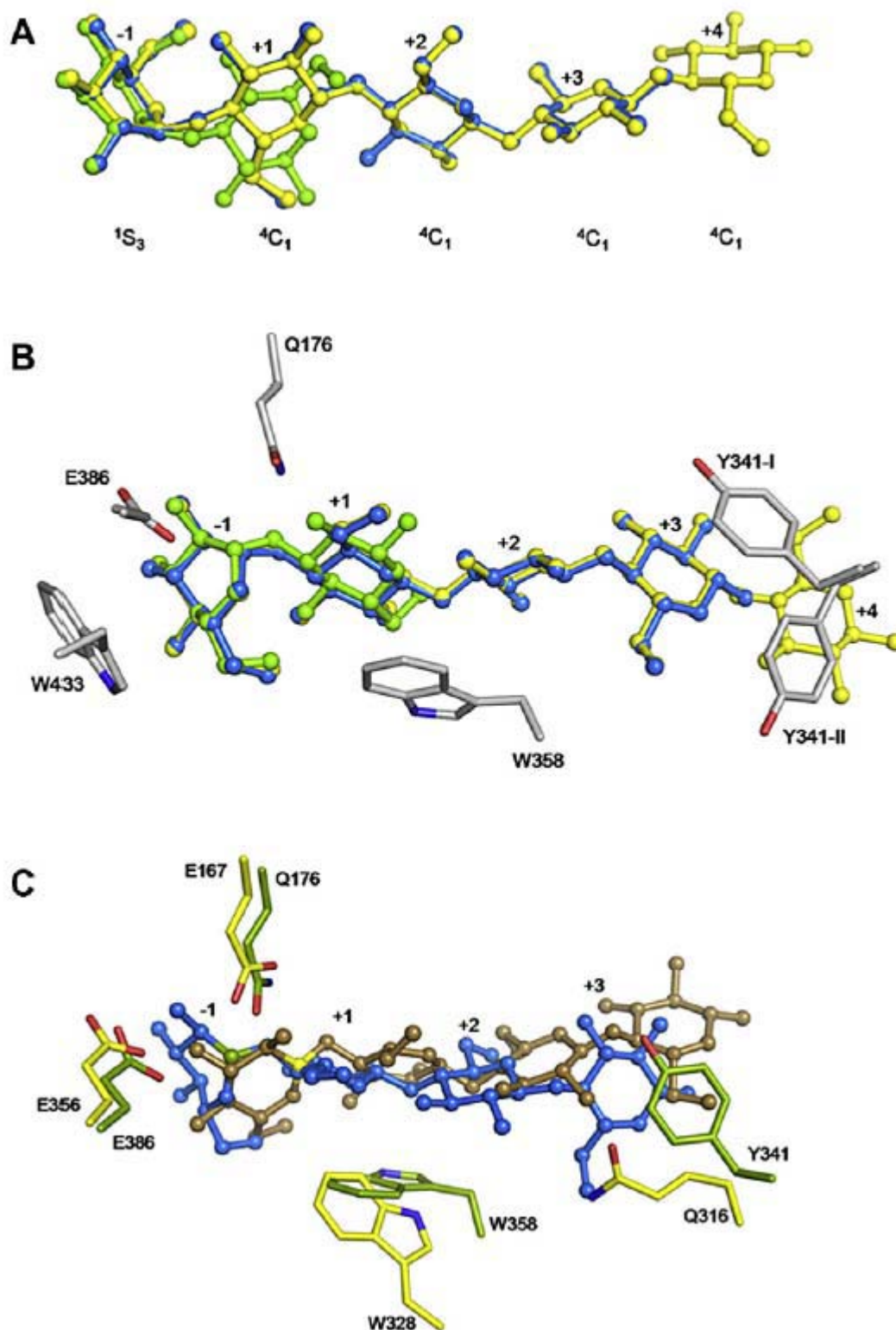


Figure 5. Comparison of oligosaccharide binding positions in the active site of Os3BGlu7. A. The final positions of the oligosaccharides in the cellobiose, cellopentaose and laminaribiose complexes are shown in the superimposed structures, showing the difference in the second glucosyl position in laminaribiose compared to the cellobiose and cellopentaose. B. The positions of cellobiose, cellopentaose and laminaribiose are shown with respect to the aromatic groups that stack the sugars in the Os3BGlu7 active site. C. The sugar positions in Os3BGlu7 and *Paenobacillus polymyxa* β -glucosidase B (Pp BgB) complexes with cellopentaose are compared. Although the nonreducing sugar is in the position for hydrolysis in Os3BGlu7, it is not in Pp BgB.

7.2. Comparison of the catalytic properties of barley rHvBII, its mutants, rice Os3BGlu7 and closely related rice isoenzymes.

7.2.1. Expression and characterization of barley rHvBII and rice isoenzymes in the same phylogenetic cluster.

Protein sequence-based phylogenetic analysis showed that four putative rice β -D-glucosidases, Os1BGlu1, Os3BGlu8, Os7BGlu26, and Os12BGlu38, cluster together with Os3BGlu7, HvBII, and tomato and Arabidopsis β -D-mannosidases (Opassiri et al., 2006; Kuntothom et al., 2009). The amino acid sequence of Os3BGlu8 is most similar to Os3BGlu7 with 74% identity over the catalytic region, but Os3BGlu8 contains a C-terminal extension, including a hydrophobic motif that could act as a membrane anchor at the C-terminus. In contrast, the Os7BGlu26 amino acid sequence shared 82% identity with that of HvBII, but only 64% with that of Os3BGlu7, while that of Os1BGlu1 shared 73% identity to HvBII, but only 64% identity with Os3BGlu7, and Os12BGlu38 is more distantly related to all these enzymes with 56-61% sequence identity. Comparison of the active site residues known from the structure of Os3BGlu7 (2RGL, Chuenchor et al., 2008) and homology modeling revealed the active site residues of Os3BGlu8 were indeed more similar to Os3BGlu7, while those of Os7BGlu26 were more similar to HvBII. To see if their activities showed corresponding similarities, HvBII, Os3BGlu8 and Os7BGlu26 were expressed as thioredoxin fusion proteins in *E. coli*, as previously described for Os3BGlu7 (Opassiri et al., 2003).

The cDNA for HvBII was cloned by RT-PCR from barley cultivar Clipper seedlings and its sequence matched that of BGQ60 β -D-glucosidase at all but four residues, two of which matched barley β -D-glucosidase isoenzyme β II (HvBII) at the residues that were different between HvBII and BGQ60 in the first 110 residues, so this clone was considered to match the HvBII protein and the protein expressed from it was designated as rHvBII (Kuntothom et al., 2009). Its activity was nearly identical to that of HvBII purified from germinating seed (Table 5). The pH and temperature optima of rHvBII were determined to be pH 4.0 and 30°C. When compared with Os3BGlu7 for mannoooligosaccharide hydrolysis, the relative β -D-mannosidase activities of both rHvBII and Os3BGlu7 increased with increasing DP of mannoooligosaccharides from 2 to 6, rHvBII had higher β -D-mannosidase activity compared to Os3BGlu7 for every mannoooligosaccharide.

The structural analysis suggested that Os1BGlu1, Os3BGlu8, and Os7BGlu26 should hydrolyse the same substrates as rHvBII and Os3BGlu7, and that Os1BGlu1 and Os7BGlu26 should show substrate specificity more similar to rHvBII, while Os3BGlu8 should be more similar to Os3BGlu7. To test this, the four rice isoenzymes Os1BGlu1, Os3BGlu8, Os7BGlu26, and Os12BGlu38 were expressed in the *E. coli* expression system that was used for expression of rHvBII and Os3BGlu7. The Os3BGlu8 and Os7BGlu26 were expressed as functional enzymes, while Os1BGlu1 and Os12BGlu38 could not be expressed in active forms in this heterologous host. The pH and temperature optima were at pH 5.0 and 30°C for Os3BGlu8, and at pH 4.5 and 40°C for Os7BGlu26. These values were comparable to the pH 5.0 and 30 °C optima for Os3BGlu7 (Opassiri et al., 2003) and pH 4.0 and 30 °C optima for rHvBII.

7.2.2 Kinetic characterization of rHvBII, Os3BGlu8 and Os7BGlu26.

In substrate hydrolysis, rHvBII, Os3BGlu8 and Os7BGlu26 were similar in hydrolysing *p*NP-glycosides, but they showed differences in hydrolysis of 4NPMann, mannoooligosaccharides and natural β -linked glucosides. While Os3BGlu8 had little activity toward *p*NPMann and mannoooligosaccharides, Os7BGlu26 and rHvBII hydrolysed *p*NPMann better than *p*NPGLc, and also hydrolysed (1,4)- β -linked mannoooligosaccharides

with DP of 2-6, although Os7BGlu26 appeared less efficient at hydrolysing mannobiose than rHvBII. Aside from *p*NPGLc and *p*NPMa, all enzymes hydrolysed *p*NP- β -D-fucoside slightly better than 4NPGLc, and 4NP- β -D-galactoside (*p*NPGal) and *p*NP- β -D-xyloside (*p*NPXyl), and *p*NP- α -L-arabinoside at low rates (Table 5). All three enzymes could hydrolyse (1,4)- β -linked glucooligosaccharides of DP 2-6 and (1,3)- β -linked glucooligosaccharides of DP 2-3, but not laminarioligosaccharides longer than 3 glucosyl residues. All enzymes hydrolysed sophorose [(1,2)- β -linked gluco-disaccharide], but only Os3BGlu8 significantly hydrolyzed gentiobiose [(1,6)- β -linked gluco-disaccharide]. Among the natural glucosides tested, rHvBII hydrolysed the broadest range of aglycones, while Os3BGlu8 could only hydrolyse esculin.

To better characterize their substrate preferences, the apparent kinetic parameters for hydrolysis of *p*NPGLc, *p*NPMa, celooligosaccharides and laminaribiose and laminaritriose by Os3BGlu8 Os7BGlu26, and rHvBII were determined, as summarised in Table 6. Os3BGlu8 hydrolysed *p*NPGLc more efficiently than *p*NPMa (80 times higher k_{cat}/K_m for 4NPGLc), as did Os3BGlu7, to which the Os3BGlu8 sequence was most similar, while Os7BGlu26 and rHvBII, which had sequences most similar to each other, hydrolysed *p*NPMa with apparent k_{cat}/K_m values that were higher than those for *p*NPGLc. All three recombinant enzymes hydrolysed laminaribiose more efficiently than laminaritriose, and hydrolysed celooligosaccharides with higher efficiency as the degree of polymerization (DP) increased from 3 to 6. The two rice isoenzymes hydrolysed cellobiose poorly, as did Os3BGlu7 (Opassiri et al., 2004), but rHvBII hydrolysed cellobiose with a higher k_{cat}/K_m value than celotriose.

To understand the differences in celooligosaccharide preferences between the rice and barley β -D-glucosidases and β -D-mannosidases, their subsite binding affinities for (1,4)- β -linked glucooligosaccharides were estimated, by the method of Hiromi et al. (1973). Figure 6 indicates that all the enzymes have six subsites for (1,4)- β -linked glucosyl residues, as predicted from the similarities of their active site residues to Os3BGlu7 and rHvBII and by molecular modelling. The main difference between rHvBII and the rice isoenzymes was the apparent negative affinity at the +2 subsite in rHvBII and the apparent relatively high positive affinity of all the rice isoenzymes at this site. The subsite affinities of Os7BGlu26 are the most similar to those of the barley enzyme HvBII, with the interaction at the +1 subsite being stronger than at the +2 subsite. The Os3BGlu8 isoenzyme was more similar to Os3BGlu7 (Opassiri et al., 2004), with stronger interaction at the +2 subsite, but still maintained relatively strong binding at the +1 subsite compared to Os3BGlu7.

Table 5 Relative rates of hydrolysis of natural and synthetic substrates by rice Os3BGlu8, Os7BGlu26 and barley rHvBII. Hydrolysis of 4NP-glycosides was detected spectrophotometrically, while hydrolysis of glucooligosaccharides was detected by glucose-oxidase assay and that of mannoooligosaccharides and glucosides by TLC, as described in the Experimental section. Rates of hydrolysis are given relative to 4NPGlc (listed as 100%), which corresponds to 2.59×10^{-4} $\mu\text{mol}/\text{min}/\mu\text{g}$ enzyme for Os1BGlu1, 1.87×10^{-4} $\mu\text{mol}/\text{min}/\mu\text{g}$ for Os3BGlu8, 1×10^{-6} $\mu\text{mol}/\text{min}/\mu\text{g}$ for Os7BGlu26, and 1.6×10^{-4} $\mu\text{mol}/\text{min}/\mu\text{g}$ for HvBII. Characters + or - indicate that hydrolysis was or was not detected, respectively, while N.T. means not tested.

Substrate	Os3BGlu8	Os7BGlu26	rHv β II
<i>p</i> NP- β -D-glucopyranoside	100 $\pm$ 3	100 $\pm$ 4	100 $\pm$ 3
<i>p</i> NP- β -D-mannopyranoside	4.6 $\pm$ 0.4	302 $\pm$ 4	238 $\pm$ 12
<i>p</i> NP- β -D-galactopyranoside	11 $\pm$ 3	12.6 $\pm$ 1.7	17 $\pm$ 3
<i>p</i> NP- β -D-xylopyranoside	2.9 $\pm$ 2.0	6.4 $\pm$ 1.5	3.1 $\pm$ 1.9
<i>p</i> NP- β -D-fucopyranoside	115 $\pm$ 7	128 $\pm$ 8	129 $\pm$ 10
<i>p</i> NP- α -L-arabinopyranoside	2.6 $\pm$ 1.1	9 $\pm$ 3	16 $\pm$ 3
Oligosaccharides			
Sophorose [β -(1,2)-linked]	57 $\pm$ 4	25 $\pm$ 4	15 $\pm$ 4
Laminaribiose [β -(1,3)]	199 $\pm$ 17	158 $\pm$ 13	13 $\pm$ 4
Laminaritriose	39 $\pm$ 10	23 $\pm$ 5	2.8 $\pm$ 1.1
Cellobiose [β -(1,4)]	17 $\pm$ 6	21 $\pm$ 5	33 $\pm$ 7
Cellotriose	83 $\pm$ 18	124 $\pm$ 18	19 $\pm$ 8
Cellotetraose	156 $\pm$ 9	194 $\pm$ 11	52 $\pm$ 9
Cellopentaose	147 $\pm$ 15	403 $\pm$ 17	225 $\pm$ 15
Cellohexaose	188 $\pm$ 16	472 $\pm$ 12	303 $\pm$ 12
Gentiobiose [β -(1,6)]	4.8 $\pm$ 1.0	-	-
Mannobiose [β -(1,4)]	-	+	+
Mannotriose	-	+	+
Mannotetraose	-	+	+
Mannopentose	-	+	+
Mannohexose	-	+	+
Natural glucosides			
Salicin	-	-	+
Esculin	+	-	+
Epiheterodendrin	-	-	+
Dhurrin	N.T.	+	+
D-amgdalin	-	+	+
Prunasin	N.T.	-	N.T.
<i>p</i> -Coumaryl alcohol glucoside	-	+	+
Coniferin	-	-	+

No hydrolytic activities were detected against 4NP- β -L-arabinopyranoside, 4NP- α -D-glucopyranoside, laminaritetraose, laminaripentaose, laminarihexaose, linamarin and indoxyl- β -D-glucoside.

Table 6 Apparent kinetic parameters of Os1BGlu1, Os3BGlu8, Os7BGlu26, rHvBII, HvBII, and Os3BGlu7 toward 4NPGlc, 4NPMAN, cellooligosaccharides with DP of 2-6 and laminaribiose and laminarotriose.

Substrate ^a	Apparent kinetic parameters	Os3BGlu8	Os7BGlu26	rHvBII	HvBII ^b	Os3BGlu7 ^c
<i>p</i> NPGlc	K_m (mM)	0.27 ± 0.04	0.27 ± 0.02	0.50 ± 0.03	0.50	0.23 ± 0.02
	k_{cat} (s ⁻¹)	3.3 ± 0.4	0.16 ± 0.003	0.50 ± 0.07	0.50	7.9 ± 0.4
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	13 ± 3	0.63 ± 0.003	1.00 ± 0.05	1.00	35 ± 1
<i>p</i> NPMAN	K_m (mM)	1.6 ± 0.3	0.52 ± 0.04	0.25 ± 0.01	N.D.	1.3 ± 0.1
	k_{cat} (s ⁻¹)	0.24 ± 0.03	1.10 ± 0.03	3.06 ± 0.02	N.D.	1.32 ± 0.05
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	0.16 ± 0.05	2.11 ± 0.20	12.7 ± 0.2	N.D.	1.01 ± 0.02
Cellobiose	K_m (mM)	26 ± 2	19.6 ± 1.9	2.76 ± 0.10	2.7 ± 0.2	32 ± 2
	k_{cat} (s ⁻¹)	0.98 ± 0.08	0.52 ± 0.03	16.1 ± 0.2	11.6 ± 0.6	1.5 ± 0.1
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	0.04 ± 0.01	$0.026 \pm$	5.8 ± 0.2	4.34 ± 0.07	$0.050 \pm$
Cellotriose	K_m (mM)	0.56 ± 0.01	0.52 ± 0.05	0.74 ± 0.06	0.97 ± 0.06	0.72 ± 0.02
	k_{cat} (s ⁻¹)	6.08 ± 0.08	0.67 ± 0.03	3.0 ± 0.4	2.0 ± 0.1	18.1 ± 0.4
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	11 ± 2	1.27 ± 0.03	3.45 ± 0.08	2.01 ± 0.01	25.4 ± 0.04
Cellotetraose	K_m (mM)	0.25 ± 0.01	0.09 ± 0.01	1.03 ± 0.02	0.89 ± 0.05	0.28 ± 0.01
	k_{cat} (s ⁻¹)	16.1 ± 0.1	0.97 ± 0.02	9.6 ± 0.4	8.9 ± 0.6	17.3 ± 0.6
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	65 ± 2	10.9 ± 0.4	9.3 ± 0.5	9.98 ± 0.08	61.1 ± 0.4
Cellopentaose	K_m (mM)	0.15 ± 0.01	0.06 ± 0.01	0.33 ± 0.02	0.41 ± 0.02	0.24 ± 0.01
	k_{cat} (s ⁻¹)	15.9 ± 0.1	1.07 ± 0.03	12.9 ± 0.4	11.7 ± 0.8	16.9 ± 0.1
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	109 ± 3	18.5 ± 0.4	40.2 ± 0.8	28.4 ± 0.4	72 ± 2
Cellohexaose	K_m (mM)	0.12 ± 0.01	0.05 ± 0.01	0.23 ± 0.01	0.29 ± 0.02	0.22 ± 0.01
	k_{cat} (s ⁻¹)	212 ± 2	1.31 ± 0.04	$13.10 \pm$	11.8 ± 0.8	16.9 ± 0.3
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	175 ± 5	29 ± 3	54 ± 2	40.7 ± 0.1	153 ± 0.5
Laminaribiose	K_m (mM)	0.32 ± 0.07	0.86 ± 0.07	4.7 ± 0.4	5.4 ± 0.4	2.05 ± 0.01
	k_{cat} (s ⁻¹)	11.1 ± 0.3	0.61 ± 0.02	11.3 ± 0.03	14.1 ± 1.0	32 ± 3
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	36 ± 8	0.70 ± 0.06	2.30 ± 0.06	2.63 ± 0.01	16 ± 2
Laminaritriose	K_m (mM)	6.0 ± 0.6	8.7 ± 0.9	2.8 ± 0.2	2.8 ± 0.2	1.92 ± 0.04
	k_{cat} (s ⁻¹)	13 ± 1	2.4 ± 0.1	2.3 ± 0.2	1.44 ± 0.10	21.2 ± 0.2
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	2.2 ± 0.3	0.28 ± 0.006	0.84 ± 0.01	0.52 ± 0.01	11.0 ± 0.2

^a From Kuntothom et al., 2009. Hydrolysis of *p*NP-glycosides was detected as described in the Experimental section.

^b Kinetics constants were taken from Hrmova et al. (1998). N.D. means not determined.

^c Kinetics constants were taken from Opassiri et al. (2004).

Table 7. Kinetic constants of hydrolysis and inhibition by Os3BGlu7 and HvBII^a.

	K_m^a mM	k_{cat}^a sec ⁻¹	$k_{cat}K_m^{-1b}$ sec ⁻¹ ·mM ⁻¹	K_i μM	ΔG^c kJ·mol ⁻¹
Os3BGlu7					
<i>p</i> NP-O-Glc	0.23±0.02	7.90±0.4	35.1±1.0		
<i>p</i> NP-O-Man	1.30±0.1	1.32±0.05	1.01±0.02		
<i>p</i> NP-S-Glc				664.3± 64.4	-18.4
<i>p</i> NP-S-Man				710.3±115.1	-18.3
HvBII					
<i>p</i> NP-O-Glc	0.50±0.03	0.50±0.07	1.0±0.05		
<i>p</i> NP-O-Man	0.25±0.01	3.10±0.02	12.7±0.2		
<i>p</i> NP-S-Glc				94.7±16.8	-23.3
<i>p</i> NP-S-Man				265.9±57.2	-20.7

^aRounded to a single decimal digit, when higher than 1.0.^bData taken from Opassiri et al. (2004) and Kuntothom et al (2009).^cCalculated according to the equation $\Delta G = -RT \ln(1/K_i)$

This table is from Kuntothom et al., 2010.

7.2.3. Inhibition by and saturation transfer difference nuclear magnetic resonance of rHvBII and Os3BGlu7 with thioglucoside and thiomannoside.

Inspection of the active sites between the closely related hydrolases within the HvBII and Os3BGlu7 cluster revealed no differences in the residues in contact with the glycon residue between those with higher β-D-mannosidase vs. higher β-D-glucosidase activities (Kuntothom et al., 2010). To further understand the differences in their interactions with mannose and glucose, the competitive inhibition constants of rHvBII and Os3BGlu7 were determined (Table 6). Surprisingly, Os3BGlu7 had similar and rather high inhibition constants for both *p*NP-thioglucoside and *p*NP-thiomannoside, despite the fact that it has a much lower K_m and higher k_{cat}/K_m for the O-linked substrate 4NPGlc compared to 4NPMann. Remarkably, rHvBII had a 3-fold lower K_i for the thioglucoside than the thiomannoside, despite the fact that its K_m for *p*NPGlc is 2 times higher than *p*NPMann, and the *p*NPGlc k_{cat}/K_m is 6-fold lower than that for *p*NPMann. Although difficult to understand from an overall perspective molecular structure and quantum mechanics and molecular mechanics (QM/MM) calculations suggested that the structure of the thio-bonded glycosides are significantly different from their O-linked counterparts, due to differences in the lengths and angles of the bonds to the sulfur, compared to the oxygen.

To further investigate the differences in binding, saturation transfer difference nuclear magnetic resonance (STD NMR) was used to look at the interactions between the thioglycoside pyranoside hydrogens, and thereby identify the shape of the sugars when bound in the active site. Although binding of the thioglycosides to Os3BGlu7 was weak and only weak signals could be detected, a clear shift in shape was seen when *p*NP-thioglucoside was bound to rHvBII, from a ⁴C₁ relaxed chair in solution, to a ¹S₃ or ³S₅ skew boat shape, while *p*NP-thiomannoside appeared to remain in the ⁴C₁ relaxed chair.

Further investigation of the thio- and O-linked glucoside and mannoside binding was done by QM/MM calculations by our colleagues Michael Raab and Igor Tvaroška (Kuntothom et al., 2010). A homology model of HvBII, based on the Os3BGlu7 structure (2RGM). The geometry of the ¹S₃, ³S₅ and ⁴C₁ conformers of O- and S- linked *p*NPGlc and

*p*NPMan were optimized at the M05-2X/6-31DG* level of theory and were docked into the active sites of the Os3BGlu7 X-ray crystallographic structure model and HvBII homology model with GLIDE. The GLIDE scores confirmed the hypothesis that the O-linked glucoside should bind in a ¹S₃ conformation, with GLIDE scores of -7.24 kcal/mol for Os3BGlu7 and -7.14 kcal/mol for HvBII in ¹S₃, while the ³S₅ conformation gave scores of -6.02 and -5.34 kcal/mol, respectively, and the ⁴C₁ relaxed chair gave scores of -5.82 and -5.20 kcal/mol, respectively. In contrast, the O-linked mannoside gave similar scores for ¹S₃ and ⁴C₁ conformations for both Os3BGlu7 (-6.66 and -6.76 kcal/mol) and HvBII (-5.77 and -5.74 kcal/mol), respectively. Thus, the calculations suggest the mannoside could bind in multiple conformations. Currently, we are trying to further explain these observations by binding of putative transition state analogues in gluco and manno configurations into rice Os3BGlu7 and Os7BGlu26 active sites to provide evidence for the shape of the transitions states that are generated by the enzyme from these initial binding conformations and see whether they match one of the proposed conformational pathways (Davies et al., 2003).

7.2.4. Mutagenesis of HvBII

To further explore the determinants of β -D-mannosidase vs. β -D-glucosidase activities, Os3BGlu7 was mutated at a few further positions to the residues in HvBII, while rHvBII active site residues were mutated to Os3BGlu7 residues. Xu et al. (2004) previously speculated that the LSENG sequence seen around the catalytic nucleophile in HvBII, as well as Arabidopsis and tomato β -mannosidases could help determine the β -mannosidase activity, since most β -glucosidases, such as Os3BGlu7, have ITENG in this position. However, when we simultaneously mutated I384 and T385 of Os3BGlu7 to L and S, respectively, no significant change in the relative hydrolysis of 4NPGlc and 4NPMan were seen. So, inspection of the active site identified several other sites of differences, and the HvBII to Os3BGlu7 mutations of V184I, A187L, L246V, V250N, and Q339S were made, in addition to two mutations to residues seen in other GH1 hydrolases: F346P (seen in Os1BGlu1) and Y136W (seen in most plant enzymes outside the HvBII and Os3BGlu7 phylogenetic cluster). Assays with *p*NPGlc and *p*NPMan showed V184I appeared to have higher activity than wild type, while Q339S was similar and all other mutants had substantially lower activity. Of the mutations that retained reasonable activity, L246V seemed to have the strongest effect in lowering the activity ratio between 4NPMan and 4NPGlc (at 1 mM) from 3.54 to 1.95, while the V240N (reciprocal of the Os3BGlu7 N245V mutant mentioned earlier) had a similar affect, lowering the ratio to 2.23. No single residue could be pinpointed that would cause the difference in glucoside vs. mannoside preference, similar to the mutations that had been made in Os3BGlu7 (Table 8).

Table 8 Effect of mutations in HvBII on relative levels of β -D-glucosidase and β -D-mannosidase activities.

	Specific Activity		<i>p</i> NPMan/ <i>p</i> NPGlc ratio
	<i>p</i> NPGlc	<i>p</i> NPMan	
WT	3.66E-03	1.30E-02	3.54
V184I	5.50E-03	4.41E-02	8.02
A187L	5.87E-04	7.91E-04	1.35
L246V	1.56E-03	3.04E-03	1.95
V250N	6.18E-04	1.38E-03	2.23
Q339S	3.92E-03	1.61E-02	2.49
F346P	3.42E-03	1.53E-03	2.23
Y136W	2.9E-04	2.15E-04	7.22

7.3 Crystal Structure of Os7BGlu26

Because attempts to crystallize rHvBII β -mannosidase failed to produce reproducible crystals, attempts were made to crystallize the rice Os7BGlu26 β -mannosidase. Os7BGlu26 was found to be expressed in higher amounts in the *E. coli* strain Rosetta-gami(DE3) than in the Origami(DE3) strain in which it was previously expressed. The protein could be purified to reasonable purity by IMAC, followed by cleavage of the fusion tag with enterokinase and adsorption of the released tag by a second IMAC step. Screening for crystallization of this protein resulted in crystals in the condition of 0.8 M sodium-potassium tartarate, 0.1 M sodium HEPES, pH 7.5 (Figure 6). Optimization around these conditions resulted in somewhat larger crystals that were used for diffraction of 1.00 Å X-rays at the NSRRC 13B beamline. The dataset was refined with a high resolution cutoff of 2.25 Å, which gave a $I/\sigma(I)$ of 17.4 (4.8 in the high resolution shell), R_{merge} of 11.1% (45.5%), 100% (99.9%) completeness and redundancy of 7.2 (7.0). The crystal was found to fall in the P212121 space group with unit cell dimensions of $a = 68.0$, $b = 73.5$, and $c = 133.3$, with one protein molecule in the asymmetric unit.

The structure was solved by molecular replacement using the Os3BGlu7 (2RGL) structure as a search model, and was found to be very similar to Os3BGlu7 (Figure 7). In the active site, few differences could be seen, although the active site itself appeared to be slightly narrower (Figure 8). Further work is underway to refine the structure further and solve the structure in complex with transition state analogue inhibitors based on mannosyl and glycosyl substrates, as well as to crystallize an active-site mutation that allows binding of substrates, as in the Os3BGlu7 E176Q mutant.

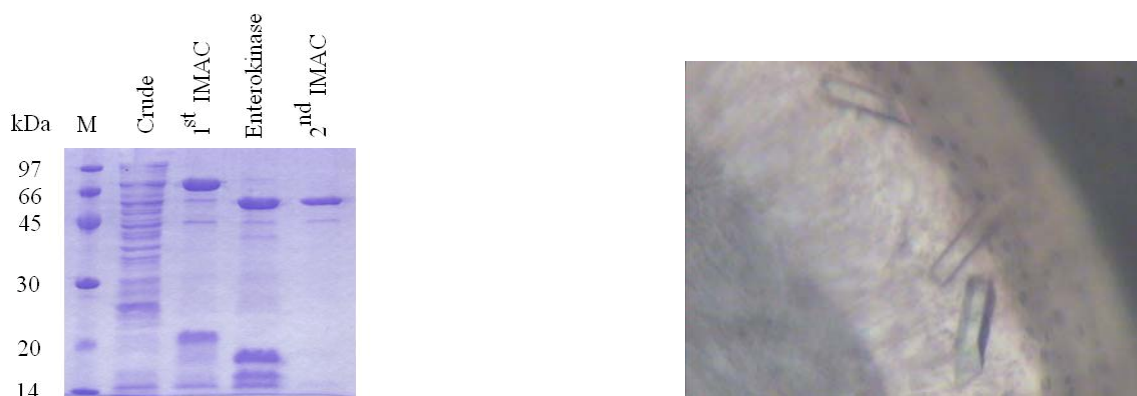


Figure 6. Purification and crystallization of Os7BGlu26 β -mannosidase. The SDS-PAGE gel shows the protein at the end of the different purification steps, from the crude extract, to the first IMAC purification, to enterokinase digest and the second (subtractive) IMAC step to remove the tag. The crystals generated in the screening condition of 0.8 M sodium-potassium tartarate, 0.1 M HEPES, pH 7.5, are shown on the right.

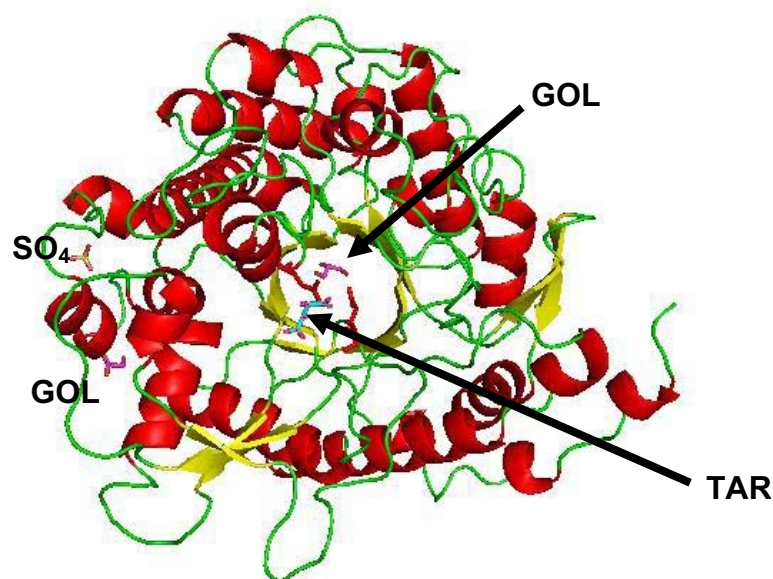


Figure 7. Overall structure of Os7BGlu26 β -mannosidase. The typical $(\beta/\alpha)_8$ -barrel of GH1 proteins is seen, with variations around the loops extending from the C-termini of the β -strands, where the active site is. The structure is very similar to that of Os3BGlu7. The active site is shown with the catalytic acid/base and nucleophile in stick, along with a tartarate (TAR) and glycerol (GOL) that were found in the active site and a second glycerol and sulfate (SO₄) found on the protein surface.

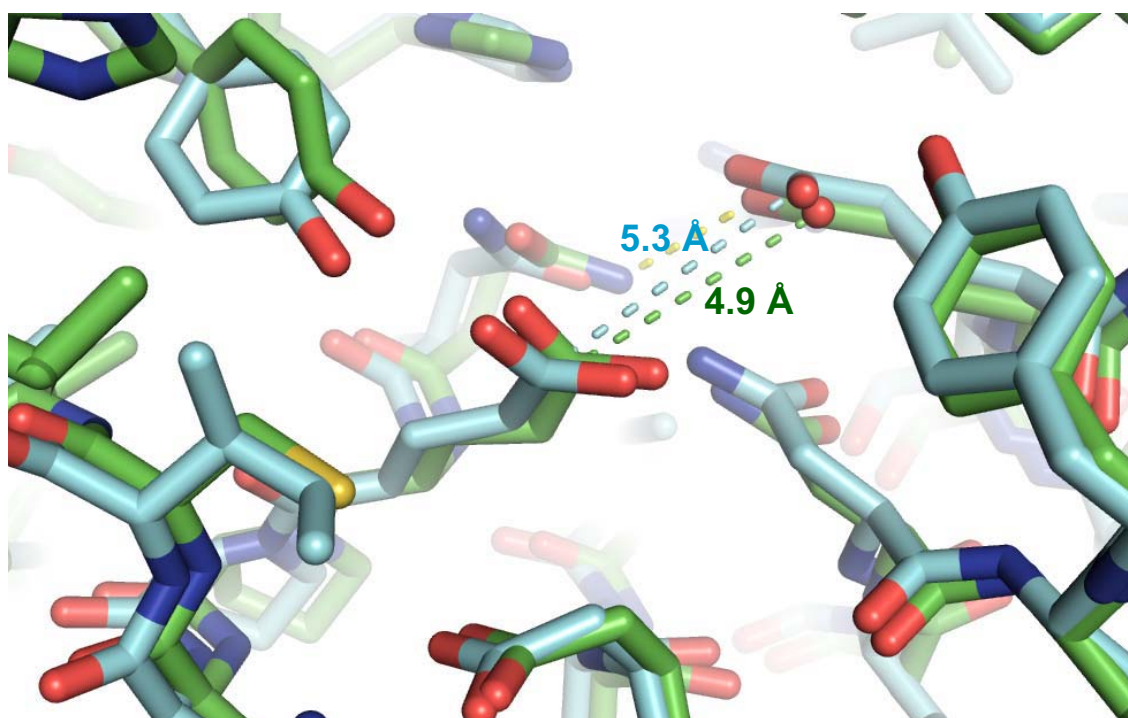


Figure 8. Comparison of the active sites of Os7BGlu26 and Os3BGlu7. The structures were superimposed with an RMSD of 0.581 Å over 2615 atoms. Although the active residues are nearly completely conserved, the active site of Os7BGlu26 appears more narrow with a distance between C δ of the catalytic acid-base and nucleophile of 4.9 Å in Os7BGlu26 compared to 5.3 Å in Os3BGlu7. This would likely allow an extra hydrogen bond to form between Asn178 (next to the catalytic acid/base) and Glu389 (the catalytic nucleophile).

Soaking of Os7BGlu26 crystals in 2,4-dinitrophenol 2-deoxy-2-fluoromannoside failed to produce any mannoside electron density in the active site, and it was observed that the crystal solution rapidly turned yellow from the hydrolysis of this slow substrate-type inhibitor. On the other hand, the 2-F-mannoside could be linked to the nucleophile in Os3BGlu7 crystals and Os3BGlu7 E176Q crystals, and nonmodified mannoside could be trapped on the nucleophile of the E176Q mutant. The electron densities of these covalent complexes clearly showed the mannoside, which appeared to be in a relaxed 4C_1 chair, despite the prediction it might form a different conformation in the mannoside complex. These structures are being further refined for determination of the absolute mannoside conformation and publication.

7.4. Recombinant expression of barley Exo I β -glucan exohydrolase in *Pichia pastoris* and purification of the derived enzyme.

7.4.1 Recombinant expression and characterization of rHvExoI

Barley β -glucan exohydrolase (exoglucanase) Exo I is a family GH3 enzyme, the structure of which has been solved from the native protein from seedlings and shown to comprise two domains, both of which contribute to the active site sandwiched between them (Varghese et al., 1999). Despite many structural studies, no recombinant protein production system was available to test the proposed roles of the active site amino acids by site-directed mutagenesis. Therefore, we amplified the gene from barley seedling cDNA and tried to express the protein in *E. coli* and *Pichia pastoris*. Attempts to produce the protein in the thioredoxin-tagged fusion protein in Origami(DE3) *E. coli* system used for the GH1 proteins resulted in protein production, but all the protein was insoluble and could not be refolded in a wide range of refolding buffers. Therefore, we turned to production in *P. pastoris*.

Initial attempts to express the ExoI protein in *P. pastoris* from the native plant cDNA fused to the yeast prepro-alpha factor sequence for secretion resulted in no increase in exoglucanase activity in the media and no obvious band of expected sized protein in the media (Luang et al., 2010a). Since we used our own designed plasmid with an N-terminal His-tag, we carried out an immunoblot of the media with anti-histidine antibody, and found that a faint band was seen at around 45 kDa, instead of at the expected 65-70 kDa. Since the band might have been derived from proteolysis or premature termination of translation due to poor codon usage for *P. pastoris* in the native gene, which caused a low level of expression as well, we expressed the protein from a synthetic codon-optimized gene in the protease-deficient *P. pastoris* strain SMD11680H. In fact, we could only get high expression of active exoglucanase Exo I in the media when the protein was expressed from the optimized gene in SMD11680H cells with induction at the low temperature of 20°C. No significant activity was seen with the native cDNA under these conditions or with the optimized gene in *P. pastoris* strain Y11430 (wildtype) or when expression was done at 28°C.

The protein could be purified from the *P. pastoris* media by ion exchange, followed by IMAC, but was seen as a smear from about 75 to 90 kDa. Western blotting verified that this was the correct protein. Upon digestion with endoglycosidase H, the protein was reduced to a single predominant band, which was measured to be 67.168 kDa by MALDI-ToF MS. This is approximately the mass expected for the protein with one N-acetylglucosaminyl residue remaining at each of the three N-linked glycosylation sites, which are glycosylated in the native protein from barley seedlings. Tryptic mapping and mass spectrometry of the purified endo H-treated protein identified 9 peptide fragments, covering 65% of the protein sequence. Three of these peptides, each of which contained a

single putative N-linked glycosylation site had masses increased by 203 mass units compared to the amino acid residues, indicating an N-acetylglucosamine residue left after removal of the remaining carbohydrate with Endoglycosidase H at each of the three sites known to be glycosylated in the native protein. Thus, although the protein was glycosylated with more sugar, it was glycosylated at the same positions as in the native protein purified from barley.

The purified glycosylated and deglycosylated proteins were tested for hydrolysis of a variety of possible exoglucanase substrates, as shown in Table 9 (Luang et al., 2010a). It appeared that the recombinant barley enzyme both before and after deglycosylation had activity on plant β -glucans and pNPGlc that were similar to the protein purified from seedlings. This suggested rHvExoI would be a good model for studying the catalytic site amino acids which had previously been hypothesized to have various functions based on X-ray crystal structure by site-directed mutagenesis. The glycosylated rHvExoI was further tested and found to hydrolyze a range of oligosaccharides, including the β -1,2-linked disaccharide sophorose, the β -1,3-linked laminaribiose to laminariheptaose, the β -1,4-linked cellobiose to cellohexaose, and the β -1,6-linked disaccharide gentiobiose. Kinetic studies verified this, although there were some slight differences that might have been derived from the nonnative glycosylation of the recombinant enzyme (Luang et al., 2010a). The main difference was that the recombinant enzyme had a higher K_m and slightly lower k_{cat} for laminarin and higher k_{cat} for barley β -glucan, compared to the native enzyme, which resulted in the recombinant enzyme hydrolyzing barley β -glucan more efficiently than laminarin, while the native enzyme showed the opposite preference. In addition to hydrolysis, the rHvExoI could also catalyze transglycosylation of pNPGlc and glucose released from barley β -glucan, to make gentiobiose as the most stable product (Figure 10).

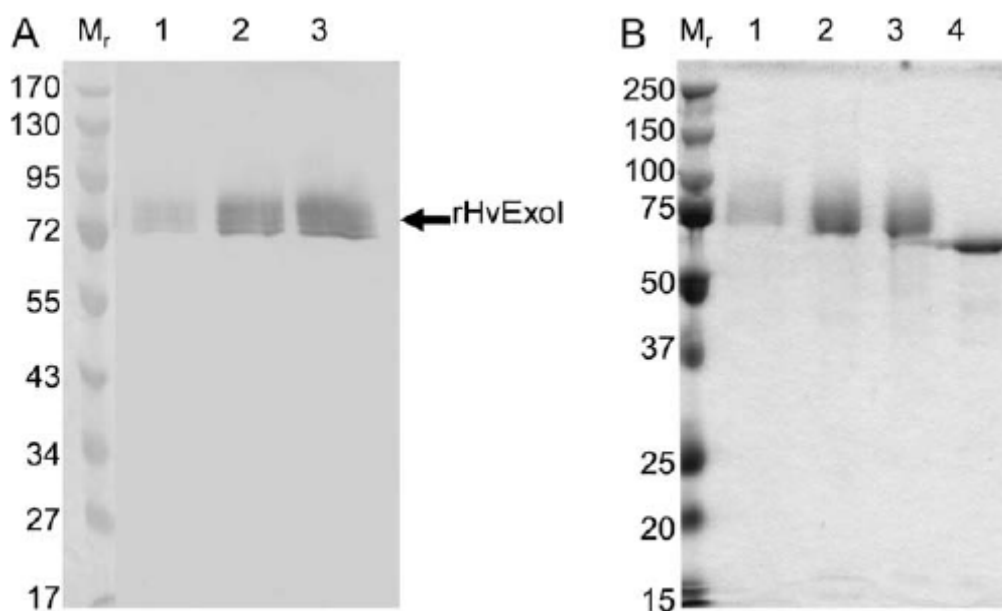


Figure 9: Detection and purification of rHvExo I from *Pichia pastoris* media. A. Western blot of *P. pastoris* media 1, 2, and 3 days after induction of HvExoI expression from the optimized gene. B. SDS-PAGE gel of 1, desalted media; 2, SP-sepharose-purified protein; 3, IMAC purified protein; and 4, protein after deglycosylation with Endo H and repurification over a second IMAC column.

Table 9 Relative activities of rHvExoI on polysaccharides, oligosaccharides and synthetic substrates.

Substrate	Relative activity (%) ^a		
	Native HvExoI ^b	Glycosylated rHvExoI	N-Deglycosylated rHvExoI
<i>polysaccharides</i>			
Laminarin (<i>L. digitata</i>)	100 ^c	98.9 ± 1.1 ^d	98.8 ± 1.2 ^d
Barley (1,3;1,4)-β-D-glucan	10	10.0 ± 0.5	10.0 ± 0.8
Lichenans (<i>C. islandica</i>)	nm ^e	1.37 ± 0.09	nm
<i>Synthetic substrate</i>			
<i>p</i> -nitrophenyl β-D-glucopyranoside	10	22.8 ± 1.4	16.8 ± 0.8

^a The numbers were rounded to three significant figures.

^b The data are from Hrmova et al., 1998.

^c The relative activity of 100% equals to 63 units/mg.

^d The relative activity of laminarin equals to 54.4 and 65.3 units/mg of glycosylated and N-deglycosylated forms of rHvExoI, respectively.

^e 'nm' indicates 'not measured'.

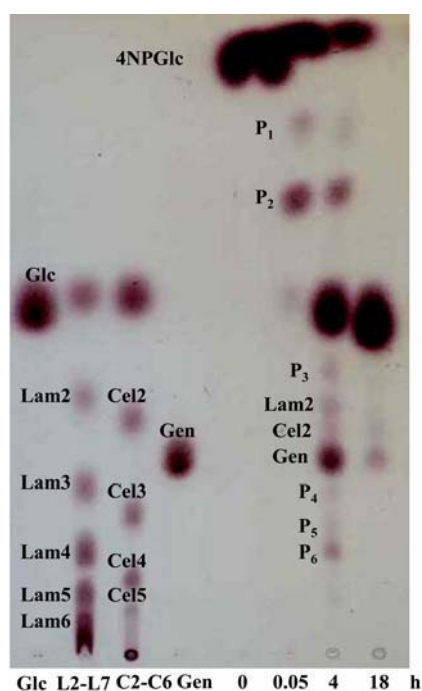


Figure 10 TLC chromatogram of hydrolysis and transglycosylation products formed by glycosylated rHvExoI with *p*NPGlc as a substrate. The enzymes were incubated in the presence of 20 mM *p*NPGlc at 30°C for 0, 3 min, 4 h and 18 h. Standards are glucose (Glc), laminari-oligosaccharides (L2-L7), cello-oligosaccharides (C2-C6), and gentiobiose (Gen). The presence of oligosaccharide products with unknown structures are indicated as P₁-P₆. The reaction times are indicated in hours below the lanes.

7.4.1. Mutagenesis of rHvExoI and characterization of mutants.

In order to characterize the functional importance of the amino acids in the active site, a series of mutations were made, which included: D95N, D95A, E220Q, E220A, E161A, E161Q, R158A, K206V, and R158A/E161A to test residues in the -1 subsite for substrate binding and W286A and W434A in the +1 subsite. Most of these mutations

destroyed the protein activity, since they are conserved active-site residues, but the E220A and W434A mutant had decreased, but relatively high beta-glucosidase activities toward *p*NPGlc. Since these two mutations are in the neighborhood of the catalytic acid/base and catalytic nucleophile, their pH optima were tested, and it was found that the W434A mutant caused a slight increase in the pH optimum, while the E220A mutant caused a narrowing of the pH range activity range at the upper end (Figure 11). This was consistent with the position of the E220 residue near the catalytic acid/base, where it is expected to shift the pKa of that group toward a higher pH in the wild type enzyme.

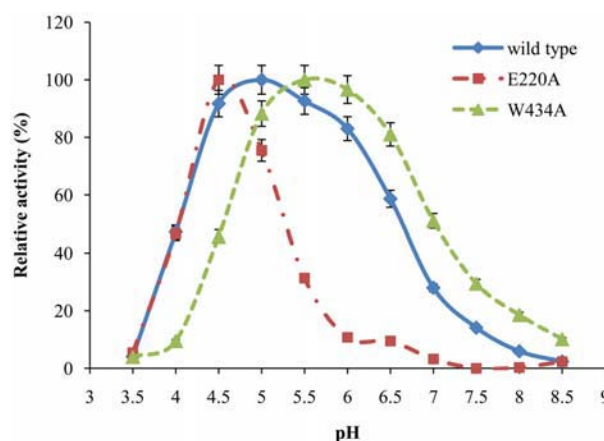


Figure 11. pH profiles of activity for wild type and mutant rHvExoI enzymes. Activity was assayed with 0.2% (w/v) 4NPGlc and 160 μ g/ml BSA in 0.1 M citric acid - 0.2 M disodium hydrogen phosphate (McIlvaine buffers) over the pH range 3.0-8.5, at 30°C for 15 min (wild type rHvExoI) or 60 min (E220A and W434A mutants of rHvExoI).

Analysis of the substrate specificity of the active mutants, showed that the rHvExoI E220A mutant retained activity to β -(1,2)-, β -(1,3)-, β -(1,4)-, and β -(1,6)-linked oligosaccharides, though at a lower hydrolytic rate than wild type (Table 10). However, it had very low activity to laminarin and no activity to barley (1,3;1,4)- β -D-glucan. For W434A mutant was more affected, in that it could not hydrolyse polysaccharides, the β -(1,2)-linked disaccharide sophorose, β -(1,4)-linked cello-oligosaccharides and the β -(1,6)-linked disaccharide gentiobiose, but only retained activity to *p*NPGlc and β -(1,3)-linked laminari-oligosaccharides. The kinetic behavior of the mutants is shown in Table 11. The k_{cat}/K_M values of E220A were decreased approximately 15-fold for laminaribiose, 16-fold for cellobiose and 11-fold for 4NPGlc compared to wild type rHvExoI. The W434A mutant had a K_M value for laminaribiose that was increased approximately 1.6-fold and its k_{cat}/K_M value was decreased approximately 32-fold compared to wild type. Surprisingly W434 seemed to improve the activity to *p*NPGlc, since although both the K_M and k_{cat} values were lower, the k_{cat}/K_M value was higher than wild type rHvExoI.

Table 10. Substrate specificity of E220A and W434A mutants compared with wild type rHvExoI.

Substrate	Activity (μ mole/min/mg protein)			Relative activity ^a (%)	
	wild type	E220A	W434A	E220A	W434A
<i>Polysaccharides</i>					
Laminarin (<i>L. digitata</i>)	54.4	0.4	nd	0.7	nd
Barley (1,3;1,4)- β -D-glucan	5.4	nd ^b	nd	nd	nd
Lichenan (<i>C. islandica</i>)	0.8	nd	nd	nd	nd
<i>Oligosaccharides</i>					
Sophorose	6.0	1.5	nd	24.8	nd
Laminaribiose	20.8	1.6	0.5	7.5	2.4
Laminaritriose	42.1	3.1	0.9	7.4	2.0
Laminaritetraose	39.1	3.3	1.1	8.3	2.7
Laminaripentose	33.8	3.3	1.0	9.7	3.1
Laminarihexaose	46.9	3.3	1.0	7.0	2.1
Laminariheptaose	43.9	3.1	1.0	7.1	2.2
Cellobiose	4.6	0.3	nd	6.7	nd
Cellotriose	8.8	0.4	nd	5.0	nd
Cellotetraose	8.6	0.5	nd	5.5	nd
Cellopentose	8.4	0.3	nd	3.8	nd
Cellohexaose	9.1	0.6	nd	6.5	nd
Gentiobiose	4.7	0.6	nd	11.9	nd
<i>Synthetic substrates</i>					
<i>p</i> -nitrophenyl β -D-glucopyranoside	12.4	1.6	6.1	12.5	48.8
<i>p</i> -nitrophenyl β -D-galactopyranoside	nd	nd	nd	nd	nd
<i>p</i> -nitrophenyl β -D-xylotopyranoside	nd	nd	nd	nd	nd
<i>p</i> -nitrophenyl β -D-fucopyranoside	nd	nd	nd	nd	nd
<i>p</i> -nitrophenyl β -D-lactoside	nd	nd	nd	nd	nd
<i>p</i> -nitrophenyl β -D-glucosiduronic acid	nd	nd	nd	nd	nd
<i>p</i> -nitrophenyl N-Acetyl- β -D-glucosaminide	nd	nd	nd	nd	nd
<i>p</i> -nitrophenyl α -L-arabinopyranoside	nd	nd	nd	nd	nd
<i>p</i> -nitrophenyl β -D-cellobioside	nd	nd	nd	nd	nd
<i>p</i> -nitrophenyl α -L-arabinopyranoside	nd	nd	nd	nd	nd
<i>p</i> -nitrophenyl β -D-cellobioside	0.9	nd	nd	nd	nd

^a Relative activity (%) compared to wild type rHvExoI hydrolysis of the same substrate.

^b 'nd' indicates as 'not detectable'.

Table 11 Kinetic parameters of the rHvExoI & its E220A and W434A mutants.

Substrate	Glycosylated rHvExoI		
	wild type	E220A	W434A
Laminarin			
K_M (mM)	0.18 ± 0.01	nm ^a	nm
k_{cat} (s ⁻¹)	42.1 ± 1.6	nm	nm
k_{cat}/K_M (mM ⁻¹ .s ⁻¹)	234 ± 14	nm	nm
Barley (1,3;1,4)-β-D-glucan			
K_M (mM)	0.04 ± 0.004	nm	nm
k_{cat} (s ⁻¹)	14.1 ± 0.7	nm	nm
k_{cat}/K_M (mM ⁻¹ .s ⁻¹)	365 ± 57	nm	nm
Laminaribiose			
K_M (mM)	1.3 ± 0.1	3.0 ± 0.2	2.1 ± 0.2
k_{cat} (s ⁻¹)	54.2 ± 2.3	8.9 ± 0.3	2.4 ± 0.1
k_{cat}/K_M (mM ⁻¹ .s ⁻¹)	41.2 ± 2.4	2.8 ± 0.1	1.3 ± 0.1
Cellobiose			
K_M (mM)	1.2 ± 0.1	1.7 ± 0.2	nm
k_{cat} (s ⁻¹)	4.2 ± 0.2	0.27 ± 0.01	nm
k_{cat}/K_M (mM ⁻¹ .s ⁻¹)	3.3 ± 0.2	0.16 ± 0.01	nm
pNPGlc			
K_M (mM)	2.0 ± 0.2	1.6 ± 0.1	0.6 ± 0.1
k_{cat} (s ⁻¹)	27.7 ± 0.9	2.1 ± 0.1	9.3 ± 0.3
k_{cat}/K_M (mM ⁻¹ .s ⁻¹)	13.9 ± 0.6	1.3 ± 0.1	16.7 ± 1.2

^a 'nm' indicates as 'not measured' because the activities were too low.

Inhibition of the glycosylated forms of the rHvExoI E220A and W434A mutants was analysed with the competitive inhibitors of wild type rHvExoI methyl-O-thiogentiobiose (G6sG-OMe), glucono δ-lactone and 2,4-dinitrophenyl 2-fluoro-2-deoxy-β-D-glucopyranoside (DNP2FG), as shown in Table 12. Mutation of E220 to alanine resulted in a 10-fold increase in the K_i value for gluconolactone inhibition. While the activity of W434A mutant was not inhibited with G6sG-OMe at 20 mM, the maximum concentration that could be tested. This agrees with the fact that rHvExoI W434A could not hydrolyse gentiobiose.

We had intended to similarly express and characterize rice GH3 enzymes, however, the gene structure for the isoenzyme most similar to rHvExoI appeared to make cloning difficult. Therefore, though we were able to PCR amplify the coding region, we could not clone it, and, since the codon usage of the native gene appeared to be poor for expression in *E. coli* and *P. pastoris*, we ordered a synthetic gene, which is currently being cloned into the expression vectors. Instead, we were able to amplify and clone the rice homologue of barley Exo II (Hrmova et al., 1996), but the cDNA for this enzyme is also in the process of subcloning and the protein has yet to be expressed. It is expected that these enzymes will have similar properties to their barley counterparts, but functional testing is necessary to confirm this.

Table 12 Inhibition of recombinant wildtype and mutant Exo I enzymes.

Inhibitor	rHvExoI			
	Glycosylated wild type	N-deglycosylated wild type	Glycosylated E220A	Glycosylated W434A
<i>G6sG-OMe</i>				
K _i (M)	109 x 10 ⁻⁶	71 x 10 ⁻⁶	221 x 10 ⁻⁶	> 0.02
ΔG (kJ mol ⁻¹) ^a	-23.0	-24.1	-21.2	-
<i>Gluconolactone</i>				
K _i (M)	10 x 10 ⁻⁶	8.3 x 10 ⁻⁶	103 x 10 ⁻⁶	15.3 x 10 ⁻⁶
ΔG (kJ mol ⁻¹)	-29.1	-29.5	-23.1	-28.0
<i>2F-DNPGlc</i>				
K _i (M)	0.5 x 10 ⁻⁶	0.6 x 10 ⁻⁶	0.5 x 10 ⁻⁶	0.6 x 10 ⁻⁶
ΔG (kJ mol ⁻¹)	-36.8	-36.2	-36.5	-36.3

^a Gibbs free energy of binding ΔG was calculated from $\Delta G = -RT \ln (K_i)$

7.4.3. Crystallization of rHvExoI (Luang et al., 2010b).

Since understanding the structure-function relationships of mutant proteins should include structures of these proteins with relevant substrates or other ligands, conditions for crystallization of the recombinant enzyme were tested. The conditions that were tested included repeating the hanging drop vapor diffusion conditions originally used for the crystallization of the native protein from the plant, screening by microbatch, robotic screening of conditions in sitting drop vapor diffusion, and seeding of native protein crystals into hanging drops with recombinant protein. In each case the deglycosylated protein, which included an extra 11 residues of AAHHHHHHHHA that was characterized in the previous section was utilized with a final step of gel filtration chromatography to improve the purity slightly.

The microbatch-under-paraffin-oil trials were done at 4°C and 14°C with recipients from the Crystal Screen 2, Crystal Screen Lite and Grid Screen Ammonium Sulfate kits to screen approximately 130 independent conditions, including those that were previously found to be successful with a native plant enzyme (Hrmova *et al.*, 1998). Previously, these conditions (1.7 M ammonium sulfate, 75 mM HEPES-NaOH buffer, pH 7, containing 7.5 mM sodium acetate and 1.2% (w/v) PEG 400) produced native protein crystals that belonged to a primitive tetragonal space group P4₃2₁2 and yielded high-resolution diffraction patterns (Hrmova *et al.*, 1998; Varghese *et al.*, 1999). However, diffraction-quality crystals did not form and, after approximately 14 days, only thin needles of approximately 40-60 μm in their longest dimensions were observed in 1.6 M magnesium sulfate at pH 6.5 and 4°C.

We set-up close to 400 trials at the Bio21 Collaborative Crystallisation Centre in Australia at 8°C and 20°C by robotic sitting-drop vapor-diffusion trials. After approximately 21 days at 8°C, we could observe the formation of short thin needle-shaped crystals that either remained dispersed throughout the droplets or formed well-organized round balls. However, after about a month, in some of the droplets with needles, we could observe truncated bi-pyramidal crystals that reached the dimensions of 80 x 40 x 60 μm after about 97 days (Fig. 12). It was of note that these crystals were only observed at 8°C

in the droplets with 1.6 to 2.2 M ammonium sulfate, containing 10 mM malate-MES-Tris buffer, pH 5. These conditions were similar to the conditions that we found previously with a native HvExoI (Hrmova *et al.*, 1998), except that here the pH value 5 was more acidic than that used previously (pH 7), and that the protein concentration was almost twice as high (12.5 mg·ml⁻¹ versus 6.8 mg·ml⁻¹). It is likely that the truncated bi-pyramidal crystals may belong to a tetragonal, such as the P4₃2₁2 space group previously seen for the native protein (Hrmova *et al.*, 1998), although at this stage we were not able to collect their diffraction patterns. It was also noted that crystals were not formed in an identical screen at 20°C.

When native crystal-derived microcrystals were used to macroseed into hanging drops containing the recombinant rHvExoI and its mutants, production of larger crystals was successful. Approximately 48 h after rHvExoI was macroseeded with the native microcrystals of the sizes between 10 x 5 x 7.5 µm and 20 x 10 x 15 µm, the original native microcrystal started growing in size. The fully grown crystals of wild-type rHvExoI reached the dimensions that varied between 100 x 50 x 75 µm (Fig. 12) and 500 x 250 x 375 µm after 5 to 7 days and these crystals were isomorphous to the native microcrystals (Fig. 12). Having succeeded in growing wild-type crystals by seeding with native microcrystals, crystals of three rHvExoI variants, Glu220Ala, Trp434Ala and Arg158Ala/Glu161Ala, were grown in the same manner. The newly grown variant crystals reached similar sizes of between 100 x 50 x 75 µm and 500 x 250 x 375 µm than the wild-type crystals and were also isomorphous to the native microcrystals. The variant crystals grew slightly slower and reached their maximum dimensions after 10-14 days.

Single wild-type and variant rHvExoI crystals were cryo-protected and subjected to diffraction at the MX1 beamline of the Australian Synchrotron. The analyses of the X-ray diffraction data sets indicated that all the datasets, collected at 0.5° oscillations through 360°, were virtually complete beyond 1.57 Å to 1.95 Å (Table 13). Autoindexing determined that the space groups of the wild-type and variant rHvExoI crystals were consistent with a primitive tetragonal space group, and preliminary molecular replacement phasing showed they belonged to the P4₃2₁2 spacegroup, similar to their native counterparts (Varghese *et al.*, 1999; Hrmova *et al.*, 2001). Thus, we were able to achieve crystals for structure determination of mutant proteins that will facilitate further studies to understand structure to function relationships in barley ExoI.

Table 13 Data collection statistics from wild-type and variant rHvExoI crystals

rHvExoI	Unique reflections	Resolution ¹ Å	Mean multiplicity ^{1,2}	Completeness ^{1,2} %	Mean I/ σ (I) ¹	R _{merge} ¹⁻³ %	a=b Å	c Å
Wild-type	119968	1.57(1.6-1.57)	28.8(26)	99(86)	71.2(5.6)	6.7 (47)	99.22	183.46
Glu220Ala	79237	1.80 (1.83-1.80)	24.8(13)	95.3(68)	48.6(2.3)	10.6 (99)	100.23	183.17
Trp434Ala	64460	1.95(1.98-1.95)	27.1(16)	100(100)	37.0(2.2)	10.1 (87)	100.05	183.57
Arg158Ala/ Glu161Ala	107601	1.65(1.68-1.65)	26.3(12)	99.8(98)	58.7(2.3)	5.6 (82)	100.78	183.16

¹ Numbers in parenthesis represent the values in the highest resolution shell.

² The numbers in parenthesis were rounded to the nearest decimal place.

³ $R_{\text{merge}} = 100[\sum(I_i - \langle I \rangle)^2 / \sum I_i^2]$ is summed over all independent reflections.

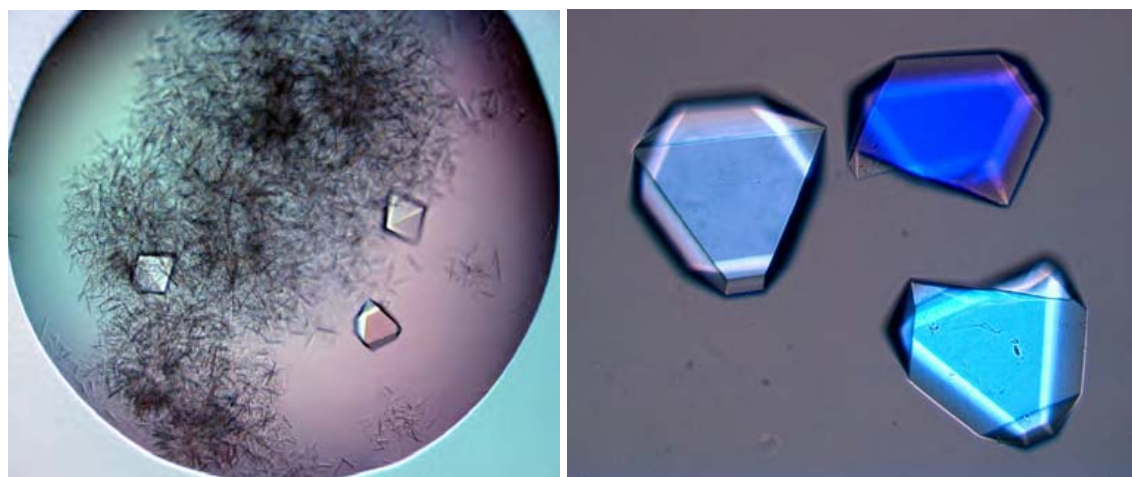


Figure 12. Crystals of rHvExoI. In part A, the crystals that grew after 97 days from the screening plate that initially grew only needles. In part B, the crystals of HvExoI formed after seeding with native ExoI microcrystals.

8. Conclusions and Benefits

8.1. Scientific Conclusions and Benefits

In this project, the structure to function relationships in two kinds of β -linked glucooligosaccharide-hydrolyzing exoglucosidases, which are critical to biomass conversion, were studied. With the increased emphasis on biomass conversion for biofuel production, the mechanism of oligosaccharide binding and specificity and ability to engineer these properties is of increasing interest. The glycoside hydrolase family 1 (GH1) rice Os3BGlu7 structure and those of its complexes with oligosaccharides and a mechanism-based inhibitor were determined. Based on these structures, mutations were made to test the importance of amino acids around the active site for cellooligosaccharide binding and substrate specificity. In addition, we looked at the determinants of specificity differences between rice Os3BGlu7 β -glucosidases, barley rHvBII β -mannosidase and the closely related rice isoenzymes Os3BGlu8 β -glucosidase and Os7BGlu26 β -mannosidase by expressing the other enzymes in the same system as Os3BGlu7 and comparing their activities and substrate specificities. We were also able to crystallize Os7BGlu26 β -mannosidase and calculate a preliminary structure, which showed it is very similar to Os3BGlu7 β -glucosidases, despite their differences in substrate specificity. In addition, we developed a recombinant expression system for the glycoside hydrolase family 3 (GH3) enzyme barley rHvExoI, which produced an enzyme with very similar properties to the native enzyme isolated from seeds. This allowed us to mutate the amino acids around the active site and verify their importance to the catalytic activity of the enzyme. In addition, we were able to develop a methodology for crystallizing the recombinantly produced enzyme and its mutants, thereby facilitating further studies on protein structure-function relationships in this enzyme.

The structures of the rice Os3BGlu7 β -glucosidase and its mutant, Os3BGlu7 E176Q, in complexes with cellotetraose, cellopentaose, and laminaribiose, gave an idea of the interactions between a glucooligosaccharide-hydrolyzing GH1 enzyme and oligosaccharides in a productive position for hydrolysis for the first time. This will allow us to engineer the active sites of GH1 enzymes to allow better hydrolysis of oligosaccharides of a particular length, which may improve their use in biomass conversion in the future. In fact, we have been able to use this information to engineer a β -1,4-linked glucosyl binding site into an enzyme that does not normally hydrolyze cellooligosaccharides (Sansenya, Maneesan and Ketudat Cairns, unpublished). Thus, this project has given us the understanding of protein-carbohydrate interactions that will of use in applications of these enzymes in the near future.

The issue of β -D-glucosidase vs. β -D-mannosidase activities have been explored and the major insight has been how complex the determination of specificity for an apparently simple difference can be. In fact, since it is believed that most β -mannosidases act through a B_{2,5} boat-like transition state, while β -glucosidases act through a ⁴H₃ half-chair-like transition state, due to differences in the stabilities of the respective sugars, it is a critical question whether an enzyme that hydrolyzes both can stabilize two different transition states or transit one of the two sugars through its less preferable transition state. In either case, it is interesting how closely related enzymes have developed different preferences and whether this relates to the transition states in their catalytic reactions. Initially, we had naively assumed, that this preference would be determined by the initial substrate binding by different interactions with the axial (mannoside) vs. equatorial (glucoside) 2-hydroxyl group by differences in the surrounding amino acids. However, inspection of the amino acids making direct interactions with the glycon showed that they are completely conserved between the rice and barley enzymes with preferences for

glucosides and mannosides. So, we tried to mutate peripheral amino acids to see if they would change the preference for glucoside vs. mannoside by secondary or aglycon-binding effects. Although some mutations, such as V246I could change decrease the ratio of hydrolysis of *p*NPMan/*p*NPGLc in rHvBII, no single mutation could turn the β -mannosidase into a β -glucosidase or the rice Os3BGlu7 β -glucosidase into a β -mannosidase. Thus, the determinants of glucoside versus mannoside selection appear to be complex.

To gain further insights into whether these closely related enzymes hydrolyze β -D-mannosides and β -D-glucosides via similar or different transition states, we studied the binding of mannosides and glucosides to the enzyme. To our surprise, the use of *p*NP-thio- β -D-glucoside and *p*NP-thiomannoside as nonhydrolyzable substrate analogues appeared to be inappropriate, since the inhibition by these compounds did not reflect the relative K_m or k_{cat}/K_m values of their corresponding O-glycosides. NMR analysis of their binding, showed that the thioglucoside appears to bind in a skew boat conformation (1S_3 or 3S_5), as expected for the O-glucoside, in rHvBII, but the thiomannoside bound to the same protein in a relaxed chair 4C_1 conformation. Binding to the rice Os3BGlu7 enzyme was very weak and did not seem to distort the normal relaxed conformations. Theoretical calculations showed that the sulfur produced longer bonds at sharper angles, compared to the O linkage, and that though several conformations were possible to dock into the active site, the O and S glucosides and mannosides showed some differences in their preferred conformation to dock into the active site. However, both enzymes tended to show the same preferences. Thus, it seems likely that differences in the catalytic preferences of the two enzymes may reside at the transition states.

To gain more insight into the active site shape and transition state selection between β -glucosidases and β -mannosidases, we are tried to solve the structures of the β -mannosidase and covalent intermediates with mannoside, as well as glucoside, and noncovalent complexes of their mutants. Since we could not produce useful crystals of rHvBII, we instead crystallized rice Os7BGlu26 as an example of a related enzyme with β -mannosidase preference. The structure, again, showed that the active site structures are very similar between the β -D-mannosidase and β -D-glucosidase, although the β -D-mannosidase is perhaps slightly narrower. The noncovalent complex of Os3BGlu7 with glucooligosaccharides showed the substrate glycon was distorted between a 1S_3 skew boat and a 4E envelope or 4H_3 half chair, while the covalent intermediate was in a relaxed chair, consistent with a transfer through a 4H3 or 4E transition state. Preliminary results on a mannoside covalent intermediate also gave the relaxed chair conformation, which may indicate the mannoside follows the same conformational pathway in the β -glucosidase. However, we have yet to determine the structure of the noncovalent mannoside complex or either a covalent or noncovalent complex with Os7BGlu26, since the 2-F-mannoside was found to be rapidly hydrolyzed by this enzyme. In the future, we will also determine the structures with gluco- and manno-imidazoles as possible transition state analogues to see how these are shaped in the active site. The understanding of the use of multiple transition states to hydrolyze different substrates in the same enzyme family or the alternative pathway for mannoside hydrolysis, will increase our ability to engineer the substrate specificities of these enzymes and to develop new processes that utilize them.

Although the GH1 enzymes allow insight into oligosaccharide binding and glycon specificity, the GH3 exohydrolases offer certain advantages in hydrolyzing glucan biomass. Among these are the broad specificity of barley Exo I β -glucan exohydrolase toward different linkages and its ability to release glucosyl residues from polysaccharides. However, in the past, it was not possible to genetically engineer this enzyme or test the

roles of certain residues in the active site, since there was no recombinant expression system available for HvExoI. We were able to develop a *Pichia pastoris* expression system that produced a recombinant protein, rHvExoI, that exhibited properties very similar to the native enzyme. This protein could also be crystallized, using microcrystals of the native protein as seeds, to produce strongly diffracting crystals, thereby allowing mutations to be characterized structurally, as well as functionally. Initial mutagenesis of all the residues in the active site, indicated that most of them are absolutely essential for detectable activity under standard conditions, but some activity remained in the E220A and W434A mutants, which allowed them to be characterized further. The sharpening of the pH optimum and decreased inhibition by the glucono- δ -lactone in the E220A mutation, suggested a role of Glu220 in increasing the pKa or the catalytic acid and binding of the transition state. The W434A mutant resulted in a loss of hydrolysis of polysaccharides, β -1,4-linked and β -1,6-linked oligosaccharides and loss of inhibition by thio-linked gentiobiose (β -1,6-linked disaccharide), suggested a role in the binding of the glucosyl residue in the +1 subsite, which is less critical for β -1,3-linked oligosaccharides. Further characterization of other mutants that may bind to substrates, but not hydrolyze them or be rescued in certain conditions will result in a thorough understanding of the HvExoI mechanism, which has been facilitated by this work. This will help with application and engineering of this enzyme for biomass conversion.

8.2. Other benefits.

This work has contributed to the publication of six research papers and one invited review and to the training of seven graduate students, some of whom have moved on to postdoctoral positions abroad and instructor positions in Thailand. Thus, it has been helpful in the development of Suranaree University of Technology as a research university and to Thailand as a respected source of scientific research and a country with well-trained researchers. The intellectual investigation has also allowed the participants to gain new understandings of the problems being addressed, which can assist with teaching as well as development of future research areas and applications. It has also allowed us to collaborate and share knowledge and abilities with researchers around the world, in order to get new ideas and make the most out of the resources available for research. Thus, this project has provided a strong foundation for further studies and scientific studies in our own lab, in our university, in Thailand and in the world.

8.3. Future benefits.

Since the world is in a crisis for development of new sources of energy sources due to the depletion of fossil fuels and the destruction that their use has caused, the investigation of enzymes involved in degradation of β -glucans, the most predominant form of biomass on the planet is especially of interest. Normally, these compounds are degraded to glucose by a combination of endoglucanases, such as cellulases, exoglucanases or cellobiosidases and β -glucosidases. Of these, the β -glucosidases tend to be limiting in the microbial sources of enzymes typically used for biomass conversion. Thus, developing better enzymes for this purpose is critical, and to engineer the enzymes one must understand the way in which they interact with their substrates. We produced the first structures of a β -glucosidase in complex with a long cellooligosaccharide in which the cellooligosaccharide was properly positioned for hydrolysis. The loose interactions that were seen can serve as a model for fashioning new glucosyl residue binding sites on enzymes to optimize their use in biomass conversion. In addition, both the GH1 and GH3 enzymes can be tested for their abilities to promote breakdown of lignocellulose waste, so that they can be engineered for better use in this application.

The understanding of glycon specificity gained in these studies is also of use for future research projects and protein engineering. The realization that transition state differences are likely to contribute to the substrate specificity will contribute to engineering of enzymes with a wide range of specificities, and perhaps generation of new specificities. It is hoped, therefore, that our specific project will have broad implications for enzyme structure and function relationships. This will allow better interpretation of the likely functions of the millions of genes with sequences determined by the genomic sequencing projects, in terms of the activities of the proteins they encode, and in engineering to develop new functionalities from known enzymes.

9. References:

- Collaborative Computing Project Number 4. (1994) The CCP4 suite: programs for protein Crystallography. *Acta Crystallog. Sect. D* 50, 760-763.
- Chuenchor W, Pengthaisong S, Yuvaniyama J, Opassiri R, Svasti J and 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 inhibitor *Acta Crystallog. Sect. F* 62, 798-801.
- Chuenchor W, Pengthaisong S, Robinson RC, Yuvaniyama J, Oonant W, Bevan DR, 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.
- Chuenchor W, Pengthaisong S, Robinson RC, Yuvaniyama J, Svasti J, Ketudat Cairns JR. 2010. The structural basis of oligosaccharide binding by rice BGlu1 beta-glucosidase. *J. Struct. Biol.* In Press (published online) doi: 10.1016/j.jsb.2010.09.021
- Davies GJ, Ducros VMA, Varrot A, Zechel DL. (2003) Mapping the conformational itinerary of beta-glycosidases by X-ray crystallography. *Biochem. Soc. Trans.* 31, 523-527.
- Hiromi K, Nitta Y, Namura C, Ono S. (1973) *Biochim. Biophys. Acta.* 302 362-375.
- Hrmova M, Harvey AJ, Wang J, Shirley NJ, Jones GP, Stone BA, Høj PB, Fincher GB. (1996) Barley β -D-glucan exohydrolases with β -D-glucosidase activity. Purification and determination of primary structure from a cDNA clone. *J. Biol. Chem.* 271, 5277-5286.
- Hrmova M, Varghese JN, Høj PB, Fincher GB. (1998) Crystallization and preliminary X-ray analysis of beta-glucan exohydrolase isoenzyme ExoI from barley (*Hordeum vulgare*). *Acta Crystallogr. Sect. D* 54, 687-689.
- Ketudat Cairns JR, Esen A. (2010) Beta-Glucosidases. *Cell Mol. Life Sci.* 67, 3389-3405.
- Kuntothom T, Luang S, Harvey AJ, Fincher GB, Opassiri R, Hrmova M, Ketudat Cairns JR. (2009) Rice family GH1 glycosyl hydrolases with β -D-glucosidase and β -D-mannosidase activities. *Arch. Biochem. Biophys.* 491, 85-95.
- Kuntothom T, Raab M, Tvaroska I, Fort S, Pengthaisong S, Cañada FJ, Calle L, Jimenez-Barbero J, Ketudat Cairns JR, Hrmova M. (2010) Binding of β -D-glucosides and β -D-

mannosides by rice and barley β -glycosidases with distinct substrate specificities. *Biochemistry* 49, 8779-8793.

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.

Leah R, Kigel J, Svendsen I, Mundy J. (1995) Biochemical and molecular characterization of a barley seed beta-glucosidase. *J. Biol. Chem.* 270, 15789-15797.

Luang S, Hrmova M, Ketudat Cairns JR. (2010a) High-level expression of barley β -D-glucan exohydrolase HvExoI from a codon-optimized cDNA in *Pichia pastoris*. *Protein Express. Purific.* 73, 90-98.

Luang S, Cairns JRK, Streltsov VA, Hrmova M. (2010b) Crystallisation of wild-type and variant forms of a recombinant plant enzyme β -D-glucan glucohydrolase from barley (*Hordeum vulgare* L.) and preliminary X-ray analysis. *Int. J. Mol. Sci.* 11, 2759-2769.

Morris GM, Goodsell DS, Halliday, RS, Huey, R, Hart WE, Belew RK, Olson AJ. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comp. Chem.* 19, 1639-1662.

Murshudov GN, Lebedev A, Vagin AA, Wilson KS, Dodson EJ. (1999) Efficient anisotropic refinement of macromolecular structures using FFT. *Acta. Crystallogr. Sect. D* 55, 247–255.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. Sect. A* 50, 157-163.

Newman J, Pham TM, Peat TS. (2008) Phoenix experiments: combining the strengths of commercial crystallization automation. *Acta Crystallog. Sect F* 64, 991-996.

Opassiri R, Ketudat Cairns JR, Akiyama T, Wara-Aswapati O, Svasti J, and Esen A. 2003. Characterization of a rice β -glucosidase genes highly expressed in flower and germinating shoot *Plant Sci.* **165**, 627-638.

Opassiri R, Hua Y, Wara-Aswapati O, Akiyama T, Svasti J, Esen A, and Ketudat Cairns JR. 2004. β -Glucosidase, exo- β -glucanase and pyridoxine transglucosylase activities of rice BGlu1. *Biochem. J.* **379**, 125-131.

Opassiri R, Pomthong B, Okoksoong 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: C.W. Carter & R.M. Sweet (Eds.), *Meth. Enzymol.* Vol. 276 Academic Press, New York, pp 307–326.

Pauling L. 1948. *Am. Sci.* 36, 50–58.

Rye, C.S. and Withers, S.G.(2000) Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* **4**, 573-580.

Toonkool P, Metheenukul P, Sujiwattanasat P, Paiboon P, Tongtubtim N, Ketudat-Cairns M, Ketudat-Cairns J, Svasti J. (2006) Expression and purification of dalcocinase, a β -glucosidase from *Dalbergia cochinchinensis* Pierre, in yeast and bacterial hosts. *Protein Expression and Purification* **48**, 195-204.

Varghese JN, Hrmova M, Fincher GB. (1999) Three-dimensional structure of a barley β -D-glucan exohydrolase, a family 3 glycosyl hydrolase. *Structure Fold. Des.* **7**, 179-190.

Verdoucq L, Moriniere 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.

Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765-784.

Wurch T, Lestienne F, Pauwels PJ. (1998) A modified overlap extension PCR method to create chimeric genes in the absence of restriction enzymes. *Biotechnol. Techniq.* **12**, 653-657.

10. Output from grant project

10.1. International Journal Publications

10.1.1. Chuenchor W, Pengthaisong S, Robinson RC, Yuvaniyama J, Oonant W, Bevan DR, Esen A, Chen C-J, Opasiri R, Svasti J, Ketudat Cairns JR. 2008. Structural insights into rice BGlu1 β -glucosidase oligosaccharide hydrolysis and transglycosylation. *Journal of Molecular Biology* **377** (4), 1200-1215. doi: 10.1016/j.jmb.2008.01.076 (ISI IF 4.146, 2008).

10.1.2. Kuntothom T, Luang S, Harvey AJ, Fincher GB, Opasiri R, Hrmova M, Ketudat Cairns JR. 2009. Rice family GH1 glycosyl hydrolases with β -D-glucosidase and β -D-mannosidase activities. *Archives of Biochemistry and Biophysics*. **491**, 85-95. doi:10.1016/j.abb.2009.09.004. (ISI IF 3.046, 2009).

10.1.3. Luang S, Hrmova M, Ketudat Cairns JR. 2010. High-level expression of barley β -D-glucan exohydrolase HvExoI from a codon-optimized cDNA in *Pichia pastoris*. *Protein Expression and Purification* **73**, 90-98. doi: 10.1016/j.pep.2010.04.011 (ISI IF 1.621, 2008).

10.1.4. Ketudat Cairns JR, Esen A. 2010. Beta-Glucosidases. *Cellular and Molecular Life Sciences*. **67**(20), 3389-3405 (Invited review). doi: 10.1007/s00018-010-0399-2 (ISI IF 6.09, 2009). (Selected for journal cover picture).

10.1.5. Luang S; Cairns JRK, Streltsov VA, Hrmova M. 2010. Crystallisation of wild-type and variant forms of a recombinant plant enzyme β -D-glucan glucohydrolase from barley (*Hordeum vulgare* L.) and preliminary X-ray analysis. *International Journal of Molecular Sciences* **11**, 2759-2769. (ISI IF 1.387, 2009).

10.1.6. Kuntothom T, Raab M, Tvaroska I, Fort S, Pengthaisong S, Cañada FJ, Calle L, Jimenez-Barbero J, Ketudat Cairns JR, Hrmova M. 2010. Binding of β -D-glucosides and β -D-mannosides by rice and barley β -glycosidases with distinct substrate specificities. *Biochemistry* **49**, 8779-8793. DOI: 10.1021/bi101112c • Publication Date (Web): 08 September 2010 (ISI IF 3.226, 2009).

10.1.7. Chuenchor W, Pengthaisong S, Robinson RC, Yuvaniyama J, Svasti J, Ketudat Cairns JR. 2010. The structural basis of oligosaccharide binding by rice BGlu1 beta-glucosidase. *Journal of Structural Biology*, **173**, 169-179. (accepted 22 Sept., 2010 by the ISI IF 3.673, 2009).

Cumulative Impact Factor: 23.189

10.2. International Meeting Presentations

10.2.1. Ketudat Cairns, J.R. 2010. Structural studies of the interactions of rice beta-glucosidases with their substrates and reaction intermediates. Institute for Protein Research Seminar on Cooperation in Protein Science between Asian and Pacific Countries. Univ. of Osaka, Osaka, Japan, 14 June, 2010. Invited Lecture.

10.2.2. Sansenya, S., Pengthaisong, S., Chuenchor, W., Seshadri, S., Opassiri, R., Ketudat Cairns, J.R. 2010. Oligosaccharide binding and Hydrolysis by rice beta-glucosidases. Plant Polysaccharide and Applied Glycoscience Workshop 2010. 29-31 July, 2010, Tokyo, Japan. Oral presentation 15.

10.2.3. Ketudat Cairns, J.R., P. Prasert, Pengthaiong, S., Tankrathok, A., Chuenchor, W. 2010. Structural investigation of β -mannosidase activity in plant GH1 glycoside hydrolases. International Carbohydrate Symposium. 1-6 August, 2010. Chiba, Japan. Poster D-P3005.

10.2.4. Kuntothom, T., Raab, M., Fort, S., Pengthaisong, S., Calle, J., Cañada, J., Tvaroška, I., Jimenez-Barbero, J., Ketudat Cairns, J.R., Hrmova, M. 2010. Structural rationale for binding of β -D-glucosides and β -D-mannosides by rice and barley β -D-glucosidases with different substrate specificities. International Carbohydrate Symposium. 1-6 August, 2010. Chiba, Japan. Oral Presentation D-O2-003.

10.2.5. Ketudat-Cairns JR, Cheunchor W, Kuntothom T, Pengthaisong S, Luang S, Hrmova M, Opassiri R, Fincher, GB. 2008. Structural basis for substrate specificity in rice and barley beta-glucosidases. Joint 4th Asian and Oceania Human Proteome Organization Congress (AOHUPO) and 2nd Pacific-Rim International Conference on Protein Science (PRICPS), Cairns, QLD, Australia. 22-26 June, 2008. Abstract 322 (poster).

10.2.6. Ketudat Cairns JR, Chuenchor W. 2008. Assessment of substrate binding with a semi-active mutant of rice BGlu1. The 22nd Symposium of The Protein Society, San Diego, CA, USA 19-23 July, 2008. Abstract 267. *Protein Science* 17 (Suppl. 1), 152-3.

10.2.7. Chuenchor W, Pengthaisong S, Robinson R, Yuvaniyama J, Ketudat Cairns J. 2008. The crystal structures of rice BGlu1 E176Q mutant in complexes with oligosaccharides reveal substrate recognition and dynamics. The 22nd Symposium of The Protein Society, San Diego, CA, USA 19-23 July, 2008. Abstract 345. *Protein Science* 17 (Suppl. 1), 181.

10.2.8. Pengthaisong S, Chuenchor W, Ketudat Cairns JR. 2008. Crystallization of rice BGlu1 β -glucosidase E176Q mutant with oligosaccharide substrates. 21st Congress and General Assembly of the International Union of Crystallography. Osaka, Japan. 23-31 August, 2008. P04.02.133(C272) (presented 28-29 August, 2008).

10.3. National Meeting Presentations

10.3.1. Ketudat-Cairns JR, Opassiri R, Ketudat-Cairns M, Tantanuch W. 2007. Rice beta-glucosidases: from genome annotation to structure-function studies. The 19th Annual Meeting of the Thai Society for Biotechnology, Thammasart University, Pathumthani, Thailand, Oct., 2007. Invited Lecture IV.

10.3.2. Kuntothom T, Luang S, Opassiri R, Hrmova M, Ketudat Cairns JR. 2008. Comparative between rice β -glucosidases and barley β -glucosidase/ β -mannosidase characterization. Third Annual Symposium of the Protein Society of Thailand. Bangkok, Thailand 28-29 August, 2008. Poster 60. Program page 80.

10.3.3. Seshadri S, Opassiri R, Kumprasert B, Akiyama T, Ketudat Cairns JR. 2009. Structure and function of rice Os3BGlu6, a second family 1 glycosyl hydrolase from rice. 2nd Symposium of the Biochemistry and Molecular Biology Section of the Science Society of Thailand. Khon Kaen University, Khon Kaen, Thailand, 7-8 May, 2009. (Invited Lecture).

10.3.4. Pengthaisong S, Kuaprasert B, Svasti J, Ketudat Cairns JR. 2009. Structure determination of rice BGlu1 glycosynthase with and without oligosaccharides. 2nd Symposium of the Biochemistry and Molecular Biology Section of the Science Society of Thailand. Khon Kaen University, Khon Kaen, Thailand, 7-8 May, 2009. Poster 058.

10.3.5. Luang S, Hrmova M, Ketudat-Cairns JR. 2009. Expression of an optimized barley exoglucanase I gene in *Pichia pastoris*. 2nd Symposium of the Biochemistry and Molecular Biology Section of the Science Society of Thailand. Khon Kaen University, Khon Kaen, Thailand, 7-8 May, 2009. Poster 069.

10.3.6. Pengthaisong S, Kuaprasert B, Svasti J, Ketudat-Cairns JR. 2009. Structural analysis of rice BGlu1 glycosynthase with and without oligosaccharides by X-ray crystallography. 4th Annual Symposium of the Protein Society of Thailand. Chulabhorn Research Institute, Bangkok, Thailand. 26-28 August, 2009. Poster 5.

10.3.7. Tankrathok A, Ketudat-Cairns JR. 2009. Improvement of rice Os7BGlu26 β -mannosidase/ β -glucosidase protein expression for purification and preliminary crystallization. 4th Annual Symposium of the Protein Society of Thailand. Chulabhorn Research Institute, Bangkok, Thailand. 26-28 August, 2009. Poster 12.

10.3.8. Kuntothom T, Luang S, Harvey AJ, Fincher GB, Opassiri R, Hrmova M, Ketudat Cairns JR. 2009. Rice family GH1 glycosyl hydrolases with β -D-glucosidase and β -D-mannosidase activities. 2nd Symposium of the Biochemistry and Molecular Biology Section of the Science Society of Thailand. Khon Kaen University, Khon Kaen, Thailand, 7-8 May, 2009. Poster 079.

10.3.9. Luang, S., Hrmova, M., Ketudat Cairns, J. R. 2010 Investigation of active site amino acid residues essential to barley β -D-glucan exohydrolase activity. The 5th Annual Symposium of the Protein Society of Thailand. Chulabhorn Research Institute, Bangkok, 24-26 June, 2010. Oral Presentation 4.

10.3.10. Sansenya, S., Opassiri, R., Ketudat Cairns, J.R. 2010. The crystal structure of Os4BGlu12, a rice β -glucosidase with significant thioglucosidase activity. The 5th Annual Symposium of the Protein Society of Thailand. Chulabhorn Research Institute, Bangkok, 24-26 June, 2010. Poster Presentation 7.

10.3.11. Prasert, P., Ketudat-Cairns, J. 2010. Preliminary structures of rice β -glucosidase BGlu1 E176Q covalent complexes with α -D-mannopyranoside and α -D-2-deoxy-2-fluoromannoside. The 5th Annual Symposium of the Protein Society of Thailand. Chulabhorn Research Institute, Bangkok, 24-26 June, 2010. Poster Presentation 13.

10.3.12. Tankrathok, A., Ketudat Cairns, J.R. 2010. Crystal structure of rice Os7BGlu26 β -mannosidase/ β -glucosidase. The 5th Annual Symposium of the Protein Society of Thailand. Chulabhorn Research Institute, Bangkok, 24-26 June, 2010. Poster Presentation 30

10.4. Graduate Students' Training

- 10.4.1. Mr. Teerachai Kuntothom, Ph.D. Suranaree University of Technology, School of Biochemistry, 2008.
- 10.4.2. Ms. Sukanya Luang, Ph.D. Suranaree University of Technology, School of Biochemistry, 2010.
- 10.4.3. Ms. Salila Pengthaisong, Studying for Ph.D. 2007-present, Suranaree University of Technology, School of Biochemistry.
- 10.4.4. Mr. Anupong Tankrathok, Studying for Ph.D. 2008-present, Suranaree University of Technology, School of Biochemistry.
- 10.4.5. Mr. Akkarawit Prawisut, Studying for Ph.D. 2008-present, Suranaree University of Technology, School of Biochemistry.
- 10.4.6. Ms. Por-ngam Prasert, Studying for M.Sc. 2008-present, Suranaree University of Technology, School of Biochemistry.
- 10.4.7. 10.4.6. Mr. Sompong Sansenya, Studying for Ph.D. May, 2010-present, Suranaree University of Technology, School of Biochemistry.

10.5. International Protein Data Bank (PDB) Entries

- 10.5.1. 3F4V.pdb: Semi-active E176Q mutant of rice BGlu1, a plant exoglucanase/ β -glucosidase; Chuenchor, W., Ketudat Cairns, J.R., Pengthaisong, S., Robinson, R.C., Yuvaniyama, J., Chen, C.-J.
- 10.5.2. 3F5J.pdb: Semi-active E176Q mutant of rice BGlu1, a plant exoglucanase/ β -glucosidase with cellotetraose. Chuenchor, W., Ketudat Cairns, J.R., Pengthaisong, S., Robinson, R.C., Yuvaniyama, J., Chen, C.-J. 2008.
- 10.5.3. 3F5K.pdb: Semi-active E176Q mutant of rice BGlu1, a plant exoglucanase/ β -glucosidase with cellopentaose. Chuenchor, W., Ketudat Cairns, J.R., Pengthaisong, S., Robinson, R.C., Yuvaniyama, J., Chen, C.-J. 2008.
- 10.5.4. 3F5L.pdb: Semi-active E176Q mutant of rice BGlu1, a plant exoglucanase/ β -glucosidase with laminaribiose (partial occupancy). Chuenchor, W., Ketudat Cairns, J.R., Pengthaisong, S., Robinson, R.C., Yuvaniyama, J., Chen, C.-J. 2008.
- 10.5.5. 3AHV.pdb: Semi-active E176Q mutant of rice bglu1 covalent complex with 2-deoxy-2-fluoroglucoside. Chuenchor, W., Pengthaisong, S., Robinson, R.C., Yuvaniyama, J., Svasti, J., Ketudat Cairns, J.R. 2009.

- 10.5.6. 3AHT.pdb: Crystal structure of rice BGlu1 E176Q mutant in complex with laminaribiose. Chuenchor, W., Pengthaisong, S., Robinson, R.C., Yuvaniyama, J., Svasti, J. Ketudat Cairns, J.R. 2009.

10.6. Collaborations

- 10.6.1. Prof. Dr. M.R. Jisnuson Svasti, Assoc. Prof. Dr. Jirundon Yuvaniyama, Assoc. Prof. Dr. Palangpol Kongsaree. Center for Protein Structure and Function, Faculty of Science, Mahidol University, Bangkok, Thailand.
- 10.6.2. Assoc. Prof. Dr. Maria Hrmova & Prof. Geoffrey B. Fincher, Australian Centre for Plant Functional Genomics, University of Adelaide, Waite Campus, Glen Osmond, SA, Australia.
- 10.6.3. Prof. Dr. Jesús Jiménez-Barbero, Centro de Investigaciones Biológicas, CSIC, Madrid, Spain.
- 10.6.4. Assoc. Prof. Dr. Robert C. Robinson, Institute of Molecular and Cell Biology, 61 Biopolis Drive, Proteos, Singapore.
- 10.6.5. Assoc. Prof. Dr. Chun-Jung Chen, National Synchrotron Radiation Research Center, Hsinchu, Taiwan.
- 10.6.6. Prof. Dr. Igor Tvaroška, Department of Structure and Function of Saccharides, Institute of Chemistry, Center for Glycomics, Slovak Academy of Sciences, Bratislava, Slovak Republic.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>

JMBAvailable online at www.sciencedirect.com ScienceDirect

Structural Insights into Rice BGlu1 β -Glucosidase Oligosaccharide Hydrolysis and Transglycosylation

Watchalee Chuenchor¹, Salila Pengthaisong¹, Robert C. Robinson², Jirundon Yuvaniyama³, Worrapoj Oonanant³, David R. Bevan⁴, Asim Esen⁵, Chun-Jung Chen⁶, Rodjana Opassiri¹, Jisnusun Svasti³ and James R. Ketudat Cairns^{1*}

¹Schools of Biochemistry & Chemistry, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand

²Institute of Molecular and Cell Biology, 61 Biopolis Drive, Proteos, Singapore 138673 Singapore

³Department of Biochemistry and Center for Excellence in Protein Structure and Function, Faculty of Science, Mahidol University, Bangkok 10400, Thailand

⁴Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0406, USA

⁵Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0406, USA

⁶Life Science Group, National Synchrotron Radiation Research Center, Hsinchu 30076, Taiwan R.O.C.

The structures of rice BGlu1 β -glucosidase, a plant β -glucosidase active in hydrolyzing cell wall-derived oligosaccharides, and its covalent intermediate with 2-deoxy-2-fluoroglucoside have been solved at 2.2 Å and 1.55 Å resolution, respectively. The structures were similar to the known structures of other glycosyl hydrolase family 1 (GH1) β -glucosidases, but showed several differences in the loops around the active site, which lead to an open active site with a narrow slot at the bottom, compatible with the hydrolysis of long β -1,4-linked oligosaccharides. Though this active site structure is somewhat similar to that of the *Paenibacillus polymyxa* β -glucosidase B, which hydrolyzes similar oligosaccharides, molecular docking studies indicate that the residues interacting with the substrate beyond the conserved -1 site are completely different, reflecting the independent evolution of plant and microbial GH1 exo- β -glucanase/ β -glucosidases. The complex with the 2-fluoroglucoside included a glycerol molecule, which appears to be in a position to make a nucleophilic attack on the anomeric carbon in a transglycosylation reaction. The coordination of the hydroxyl groups suggests that sugars are positioned as acceptors for transglycosylation by their interactions with E176, the catalytic acid/base, and Y131, which is conserved in barley BGQ60/ β -II β -glucosidase, that has oligosaccharide hydrolysis and transglycosylation activity similar to rice BGlu1. As the rice and barley enzymes have different preferences for cellobiose and cellotriose, residues that appeared to interact with docked oligosaccharides were mutated to those of the barley enzyme to see if the relative activities of rice BGlu1 toward these substrates could be changed to those of BGQ60. Although no single residue appeared to be responsible for these differences, I179, N190 and N245 did appear to interact with the substrates.

© 2008 Elsevier Ltd. All rights reserved.

*Corresponding author. E-mail address: cairns@sut.ac.th.

Abbreviations used: ABTS, 2,2'-azino-bis(3-ethylbenzthiazolinesulfonic acid); CC, correlation coefficient; DLS, dynamic light-scattering; DIMBOAGlc, 2-O- β -D-glucopyranosyl-4-hydroxy-7-methoxy-1,4-benzoxazin-3-one; DNP2FG, 2-deoxy-2-fluoro- β -D-glucoside; DP, degree of polymerization; free BGlu1, rice BGlu1 β -glucosidase with no substrate-based inhibitor; G2F, 2-deoxy-2-fluoro- β -D-glucoside; GH1, glycosyl hydrolase family 1; IMAC, immobilized metal affinity chromatography; LGA, Lamarckian genetic algorithm; Mes, 2-morpholinoethanesulfonic acid; NCS, non-crystallographic symmetry; PDB, Protein Data Bank; PEG MME, polyethylene glycol monomethyl ether; *p*NP, *p*-nitrophenol; *p*NPG, *p*-nitrophenyl β -D-glucopyranoside; PpBglB, *Paenibacillus polymyxa* β -glucosidase BglB; RMSD, root-mean-square deviation; SbDhr1, *Sorghum bicolor* β -glucosidase isozyme 1; *V*₀, initial velocity; ZmGlu1, *Zea mays* β -glucosidase isozyme 1.

Received 17 October 2007;
received in revised form
7 January 2008;
accepted 24 January 2008
Available online
4 February 2008

Edited by G. Schulz

Keywords: β -glucosidase; X-ray crystallography; cellooligosaccharides; enzyme-substrate interactions; structure function relationship

Introduction

Cellulose and the other β -glucans of plant cell walls together are the most abundant biopolymers on earth. They serve to define the shape and structural strength of plant cells and act as sources of energy for a variety of organisms, primarily bacteria and fungi. Since the emergence of plants, bacteria and fungi have apparently evolved enzymes to degrade cellulose and other β -glucans, including cellulases and other endoglucanases, cellobiosidases, which cut cellobiose from the non-reducing ends, and cellobiases and other β -glucan β -glucosidases, which remove one glucose unit at a time from the non-reducing end of β -glucan-derived oligosaccharides.¹ Plants are likely to have independently developed these enzymes, which are necessary for cell growth and remodeling of cell walls and for releasing glucose for the metabolic needs of the plant.

Glycosyl hydrolase family 1 (GH1) is a family of enzymes including β -glucosidases involved in β -glucan hydrolysis, as well as many other functions.² GH1 belongs to glycosyl hydrolase clan A, for which the basic structure is a $(\beta/\alpha)_8$ barrel with the conserved glutamate residues at the carboxy-terminal ends of strands 4 and 7 serving as the catalytic acid/base and catalytic nucleophile, respectively.^{3,4} Despite their similar structures, GH1 enzymes have a wide range of specificities, including β -galactosidase, β -mannosidase, β -fucosidase, phospho- β -galactosidase, β -glucuronidase, phospho- β -glucosidase, diglycosidases like primeverosidase, furcatin hydrolase and isoflavone 7-O- β -apiosyl- β -1,6-glucosidase, β -glycosidase, and hydroxyisourate hydrolase, which hydrolyzes a purine ring rather than a glycosidic bond (CAZY $\dagger$).⁵ The GH1 glycosyl hydrolases also display a range of aglycone specificities, with some showing nearly absolute specificity for one sugar and one aglycone, while others accept a range of either glycone or aglycone or both. As such, GH1 provides a useful model for the study of structure-function relationships in enzyme specificity.

To date, coordinates for the structures of 20 GH1 proteins have been deposited in the Protein Data Bank (PDB), three from Archaea, eight from bacteria, two from animals, one from a fungus and six from plants, representing β -glycosidases, a 6-phospho- β -galactosidase, β -glucosidases with var-

ious aglycone specificities and myrosinases (CAZY). In addition, several complexes with inhibitors and mutants are available, which provide insights into the mechanisms of both catalysis and substrate recognition. However, the loops at the C-terminal ends of the β -strands vary greatly in both length and composition resulting in active sites that range from short, narrow slots to deep, wide pockets, and the solution of one enzyme for binding of a certain substrate may not be the same as that of an enzyme from another organism that has evolved separately. Thus, many GH1 enzymes are still under investigation and hopefully will yield new insights into sugar and aglycone binding. Recently, structures for a bacterial cellooligosaccharide-hydrolyzing β -glucosidase⁶ and a fungal β -glucosidase similar to fungal oligosaccharide-degrading enzymes⁷ have been reported, but a GH1 plant β -glucan β -glucosidase structure has not been described.

Rice (*Oryza sativa* L.) BGlu1 β -glucosidase (also called Os3bglu7) is a GH1 enzyme widely expressed in rice tissues, especially flower and seedling shoot, that has a high level of activity toward β -glucan-derived oligosaccharides.^{8–10} It is over 66% identical in sequence with barley (*Hordeum vulgare* L.) BGQ60 (also called β -II) β -mannosidase/ β -glucosidase (Fig. 1) and falls in the same phylogenetic cluster, but has significantly different substrate specificity.¹¹ While both enzymes hydrolyze short 1,3-linked oligosaccharides with degrees of polymerization (DP) of 2 or 3 and long (DP 4–6) cellooligosaccharides (β -1,4-linked glucose polymers) with at least six apparent glucosyl residue-binding subsites, they show distinct preferences for short cellooligosaccharides.^{8,9,12} Barley BGQ60 prefers to hydrolyze cellobiose over cellotriose, with an apparent disruptive interaction at the third glucosyl residue of cellotriose in the +2 subsite, while rice BGlu1 hydrolyzes cellobiose very poorly and has a high level of affinity for the third glucosyl residue of cellotriose at the +2 site. The BGlu1 enzyme can also catalyze transglycosylation reactions and, interestingly, a glycosyl synthase mutant of BGlu1 can synthesize long 1,4-linked oligosaccharides, but only with acceptors that occupied at least three subsites,¹³ which is consistent with stronger binding at subsites +2 and +3 compared to +1.

In order to determine the structural basis of oligosaccharide-binding by rice BGlu1 exo-glucanase/ β -glucosidase, we crystallized the protein and initiated diffraction trials.¹⁴ Here, we report

$\dagger$ www.cazy.org

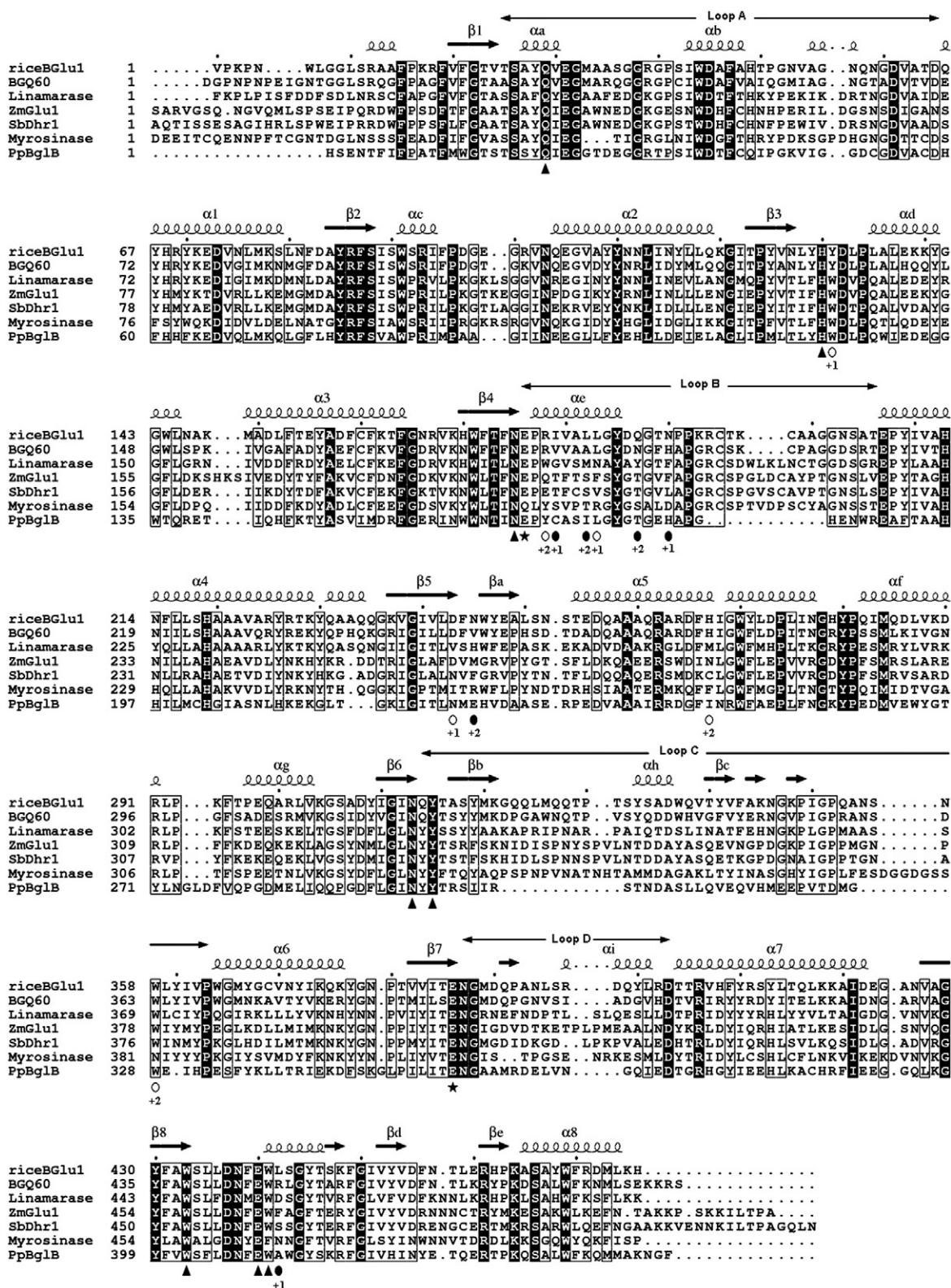


Fig. 1. Amino acid sequence alignment of rice BGlul1, BGQ60: barley β -glucosidase BGQ60, linamarase: cyanogenic β -glucosidase from *Trifolium repens* (Genbank accession 1CBG), ZmGlul1: *Zea mays* β -glucosidase isozyme I (1E1E_A), SbDhr1: dhurrinase isozyme I from *Sorghum bicolor* (1V02_A), myrosinase: *Sinapis alba* myrosinase (1MYR), and PpBglB: *Paenibacillus polymyxa* β -glucosidase BglB (2JIE_A). The alignment was generated with ClustalW, and the secondary structure elements of rice BGlul1 were aligned at the top with the ESPrpt program.⁴¹ Stars indicate acid/base and nucleophilic catalytic residues, black arrowheads mark the amino acids in close contact with the glucose residue at subsite -1, and ovals highlight the amino acids predicted to be aglycone binding residues with indication of subsites +1 and +2 below the sequence. Aglycone-binding residues of rice BGlul1 that different from those of BGQ60 are presented as black ovals.

the structures of BGLu1 with no ligands added (free BGLu1) and BGLu1 in complex with a covalent 2-deoxy-2-fluoroglucoside inhibitor (G2F). Computational docking studies were done to identify the likely position for cellobiose binding in the active site. Potential oligosaccharide-binding residues were mutated to those of barley BGQ60 and the activities of the mutants were characterized.

Results and Discussion

Overall structure and quality of models

Crystals of free BGLu1 enzyme and BGLu1 bound to G2F are isomorphous, and belong to the orthorhombic $P2_12_12_1$ space group. The free BGLu1 structure was refined against 2.20 Å resolution data to yield a final model with an R -factor of 16.2% and an R_{free} of 20.5% (Table 1). The structure of the G2F complex was finally refined with 1.55 Å data to give a model characterized by an R -factor of 17.1% and an R_{free} of 19.7% (Table 1).

The overall structure of rice BGLu1 was similar to other GH1 structures, with a $(\beta/\alpha)_8$ or TIM barrel consisting of a core of eight twisted parallel β -strands connected by long loops and eight α -helices that form the outer layer of the core structure.¹⁵ The active site at the carboxyl ends of the β -strands has an approximate volume of 123.5 Å³. The catalytic acid/base, E176, and nucleophile, E386, are located at the ends of β -strands 4 and 7 on opposite sides of a cleft at the bottom of the active site. The two molecules in the asymmetric unit are related by 2-fold symmetry, and are linked by a single metal ion (Fig. 2a).

The quality of electron density at the four variable loop regions, including residues T25–D65 of loop A, P177–T206 of loop B, Q314–P363 of loop C, and N387–D403 of loop D, was almost as good as in the core region for both the main chain and the side chains. However, the densities of a few side chains that are fully exposed to the solvent were unclear. The poorest density was observed in loop C, where some residues, especially Q323, Q324, and Q327, had ambiguous side-chain densities, but the main-chain density was clear and continuous. Linking C195 and C198 in loop B is a disulfide bond that is conserved in plant GH1 enzymes. Two *cis*-peptide bonds were found between P191 and P192, and between W433 and S434. The latter non-proline *cis*-peptide bond occurred after W433, which is a residue crucial to binding of the glycone at subsite –1. The *cis*-peptide bonds are found at the conserved amino acids in the 1CBG template at these same positions. Although N-terminal amino acid sequencing of the first eight residues of the crystallized protein gave the sequence AMADVVPK, corresponding to five residues of the linker to the fusion protein and the first three residues of BGLu1, no electron density was observed before BGLu1 W6. Therefore, it is likely that the N terminus, which is exposed to the solvent, has great

Table 1. Data collection, processing and refinement statistics

	Free enzyme (2RGL)	Inhibitor complex (2RGM)
A. Data collection and processing statistics		
Beamline	BL13B1	BL13B1
Wavelength (Å)	0.98	0.98
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)		
<i>a</i>	80.05	80.29
<i>b</i>	100.67	101.29
<i>c</i>	127.02	127.25
Resolution range (Å)	30–2.20	30–1.52
Resolution range outer shell (Å)	2.28–2.20	1.57–1.52
No. unique reflections	53,089	157,799
No. observed reflections	363,672	491,141
Averaged multiplicity	6.9 (6.8)	3.1 (2.0)
Completeness (%)	99.9 (99.5)	98.8 (96.3)
R_{merge} (%)	13.2 (40.4)	5.2 (40.3)
$\langle I/\sigma(I) \rangle$	15.5 (5.2)	17.5 (2.29)
B. Refinement and final model statistics		
PDB accession code	2RGL	2RGM
Resolution used in refinement (Å)	29.7–2.20	28.2–1.55
No. reflections		
Working set	50,374	141,569
Test set	2652	7487
NCS restraints		
Positional, RMS (Å), weight	0.02, 0.05	0.05, 5.0
Thermal, RMS (Å ²), weight	0.06, 0.50	0.59, 10.0
R_{factor} (%)	16.2	17.1
R_{free} (%)	20.5	19.7
No. of residues in protein ^d	952 (2 molecules)	952 (2 molecules)
No. protein atoms	7602 (2 molecules)	7602 (2 molecules)
No. water molecules	1187	1352
No. ligand atoms	None	22
No. other hetero-atoms	47	47
Mean B -factor		
Protein (Å ²)	10.43	10.74
Nonsolvent	24.81	14.21
heteroatoms (Å ²)		
Solvent (Å ²)	23.40	27.77
r.m.s. bond deviation (Å)	0.010	0.011
r.m.s. angle deviations (deg.)	1.308	1.402
Ramachandran plot		
Residues in most favorable regions (%)	89.0	89.0

^a $R_{\text{merge}} = (\sum |I - \langle I \rangle| / \sum \langle I \rangle)$.

^b $R_{\text{factor}} = (\sum |F_o| - |F_c| / \sum |F_o|)$.

^c Based on 5% of the data.

^d The protein structures cover residues 6–476 of the mature rice BGLu1 protein sequence.

flexibility in this crystal form. One sulfate ion and one Mes molecule from the precipitant solution were associated with each monomer in the asymmetric unit, and a single zinc ion was shared between molecules A and B (Fig. 2a).

Monomer and crystal contacts mediated by zinc

Rice BGLu1 is a monomeric protein in solution,⁸ whereas other plant GH1 enzymes for which structures have been solved are dimeric, for the white

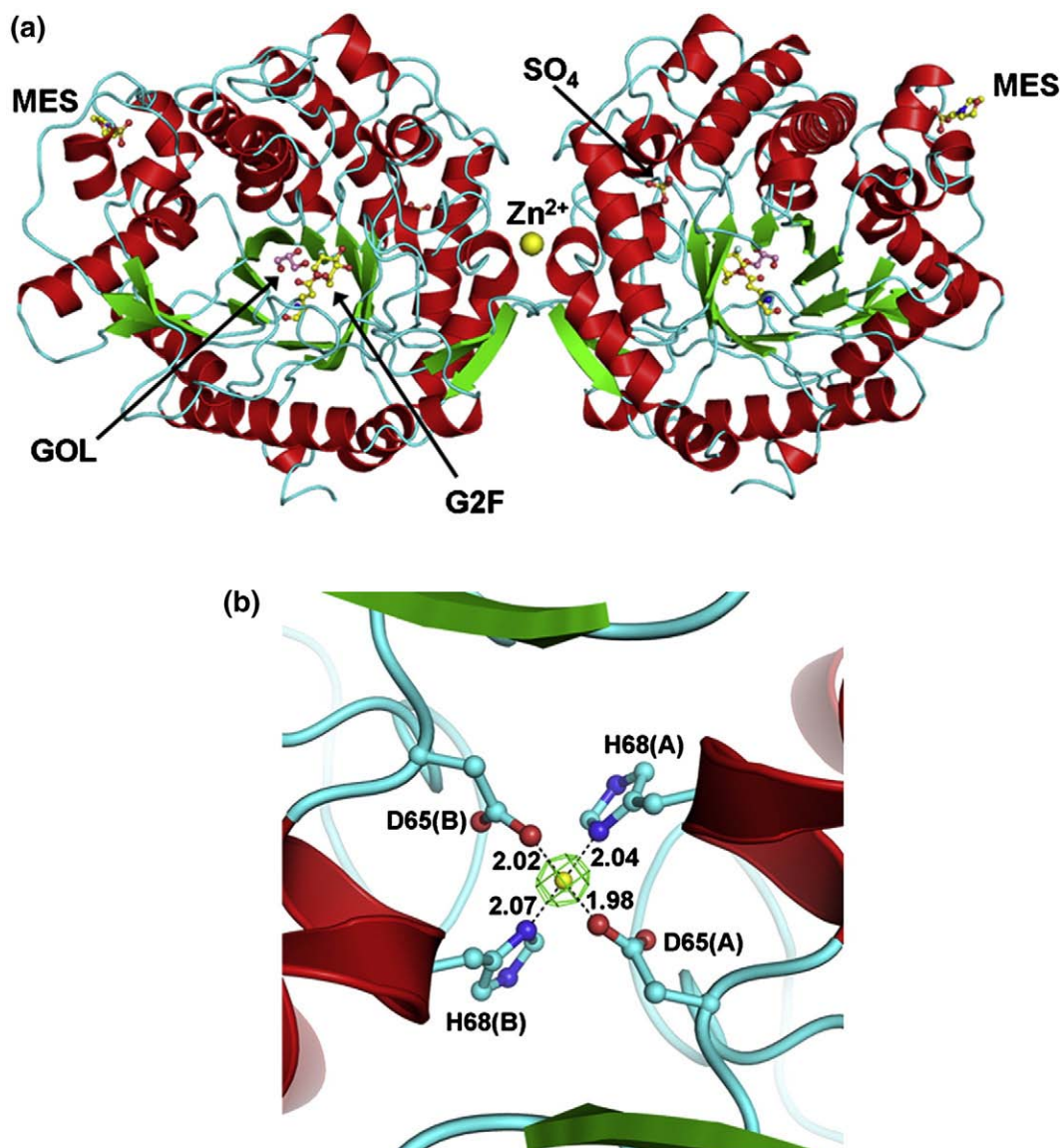


Fig. 2. Asymmetric unit of the rice BGlu1/G2F complex structure. (a) A ribbon diagram representing the overall structure of the BGlu1/G2F inhibitor complex asymmetric unit. The β -strands are colored green, α -helices red and loops cyan. The nucleophilic catalytic residue E386, which is covalently bound to the G2F inhibitor is shown as ball and stick and colored by atoms with carbon in yellow. Hetero atoms found in the crystal structure: GOL (glycerol), MES, and SO_4 are drawn as ball and stick (colored by atoms) and a yellow sphere represents Zn^{2+} . (b) Crystallographic contacts mediated by Zn^{2+} . The two protein molecules in the asymmetric unit are linked by D65 and H68 (ball and stick, with nitrogen blue and oxygen red) from molecule A and B via Zn^{2+} (yellow sphere). Broken lines represent the chelating interactions with their distances given in ångström units.

clover (*Trifolium repens*) cyanogenic β -glucosidase (linamarase),¹⁵ white mustard (*Sinapsis alba*) myrosinase,^{16,17} maize (*Zea mays*) β -glucosidase isozyme I (ZmGlu1),^{18,19} and sorghum²⁰ (*Sorghum bicolor*) dhurrinase isozyme 1 (SbDhr1), or hexameric for wheat (*Triticum aestivum*) β -glucosidase.²¹ However, in the rice BGlu1 crystal structures, two molecules related by a non-crystallographic 2-fold axis were linked through a single zinc ion, which is bound symmetrically between the D65 and H68 residues of the two molecules (Fig. 2b). The Zn^{2+} is well defined with temperature (*B*) factors of 7.36 Å² in the free BGlu1 structure and 7.07 Å² in the BGlu1–G2F complex, and a peak at the Zn^{2+} absorbance edge

was observed in an X-ray absorbance/fluorescence scan, while no peak was seen at the edge of Co^{2+} , which might be left from immobilized metal affinity chromatography (IMAC) purification (data not shown), thereby confirming the identification of the metal ion. The Zn^{2+} was evidently left from seeding crystal conditions, or from contamination in the chemicals or water used for purification or crystallization, since no Zn^{2+} was used in the current crystal growth conditions. The presence of Zn^{2+} was reported to stabilize a dimer in the white mustard myrosinase structure, in which zinc was similarly bound between His and Asp residues, H56 and D70 from each myrosinase molecule.¹⁶ However, the

interfacial buried surface area in the BGlu1 structure is quite small ($270 \text{ \AA}^2$, which is approximately 1.7% of the monomer surface) and few direct interactions linking the dimers are seen, aside from those through the zinc ion. In addition, 95% of the BGlu1 protein eluted as monomer during gel-filtration, dynamic light-scattering (DLS) gave a single peak at a diameter of 7.115 nm, corresponding to the expected size of the monomer, and the enzyme activity was neither inhibited by 10 mM EDTA nor stimulated by Zn^{2+} . None-the-less, the PISA program²² predicted that the complex of Zn^{2+} with the two monomers would be stable in solution with an energy of -43.2 kcal/M , the bulk of which (-38.9 kcal/M) is contributed by interactions between the zinc ion and the monomers. So, the protein may form dimers in the presence of Zn^{2+} in solution, but the monomer appears to be equally active and the functional relevance of any such dimerization is not clear.

Overall structure comparison of rice BGlu1 with other structures in GH1 family

The known 3D structures of the GH family 1 white clover linamarase (1CBG),¹⁵ ZmGlu1 (1E1E),¹⁹ SbDhr1 (1V02),²⁰ white mustard myrosinase (1MYR),¹⁶ and the bacterial *Paenibacillus polymyxa* β -glucosidase BglB (PpBglB, 2JIE)⁶ were superimposed on the rice BGlu1 structure for comparison (Fig. 3). The 1CBG, 1E1E and 1V02 structures represent plant O-glucosidases, 1MYR is a plant

S-glucosidase and 2JIE is a bacterial enzyme that hydrolyzes cellooligosaccharides with high degrees of polymerization as substrates, similar to BGlu1. These structures share amino acid sequence identity of 37–48% with BGlu1. The C^α atoms of molecule A from the 1CBG, 1E1E, 1V02, 1MYR, and 2JIE structures superimposed on BGlu1 with RMSD of 0.80 $\text{\AA}$, 0.96 $\text{\AA}$, 0.91 $\text{\AA}$, 0.96 $\text{\AA}$, and 1.44 $\text{\AA}$, respectively (based on the overlap of 424, 412, 411, 399, and 375 C^α atoms, respectively). The 1CBG structure, which was used as the search model had the greatest sequence identity (48%) and the lowest rmsd values of the overall structure and loops. Inversely, 2JIE, the bacterial PpBglB structure, had the largest differences from our structure.

Structural superimposition showed that the core $(\beta/\alpha)_8$ structures of the GH1 structures are similar in shape and size, while the largest differences are seen in the loop regions surrounding the active site. Four variable loop regions have been described that connect the β -strands to α -helices of the $(\beta/\alpha)_8$ -barrel: loop A (loop 1) between $\beta 1$ and $\alpha 1$, loop B (loop 4) between $\beta 4$ and $\alpha 4$, loop C (loop 6) between $\beta 6$ and $\alpha 6$, and loop D (loop 7) between $\beta 7$ and $\alpha 7$.^{15,16,23} Loop A is little different in length or structural alignment for the six structures, though it contains the residues involved in Zn^{2+} binding between monomers in BGlu1 and myrosinase. Loop B contains the conserved disulfide bond in plant GH1 enzymes and contributes part of the aglycone binding pocket. The hairpin structure closed by

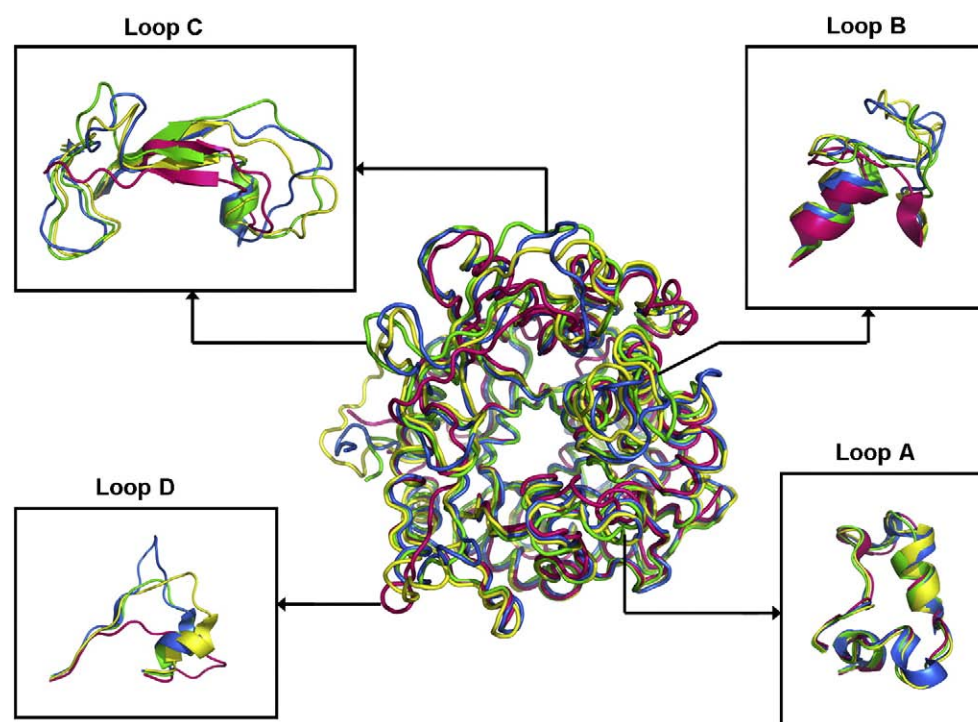


Fig. 3. Superimposition of the structures of rice BGlu1 (green), linamarase from *Trifolium repens* (1CBG, yellow), *Zea mays* β -glucosidase isozyme I (1E1E, blue), and *Paenibacillus polymyxa* β -glucosidase BglB (2JIE, pink). Loops A–D, which constitute the doorway to the active site are expanded to the side, as indicated by the arrows, to show the differences in loop structures. *Sorghum bicolor* dhurrinase isozyme I (1V02) and *Synapis alba* myrosinase (1MYR) were similarly superimposed, but are not shown for the sake of clarity.

this disulfide bridge is several residues shorter in rice BGlu1 than in other plant enzymes. The largest differences are in loop C, as seen in the sequence (Fig. 1) and structural comparisons (Fig. 3). This is the largest loop and it covers almost a half of the entrance to the active site and forms part of the aglycone-binding site. In comparison to rice BGlu1, two extra residues are observed in ZmGlu1, SbDhr1, and myrosinase, and another insertion of six extra amino acid residues is found only in myrosinase. The shorter loop C length in rice BGlu1 contributes to its more open and extended active site, which is discussed below. This loop (D342-Y357) was also reported to form a dimer interface in the ZmGlu1 structure.¹⁹ Finally, loop D, which is the smallest loop (12–16 amino acids), also showed great variation, but it is not close to the substrate-binding site.

The loop structure of PpBglB is extremely different from that of the GH1 plant enzymes. A disulfide linkage is found at loop A instead of loop C. Both loops B and C are much shorter than in the plant enzymes, but their folds are similar, except for the absence of the long coiled structures linking the short β -strands and α -helices. The shorter loops B and C in BGlu1 and PpBglB provide for an open active site, which is compatible with their activity on long cellooligosaccharides. However, the structure of loop D in PpBglB is different from BGlu1 and the other plant enzymes.

Enzyme active site

Rice BGlu1 has a deep, narrow and straight binding cleft, which is fully encircled by negatively charged residues around the deepest and narrowest part, with a broader gate (10 Å $\times$ 23 Å), the top of

which has positively charged residues along one side (Fig. 4a). These features are suitable for binding long oligosaccharide chains, the large number of hydroxyl groups of which can form hydrogen bonds with the polar residues. Several aromatic residues are also aligned along the path from the bottom to the entrance of the active site to form stacking interactions with the glucopyranoside rings of oligosaccharides binding in the active site (Fig. 4b). Although the distance of 18 Å from W433 at the bottom of the active site to the surface is within the range of other plant GH1 enzymes with shorter substrates (from 14.4 Å in ZmGlu1 to 21 Å in clover linamarase), a groove along one side of the active site extends 35 Å from W433 to Q328, which is suitable for binding long oligosaccharide substrates of up to six residues.⁹ The active site geometry is consistent with previous kinetic data that rice BGlu1 hydrolyzes β -1,4-linked oligosaccharides of two to six glucose moieties but not β -1,3-linked oligosaccharides with DP > 3, since their more twisted shape may not fit in the narrow cleft at the inside of the active site.

The covalently bound 2-deoxy-2-fluoro-glycoside inhibitor (G2F) in the complex structure showed strong electron density, similar to that found at surrounding active site residues. A standard relaxed ⁴C₁ chair conformation of the sugar ring fit well into the density at subsite -1, which is formed by Q29, H130, Y131, E175, E176, Y315, E386, W433, E440, W441, and W433, with an average *B*-factor for G2F of 7.43 Å². The overall conformation of free BGlu1 and its G2F complex are very similar, with rmsd of 0.093 Å in molecule A and 0.096 Å in molecule B (calculated over 434 residues each) and 0.121 Å for both molecules (881 residues). The few differences

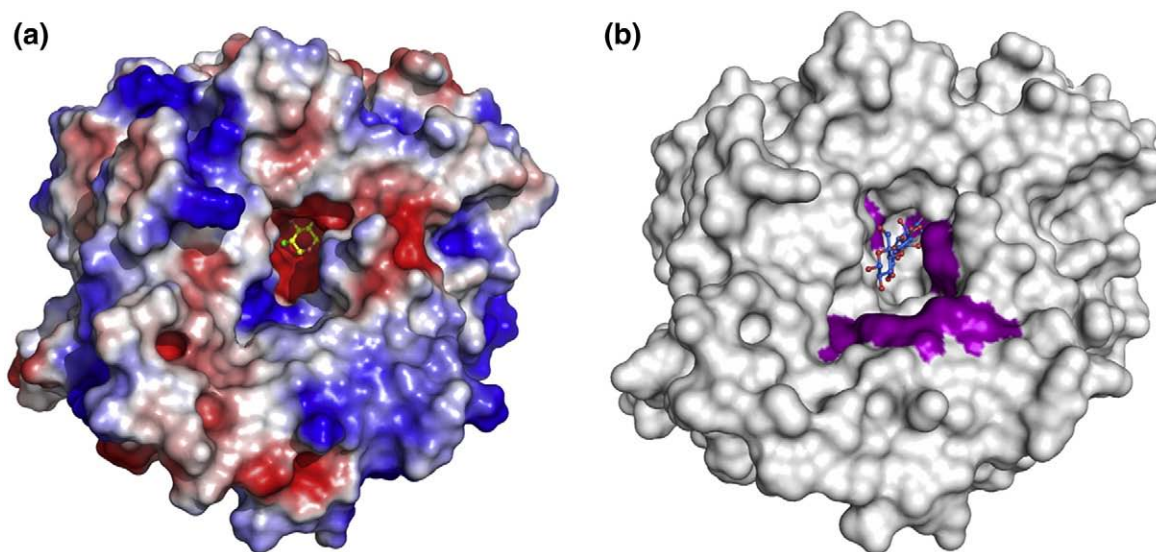


Fig. 4. Molecular surface of the G2F inhibitor complex and free structure with a docked cellotriose. (a) The G2F complex structure is shown as an electrostatic surface with positively charged, negatively charged, and neutral regions colored in blue, red and white, respectively. The G2F inhibitor (drawn with carbon yellow, oxygen red and fluorine green) is bound at the deepest part of the active site pocket. (b) The surface structure of the BGlu1 structure with the docked cellotriose (carbon blue, oxygen red) drawn inside the pocket and the aromatic amino acids that can form sugar-binding platforms lining the active site colored purple.

that are seen are mainly at the active site pocket. In total, five hydrogen bonds were formed between G2F and amino acids (Fig. 5a). The O3 atom of G2F is hydrogen bonded to W441 N $^{\epsilon 1}$, and Q29 O $^{\epsilon 1}$, while the O4 atom forms hydrogen bonds with Q29 N $^{\epsilon 2}$ and E440 O $^{\epsilon 1}$, and O6 interacts with O $^{\epsilon 2}$ of E440. A hydrophobic stacking interaction between glucose and the indole ring of W433 is conserved with other GH1 β -glucosidases. The nucleophile E386 is rotated $\sim 60^\circ$ around the C $^\alpha$ to C $^\beta$ bond with respect to the free enzyme in order to form the glycosyl-enzyme complex (Fig. 5b). A slight rotation of the 2-fluoroglucose-bound nucleophile was also observed in the *Thermotoga maritima* β -glucosidase, where it

was noted to avoid a clash between the nucleophile carboxyl and the 2-fluoro group.²⁴ The distances of the E386 O $^{\epsilon 1}$ and O $^{\epsilon 2}$ to the E176 O $^{\epsilon 2}$ are 3.27 Å and 4.22 Å, respectively, which are closer than in the free enzyme (4.04 Å and 4.67 Å). Y315, which is found within hydrogen bonding distance of the O $^{\epsilon 1}$ of E386 (2.60 Å) in the free enzyme, is shifted away in the G2F complex, and its hydroxyl group forms a hydrogen bond with O5 of G2F (2.90 Å) instead. The O $^{\epsilon 1}$ instead hydrogen bonds to a water molecule that is also hydrogen bonded to N175 O $^{\delta 1}$.

A glycerol molecule was observed with its 2-OH group at a distance of 3.37 Å from the anomeric carbon, 2.68 Å from O $^{\epsilon 1}$ and 3.36 Å from O $^{\epsilon 2}$ of E176,

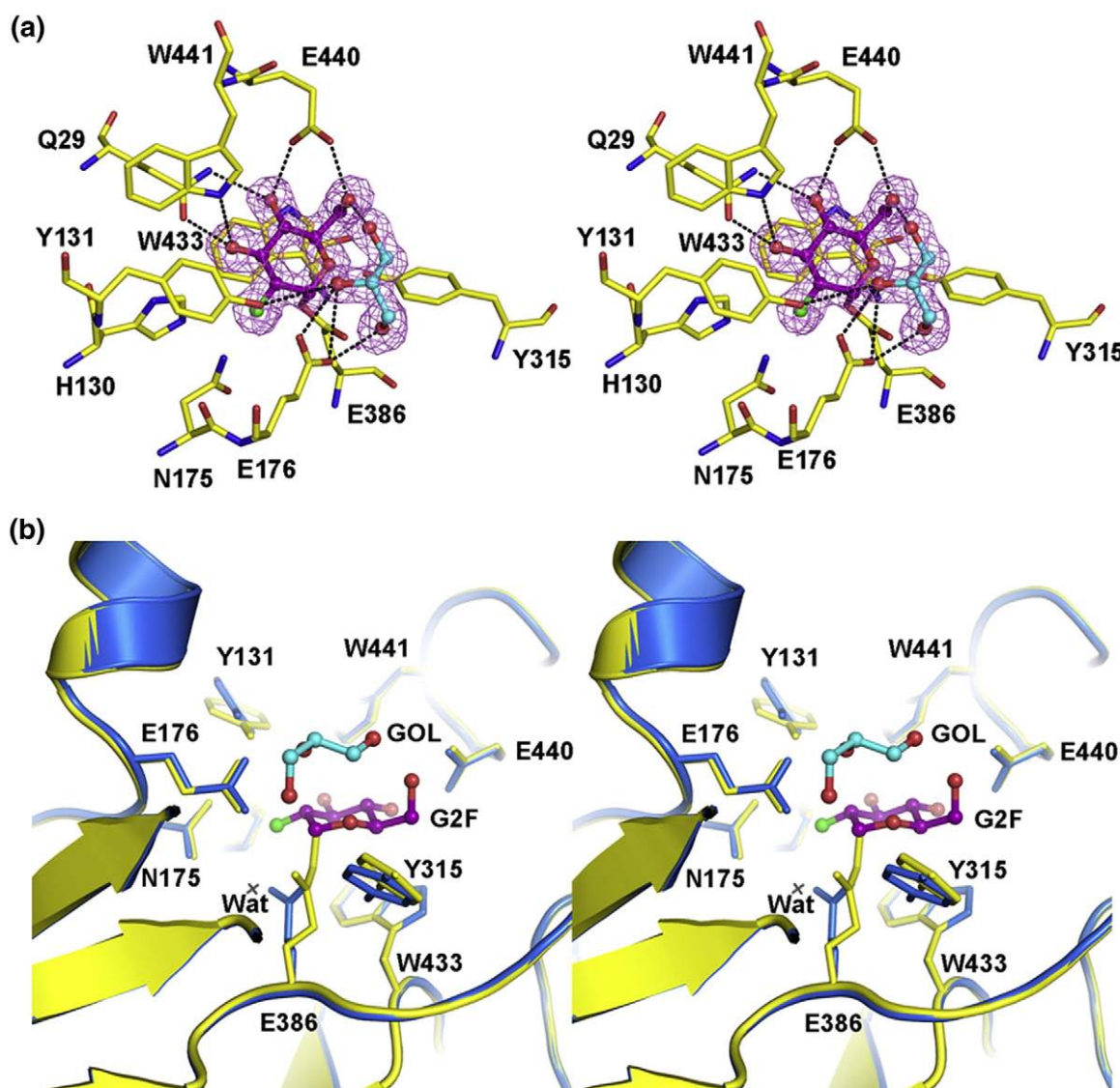


Fig. 5. Stereoview of the active site of rice BGlu1. (a) Protein-ligand interactions in the active site of the BGlu1-G2F complex. The amino acid residues surrounding G2F are presented in stick representation, with carbon yellow, nitrogen blue, oxygen red, and fluoride green. G2F and glycerol (GOL) are drawn in ball and stick representation in the same colors, except carbon atoms are purple and cyan, respectively. The purple $|F_o| - |F_c|$ omit map for G2F and GOL is shown contoured at 3 σ . Hydrogen bonds between the protein and the glycone at subsite -1 and glycerol at subsite +1 are drawn as broken black lines. (b) Superimposition of subsite -1 of the active site of free rice BGlu1 (blue) and covalently bound G2F complex (yellow). The covalently bound G2F and solvent glycerol from the G2F complex are drawn as ball and stick with carbon colored purple and cyan, respectively. The water A1755 from the G2F complex (Wat), which was hydrogen bonded with E386 (the catalytic nucleophile) and N175, is displayed as a cross.

the catalytic acid/base residue, and 3.42 Å from the Y131 OH oxygen, which is twisted slightly compared to its position in the free enzyme (Fig. 5b). The glycerol 1-OH was also within hydrogen bonding distance of both E176 and E386, but was further (3.86 Å) from the anomeric carbon. This glycerol molecule, which is well defined with an average *B*-factor of 14.21 Å², seems to be positioned for deprotonation by the acid/base residue for nucleophilic attack at anomeric carbon in the deglycosylation step, which would lead to transglycosylation. This is consistent with the high level of transglycosylation activity of the enzyme with oligosaccharides at relatively low concentrations,⁹ since the glycerol is equivalent to half a hexose sugar and the coordination of the 1-OH positions the 2-OH group in a way that could act in hexopyranose sugars as well.

The covalent intermediate of the double displacement mechanism of retaining glycosyl hydrolases may be released by either hydrolysis, if a water molecule displaces the enzyme from the anomeric carbon, or by transglycosylation, if an alcohol or other nucleophile displaces it.¹² As far as we know, the BGlu1–G2F complex is the first structure showing a putative transglycosylation complex of a glycosyl hydrolase bound to G2F, but a water molecule is in a position similar to that between the anomeric carbon and an ascorbate that takes the place of the catalytic acid/base in the 1E73 myrosinase structure, which is also a G2F intermediate (Fig. 6).¹⁷ In that case, the enzyme was trapped in the process of hydrolysis, instead of the transglycosylation complex formed with the glycerol in the current structure. In the myrosinase complex, the water approaches from a slightly different angle than that of the glycerol 2-OH in the BGlu1–G2F complex structure, but both are nearly opposite (162° for

BGlu1 and 170° for myrosinase) the bond from the anomeric carbon to O⁶² of the catalytic nucleophile (E386 in BGlu1), thereby avoiding clashes with the G2F C2 (84° from the line between GOL O2 and G2F C1), O5 (80°) and H1 (48°) that are tetrahedrally arranged around C1. The approach of this glycerol molecule appears to be more affected by the steric properties of these surrounding groups than the smaller water molecule in myrosinase, which approaches the G2F closer to C2 (65°) and O5 (58°) and further from H1 (89°). So, the small differences seen in the relative positions of the glucose and acceptor between the putative hydrolysis and transglycosylation complexes can be explained by the sizes of the acceptor molecules.

Docking studies

Until now, the crystal structure of a plant GH family 1 β -glucan exo- β -glucosidase in complex with β -1,4-linked oligosaccharides has not been characterized. In order to find the positions of the β -1,4-linked glucosyl residue subsites, automated docking was used to predict where cellotriose would bind in the rice BGlu1 active site. However, cellotriose with a standard relaxed chair conformation (⁴C₁) at the non-reducing end was found to dock into the active site of rice BGlu1 with its non-reducing glucosyl residue either flipped 180° from the G2F position in the complex structure or far from the catalytic amino acids.

The structure of sorghum dhurrinase E189D (acid/base mutant) bound to dhurrin shows the glucose residue in the ¹S₃ skew boat conformation in the Michaelis complex.²⁰ Molecular dynamics simulations suggest that this adaptation of the sugar conformation at subsite –1 from a ⁴C₁ relaxed chair to a ¹S₃ skew boat conformation is likely to occur

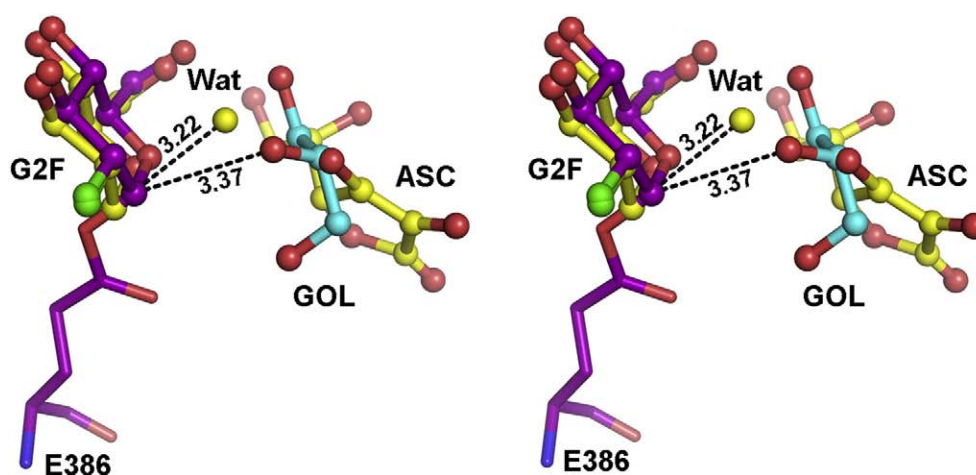


Fig. 6. Superimposition of the active site of BGlu1–G2F complex with the myrosinase–G2F–ascorbate complex structure 1E73. All ligands are drawn as ball and stick, the BGlu1 E386 nucleophilic residue as a stick, and water (Wat) as a non-bonded sphere. The carbon atoms of the covalently bound G2F and solvent glycerol (GOL) from the BGlu1–G2F complex are colored purple and cyan, respectively, while Wat and carbon of the covalently bound G2F and ascorbate (ASC) of the myrosinase complex are colored yellow. Broken lines show the distances (in ångström units) from the anomeric carbon of the covalently bound G2F to the glycerol in the BGlu1–G2F complex and to a water molecule in the myrosinase complex.

upon binding to the enzyme.²⁵ A chair conformation is then recovered after release of the aglycone. Thus, cellotriose with a skew boat conformation at the non-reducing end was docked to the active site, and it docked with the non-reducing glucosyl residue nearly at the same position as the 2-fluoroglucosyl residue at subsite -1 in the covalent complex structure with an rmsd between the two rings of 0.90 Å. The lowest total docking energy of -8.20 kcal/mol was obtained for a cellotriose structure in the highest ranked cluster, which contained the lowest energy conformation of 33 out of 100 runs and had a mean docking energy of -6.65 kcal/mol. The second-ranked cluster (which accounted for 21 of the lowest-energy conformations and had a lowest and mean binding energies of -7.85 and -6.52 kcal/mol) was positioned similarly, but was twisted slightly compared to the first-ranked result. Therefore, the lowest energy structure, shown in Fig. 7a, is likely to indicate the general positions of binding for glucosyl residues at subsites +1 and +2.

Surprisingly few hydrogen bonds were likely between the enzyme and the docked cellotriose at subsites +1 and +2, while subsite -1 was strongly stabilized by four of the five hydrogen bonds seen in the BGlu1-G2F complex, and additional interactions of O2 with E386 O ϵ 2 (2.06 Å) and N175 N δ 2 (3.22 Å), not seen with G2F due to the F in that position.²⁴ At subsites +1 and +2, E176, R178, Q187, N190, N245, T316, and H267, are within range to make direct or water-mediated hydrogen bonds with Glc2 and Glc3. Non-polar interactions could occur between the sugars and I179, L182, L183, L442 and D243 (in which the fully charged side chain is turned away from the substrate). Moreover, both Glc2 and Glc3 at subsites +1 and +2 could have stacking interactions with the aromatic ring of W358, which is a conserved Trp shown to position the aglycone of other plant GH1 β -glucosidases.^{18,20,26}

It should be noted that many of the hydrogen bonding interactions with the sugar hydroxyl groups

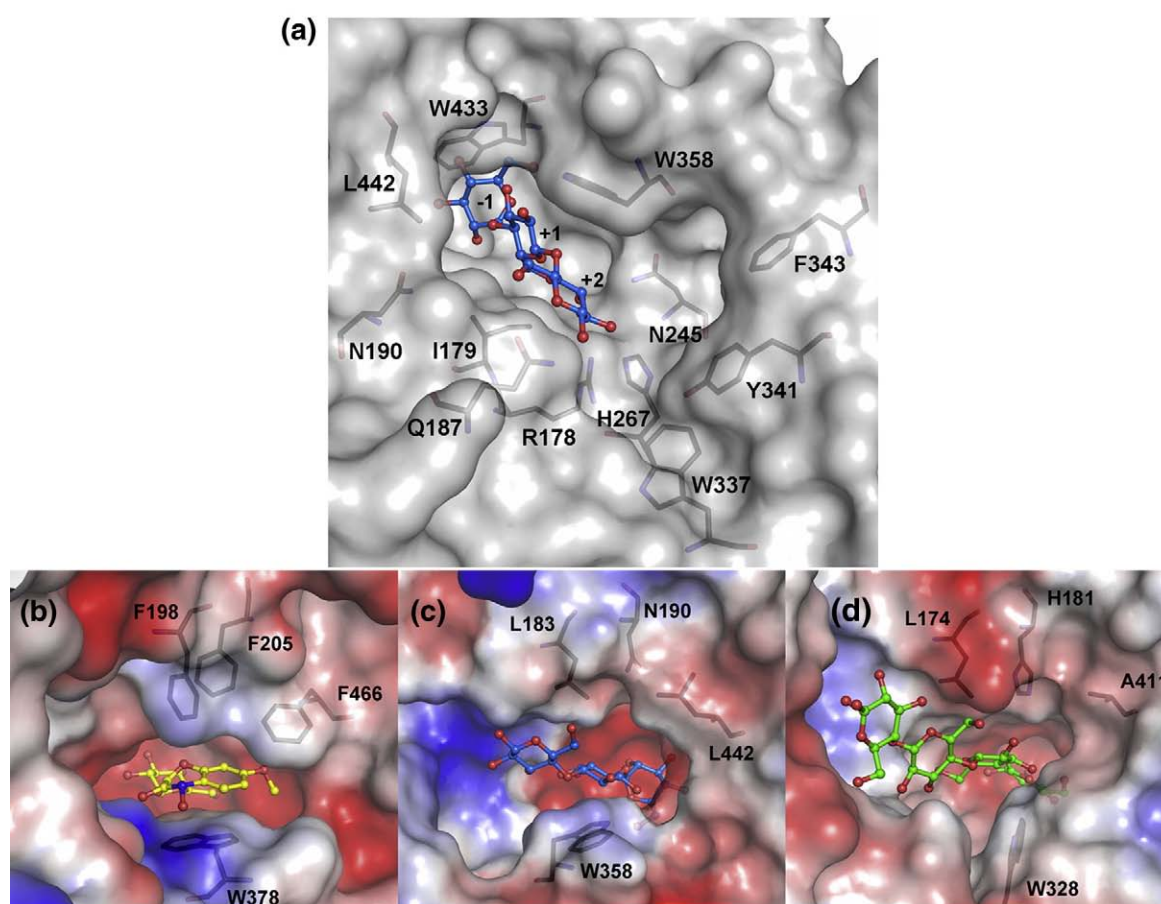


Fig. 7. Substrate contacts in the active site. (a) Molecular surface structure at the active site of the free rice BGlu1 enzyme with cellotriose (blue) positioned in the binding cleft by automated docking. Amino acid residues at subsite -1 and putative aglycone-binding residues at subsites +1 and +2, and possible positions of more distal subsites are shown as stick diagrams below the surface. (b) Crystal structure of maize ZmGlu1 β -glucosidase in complex with DIMBOAGlc used for comparison of the active site geometries. The DIMBOAGlc glucosyl residue with the standard chair conformation is drawn in the active site.¹⁸ (c) The lowest energy conformer of cellotriose (blue) with the skew boat conformation at subsite -1 docked into the free BGlu1 crystal structure active site. (d) Cellotetraose bound in the PpBgIB active site (2Z1S).⁶ Comparison of b, c and d shows that the large aromatic residues F198, F205, and F66 in ZmGlu1 are replaced by smaller L, N or H, and L or A residues in rice BGlu1 and PpBgIB, respectively, while the conserved tryptophan, W358 in rice BGlu1, adopts distinct orientations in these three structures. These changes result in somewhat more open clefts in the latter two structures.

may be mediated by water molecules, such as is seen between the aromatic hydroxyl of dhurrin and a serine in the dhurrinase complex,²⁰ but water molecules were not included in the docking experiment. Nonetheless, the docking is likely to give a good approximation for the positions of glucose residues in subsites +1 and +2, since the non-reducing glucose appears to be docked correctly in subsite -1. The position of the cellotriose chain also suggested that an array of aromatic amino acids around the entrance, including F343, Y341, and W337, may act as platforms for additional glucose residues (Fig. 7a).

Substrate specificity

GH1 enzymes hydrolyze substrates with great variation in their aglycone, so the subsite occupied by the glycone part is remarkably similar in the family, while the aglycone-binding pocket is different for each enzyme. To explain the substrate specificity, the putative substrate-binding residues at the aglycone-binding pocket were inspected carefully.

Rice BGlul and ZmGlul

The substrate specificity of maize ZmGlul β -glucosidase is well characterized, and has been structurally investigated with crystal structures of wild type and inactive mutant enzymes in complex with several ligands.^{18,19} Four amino acid residues were observed to be important for substrate recognition, W378, which is strictly conserved in O-glycosidases, and three phenylalanine residues, F198, F205, and F466, which are located on the opposite side of the active site cleft from W378 and vary among GH1 enzymes (Fig. 7b). A comparison of the BGlul and ZmGlul active sites is shown in Fig. 7b and c. 2-O- β -D-Glucopyranosyl-4-hydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOAGlc) and cellotriose are stacked onto the hydrophobic face of the conserved tryptophan (W358 in BGlul). The three opposing phenylalanine residues in maize are replaced in rice BGlul by L183, N190, and L442, respectively. The result of these differences is that the active site of BGlul is wider, 10 Å from W358 to L182, while that of ZmGlul is narrower at the entrance, measuring 7.6 Å from W378 to F198. As reported recently, the PpBglB structure has similar smaller residues (L174, H181, and A411) in place of the three bulky aromatic residues (Fig. 7d), which makes the entrance between W328 and L174 about 1 Å wider than ZmGlul, and thereby extends the binding cleft for longer oligosaccharides to at least 25 Å.⁶

The superimposition of the two substrate complex structures showed clearly that the reducing end of cellotriose occupied a different aglycone pocket site from DIMBOAGlc. The aglycone part of DIMBOAGlc was bound in a pocket adjacent to the three aromatic phenylalanine residues, while the reducing end of the more hydrophilic cellotriose appears to bind another site (compare Fig. 7b and c).

The result is that, though F446 is in contact with DIMBOA in the ZmGlul structure, the corresponding L442 is far from cellotriose in the predicted BGlul-cellotriose structure. The highly conserved tryptophan (W358 BGlul and W378 in ZmGlul) forms a platform for binding both the second and third glucose residues and DIMBOA, but it is tilted differently in the two enzymes and allows suitable substrate alignment with other interacting side chains.

Rice BGlul and barley BGQ60 β -glucosidases

Of the well-characterized glycosyl hydrolase family 1 enzymes, rice BGlul is most similar to barley BGQ60 β -glucosidase/ β -mannosidase,^{8,9,11,12,27} and both hydrolyze short 1,3- and longer 1,4- β -oligosaccharides. However, while BGQ60 appears to be an exo- β -mannanase²⁸ and prefers mannoside to glucoside and cellobiose to cellotriose, rice BGlul shows the opposite preferences. Differences in the affinities of their +1 and +2 sites may explain their differences in activities toward cello-oligosaccharides.^{9,12} The docking of cellotriose into the BGlul active site indicated that the amino acid residues likely to be involved in hydrogen bonds and hydrophobic interactions with cellotriose at the two reducing end glucosyl residues were located mainly in loop B and loop C, and none of them was conserved among GH1 members, so a comparison of these residues in BGlul and BGQ60 should be informative regarding substrate binding. Though the BGQ60 structure remains to be determined, these residues should be in positions similar to those in BGlul, with which it shares over 66% amino acid sequence identity.

The BGlul residues Y131 (subsite +1), R178 (+2), L183 (+1), D243 (+1), H267 (+2), and W337 (+3) might be crucial for substrate recognition of oligosaccharides, since the corresponding residues are the same in BGQ60, but not in other β -glucosidases that do not hydrolyze oligosaccharides well (Fig. 1). However, the equivalent residues in PpBglB (W123, Y169, L174, N223, I247, and S310) were almost completely different, indicating the bacterial β -glucosidase has evolved a different set of interactions for binding oligosaccharides. The residues I179, L182, N190, N245, and L442 in BGlul are replaced by V184, A187, H195, V250, and R447, respectively, in BGQ60 (Fig. 1), so they might contribute to substrate specificity differences between the rice and barley enzymes. In order to probe the basis of differential binding of cellobiose and cellotriose between rice BGlul and barley BGQ60, mutations were made at four residues found in the putative aglycone binding region of BGlul that differ from those at the same positions in BGQ60.

Kinetic analysis of substrate-binding pocket mutants

At the putative subsite +1 position, BGlul residues I179, N190 and L442 (which is close to subsite

+1, but not close to the docked cellotriose) were mutated to match BGQ60, while at subsite +2, N245, the only residue that is polar in BGLu1 that is non-polar in BGQ60, was mutated. The mutations of BGLu1 substrate binding residues to their counterparts in BGQ60 were expected to change the activity of BGLu1 to be more similar to that of BGQ60, with improvement in the ratio of $k_{\text{cat}}/K_{\text{m}}$ (cellobiose) to $k_{\text{cat}}/K_{\text{m}}$ (cellotriose) and a decreased rate of hydrolysis of *p*NPG.^{12,27} Instead, the L442R mutant had little effect, while the N190H, I179V and N245V mutants each had similar effects on K_{m} and k_{cat} for all substrates (*p*NPG, cellobiose and cellotriose), with little change in substrate selectivity (Table 2). The N190H mutation had the greatest effect and decreased the K_{m} and increased the k_{cat} for all substrates, while the I179V and N245V mutant enzymes exhibited higher values of K_{m} and k_{cat} for almost all substrates, except that N245V had a k_{cat} for *p*NPG similar to wild type.

The expected improvement in cellobiase activity was not detectable for any of the mutants (the $k_{\text{cat}}/K_{\text{m}}$ of the mutants was very low compared to BGQ60). The N245V mutation gave a sixfold increase in k_{cat} for cellobiose to a value similar to that of BGQ60, but the K_{m} also increased over threefold, so the catalytic efficiency ($k_{\text{cat}}/K_{\text{m}}$) was only twofold higher than wild type. Since the N245V mutant also has a high K_{m} for cellotriose and its $k_{\text{cat}}/K_{\text{m}}$ for cellotriose is 15-fold lower than wild type BGLu1, its ratio of $k_{\text{cat}}/K_{\text{m}}$ for cellobiose to $k_{\text{cat}}/K_{\text{m}}$ for cellotriose is 0.14, 28-fold higher than the wild type ratio of 0.005. However, this is still well below the cellobiose/cellotriose catalytic efficiency ratio of 2.15 for BGQ60 and comes at a high cost of substrate affinity for both oligosaccharides. The effects of I179V were similar, and its ratio of $k_{\text{cat}}/K_{\text{m}}$ for hydrolysis of cellobiose to $k_{\text{cat}}/K_{\text{m}}$ for hydrolysis of cellotriose was 0.0167, about threefold higher than wild type. On the other hand, the N190H mutant showed a ratio of $k_{\text{cat}}/K_{\text{m}}$ for cellobiose to $k_{\text{cat}}/K_{\text{m}}$ for cellotriose slightly higher than wild type at 0.008. Although the difference is small, it is possible that this change that increases efficiency, combined with the N245V and I179V changes, which decreased efficiency but gave greater gains in selectivity for cellobiose, could yield the more efficient cellobiase

activity seen in BGQ60. As noted by Mildvan et al.,²⁹ the effects of such mutations on the K_{m} and k_{cat} can be synergistic due to cooperativity between residues acting at the same step of catalysis. Therefore, it should be informative to investigate the double and triple mutations of these residues in the future.

In this study, we have described the structure of rice BGLu1 β -glucosidase, which is an example of how plants have adapted β -glucosidases for hydrolysis of their cell wall-derived oligosaccharides. The active site shows some similarity in shape to the recently described bacterial β -glucosidase, PpBglB, but the interactions with the sugar residues in the +1 and +2 sites are quite different, as demonstrated from comparison of residues near a docked cellotriose substrate. The covalent 2-F-glucoside-bound complex also provides a snap shot of the transglycosylation reaction, which is catalyzed efficiently by BGLu1, with a glycerol molecule from the cryoprotectant playing the role of the incoming acceptor sugar. Mutations of the residues I179, N190 and N245 were shown to have effects on both substrate binding and catalytic rate, confirming their interactions with the substrate, while mutation of L442 to R had little effect. These kinetic results support the automated docking results that placed cellotriose near the first three residues, but away from L442. However, no single change could account for the differences in oligosaccharide preferences between rice BGLu1 and barley BGQ60 β -glucosidases, suggesting multiple residues are involved.

Materials and Methods

Protein expression, purification, and crystallization

Expression and purification of recombinant BGLu1 protein was as described.¹⁴ The immobilized metal-affinity chromatography (IMAC) was performed on Talon Co²⁺ resin (Clontech, Palo Alto, CA) instead of nickel resin for BGLu1 crystallized in complex with 2-deoxy-2-fluoro- β -D-glucoside (DNP2FG), since this yielded purer protein. The free BGLu1 crystal from which data were refined was formed in a hanging-drop, vapor-diffusion drop of 3 μ l of protein (11.5 mg ml⁻¹ in 20 mM Tris-HCl (pH 8.0), 150 mM NaCl) and 1 μ l of precipitant (23% (w/v) polyethylene

Table 2. Kinetic parameters of wild type rice BGLu1, barley β -glucosidase isozyme II; BGQ60, and putative aglycone binding residue mutants for hydrolysis of *p*NPG, cellobiose, and cellotriose

Substrate	Kinetic parameters	Wild type	L442R	N190H	I179V	N245V	BGQ60 ^{a,b}
<i>p</i> NPG	K_{m} (mM)	0.201 $\pm$ 0.006	0.201 $\pm$ 0.008	0.139 $\pm$ 0.005	0.95 $\pm$ 0.08	2.10 $\pm$ 0.15	0.50 ^a
	k_{cat} (s ⁻¹)	4.69 $\pm$ 0.30	4.57 $\pm$ 0.22	6.76 $\pm$ 0.32	6.08 $\pm$ 0.46	4.70 $\pm$ 0.34	0.50 ^a
Cellobiose	$k_{\text{cat}}/K_{\text{m}}$ (s ⁻¹ mM ⁻¹)	23.3	22.7	48.6	6.40	2.24	1 ^a
	K_{m} (mM)	22.0 $\pm$ 2.6	21.9 $\pm$ 1.9	14.3 $\pm$ 0.6	61.1 $\pm$ 2.6	74.4 $\pm$ 6.4	2.67 $\pm$ 0.19 ^b
Cellotriose	k_{cat} (s ⁻¹)	1.16 $\pm$ 0.07	1.12 $\pm$ 0.03	2.09 $\pm$ 0.06	3.42 $\pm$ 0.29	7.17 $\pm$ 0.11	11.58 $\pm$ 0.63 ^b
	$k_{\text{cat}}/K_{\text{m}}$ (s ⁻¹ mM ⁻¹)	0.05	0.05	0.15	0.06	0.10	4.34 ^b
Cellotriose	K_{m} (mM)	0.22 $\pm$ 0.02	0.23 $\pm$ 0.01	0.160 $\pm$ 0.004	1.08 $\pm$ 0.06	9.25 $\pm$ 0.83	0.97 $\pm$ 0.06 ^b
	k_{cat} (s ⁻¹)	2.35 $\pm$ 0.29	2.58 $\pm$ 0.14	3.05 $\pm$ 0.29	3.30 $\pm$ 0.30	6.58 $\pm$ 0.25	1.95 $\pm$ 0.12 ^b
	$k_{\text{cat}}/K_{\text{m}}$ (s ⁻¹ mM ⁻¹)	10.7	11.2	19.1	3.06	0.71	2.01 ^b

^a Kinetic data obtained from Ref. 27.

^b Kinetic data obtained from Ref. 12.

glycol monomethyl ether (PEG MME) 5000, 0.1 M Mes (pH 6.7), 0.17 M ammonium sulfate). The crystal for the final structure of BGlu1 after reaction with DNP2FG was grown by hanging-drop, vapor-diffusion with micro-seeding, as described.¹⁴ The drop containing 2 μ l of protein (2 mg ml⁻¹ in 20 mM Tris-HCl (pH 8.0), 150 mM NaCl containing 0.2 mM DNP2FG) and 1 μ l of precipitant (23% (w/v) PEG MME 5000, 0.1 M Mes (pH 6.7), 0.2 M ammonium sulfate) was equilibrated with the reservoir for 2 h before seeding. The crystals were soaked briefly in 18% (v/v) glycerol in precipitant solution before flash vitrification in liquid nitrogen for diffraction.

Data collection and processing

Data sets used for refinement of the final models of free BGlu1 and its complex with DNP2FG were collected at 2.2 Å and 1.52 Å resolution, respectively, on beamline BL13B1 at the National Synchrotron Radiation Research Center (NSRRC, Hsinchu, Taiwan) with an ADSC Quantum 315 CCD detector. The wavelength was set at 0.98 Å and crystals were maintained at 105 K during diffraction with a nitrogen cold stream. Data were processed and scaled with the HKL-2000 package.³⁰ The X-ray absorbance and fluorescence of Zn²⁺ was scanned over energies within 50 eV of the 9656 eV Zn²⁺ K-edge.

Structure solution and refinement

The structure of free BGlu1 was first solved by molecular replacement with the cyanogenic β -glucosidase from white clover (PDB code 1CBG)¹⁵ as a search model in the AMoRe program,³¹ based on an initial 2.75 Å resolution data set.¹⁴ The rotation and translation functions gave clear solutions for two molecules per asymmetric unit with a correlation coefficient on amplitude (CC) of 56.2% and an *R*-factor of 45.2% in the solution range 10.00–4.00 Å. The σ_A -weighted $|2|F_o| - |F_c||$ and $|F_o| - |F_c||$ electron density maps were calculated at 2.8 Å resolution using the CNS program.³² The amino acids in the model were subsequently changed from those of 1CBG to those of rice BGlu1 and manually adjusted into the electron density using O.³³

The preliminary structure was refined using CNS with the 2.8 Å resolution free BGlu1 dataset. The residues V1-A14, K197-C198, R291-P293, G322-T329, and F343-K348 were deleted from the model, since no electron density was observed at main-chain atoms in these loops. Other residues with poor side-chain electron density were initially mutated to alanine. The derived model, which had an *R* value of 39.5% and an *R*_{free} of 41.8%, was used to solve the structure of the 2-fluoroglucoside (G2F) complex by rigid body refinement with the same *R*_{free} test set as the initial free model. The G2F inhibitor was built in the standard ⁴C₁ chair conformation and its torsion angles defined with ChemOffice (CambridgeSoft, Cambridge, MA). This covalently bound inhibitor was modeled into the $|F_o| - |F_c||$ electron density map at O^{e2} of E386, the catalytic nucleophile. During refinement, water molecules positioned within hydrogen-bonding distance of the model were identified from electron density peaks in the $|F_o| - |F_c||$ map and included in further refinement steps. After several cycles of rebuilding and refinement, the model of the G2F complex gave an *R* value of 25.9% and an *R*_{free} of 30.0% at 2.3 Å resolution.

The latter model was placed into the G2F complex and free enzyme datasets collected at the NSRRC synchrotron (Hsinchu, Taiwan) to 1.52 Å and 2.2 Å resolution,

respectively, by rigid body refinement with the two molecules in the asymmetric unit refined as independent domains. The refinement was executed with REFMAC5 in the CCP4 suite.³⁴ Further refinement was initially done with tight non-crystallographic symmetry (NCS) restraints for both structures. These NCS restraints were maintained throughout the refinement of the free BGlu1 structure, but loose main-chain and side-chain NCS were applied at the final stages of the G2F complex refinement. NCS restraint parameters and final refinement values are listed in Table 1. The stereochemical quality of the final models was assessed with PROCHECK.³⁵ In the Ramachandran plots of the free BGlu1 enzyme and its G2F complex, 89.0% of non-proline and non-glycine residues were found in the most favored regions and a further 10.2% in additional allowed regions.

The figures of protein structures were generated by PyMol (Delano Scientific LLC). Analysis of the interfacial surface and molecular assembly stability was done with the Protein interfaces, surfaces and assemblies service PISA at the European Bioinformatics Institute.²²

Automated docking of cellotriose into the BGlu1 structure

The models of the BGlu1 molecule from the G2F complex structure and the oligosaccharide ligand were prepared for docking studies with the Sybyl 7.1 program (Tripos, Inc., St. Louis, MO). All hetero atoms and molecule B residues were removed from the structure. The N and C termini were fixed by adding blocking groups on the charged N-terminal amine (AMN) and C-terminal carboxylate (CXL) groups on V6 and H476, respectively, and polar hydrogen atoms were added. Charges were assigned by the Kollman united atom method,³⁶ and then the energy of the hydrogen atoms was minimized. The cellotriose ligand used for docking was retrieved from the ligand-bound protein complex of PDB database file 1FBW.³⁷ All hydrogen atoms were added to the ligand and charges were assigned by the Gasteiger-Huckel method³⁸ followed by energy minimization. The ligand was checked for accuracy and its non-rotatable and rotatable bonds were defined, then converted from mol2 to pdbqt format in the Autodock Tools program. All bonds of hydroxyl groups and glycosidic linkages, but not those of the rigid glucose pyranose ring were defined as rotatable bonds. A skew boat conformation was applied to the glucose residue at the non-reducing end with the Spartan program (Wavefunction Inc., Irvine, CA), by which the dihedral angle defined by C2, C1, O5, and C5 ring atoms was changed from -61.31° to +45.00°, as reported.²⁵

Autodock 4.0 was executed 100 times with the Lamarckian genetic algorithm (LGA),³⁹ a population size of 150, elitism set at 1, mutation rate at 0.02, and crossover rate of 0.80. Simulations were performed with a maximum of 2,500,000 energy evaluations and a maximum of 27,000 generations. Docking results were clustered using a cutoff of 2 Å RMSD.

Site-directed mutagenesis

Mutations were constructed with the QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA)

‡ http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html

according to the manufacturer's protocol. The following oligonucleotides were used for mutagenesis. I179V: 5'-TACATTTAATGAGCCAAGGGTAGTACACTTCTTGGTTATG-3', 5'-CATAACCAAGAAGTGCTACTACCTTG-GCTCATTAATGTA-3'; N190H: 5'-GTTATGACCAAG-GAACACATCCTCCTAAAAGGTGC-3', 5'-GCACCTTT-TAGGAGGATGTGTTCTTGGTCATAAC-3'; N245V: 5'-AGTTGGAATAGTTCTGGACTTCGTATGGTAT-GAAGCTCTTTCCAACCTC-3', 5'-GAGTTGGAAAGA-GCTTCATACCATAACGAAGTCCAGAACTATTC-CAACTT-3'; L442R: 5'-ACAACTTCGAGTGGCGGT-CAGGTTACACGTC-3', 5'-GACGTGTAACCTGACCGC-CACTCGAAGTTGT-3' (the mutated codons are underlined). All mutations were confirmed by DNA sequencing.

Kinetic studies

Wild type and mutant BGlu1 enzymes were purified as described.⁹ Purified protein concentration was determined by the Bio-Rad Bradford protein assay (Bio-Rad Corp., Hercules, CA) with BSA as the standard. High concentrations of protein were diluted in 50 mM sodium acetate buffer (pH 5.0), to concentrations appropriate for kinetic assays. To determine the kinetic parameters, wild type and mutant BGlu1 were assayed in 50 mM sodium acetate buffer (pH 5.0) at 30 °C with seven substrate concentrations (~0.1–sevenfold of the K_m value, when practical) for a time that gave an initial velocity (V_0). Blanks containing buffer and substrate were set up at every concentration of substrate. The kinetic parameters were calculated by non-linear regression of Michaelis–Menten plots with the GraFit 5.0 program.⁴⁰ Linear regression of Lineweaver–Burk plots was used for some mutants that had very high K_m , for which a saturation curve could not be obtained due to limitations of substrate solubility (>200 mM).

For *p*-nitrophenyl β -D-glucopyranoside (*p*NPG), 70 μ l of each substrate concentration was mixed with 70 μ l of enzyme (1.5 nmol) in microtiter plate wells. The reactions were incubated at 30 °C for 20 min., and then stopped by adding 70 μ l of 0.4 M NaCO₃. The 405 nm absorbance was measured and the liberated *p*NP was quantified by comparison to a *p*-nitrophenol (*p*NP) standard. To analyze the effects of Zn²⁺ and EDTA on BGlu1 activity, the wild type BGlu1 enzyme was pre-incubated in the presence of 10 mM ZnSO₄ or EDTA for 30 min. It was then diluted 70-fold in 50 mM sodium acetate (pH 5.0), and assayed for hydrolysis of 1 mM *p*NPG, as described above.

Glucose released from oligosaccharide substrates was assayed with a coupled peroxidase-glucose oxidase reaction. Samples (25 μ l each) of substrate (1.5 nmol of celotriose or 30 nmol of cellobiose) and protein were mixed well in 1.5 ml tubes and incubated at 30 °C for 20 min. The reaction was stopped by heating at 80 °C for 5 min, and then transferred to a microtiter plate for measurement of the glucose: 100 μ l of peroxidase/glucose oxidase enzyme and 50 μ l of ABTS were added to each microtiter plate well and the plate was incubated at 37 °C for 30 min. The 405 nm absorbance was measured and compared to a glucose standard curve to quantify the amount of glucose.

Determination of solution molecular mass

Protein purified as described above was subjected to chromatography on Sephacryl S-200 (for fusion protein) and on a Superdex S-200 (for BGlu1 after removal of the

tag), in 20 mM Tris (pH 8.0), 150 mM NaCl, and the elution volume compared to 12.4–200 kDa gel-filtration molecular mass markers (Sigma) to determine the apparent native molecular mass. The major peak eluted at a similar time as bovine serum albumin (66 kDa) for the fusion protein and slightly later for the BGlu1 with the fusion tag removed. In addition, 0.5 mg ml⁻¹ of fusion protein dissolved in 20 mM Tris–HCl (pH 8.0), 150 mM NaCl was tested for dynamic light-scattering on a Zetasizer nano system (Malvern Instruments Ltd, UK) with 20 rounds of determination, in which a single peak was observed.

Protein Data Bank accession numbers

The coordinates and structure factors of the free BGlu1 and BGlu1-G2F complex structures have been deposited in the RSCB Protein Data Bank with the accession numbers 2RGL and 2RGM, respectively.

Acknowledgements

We thank Catleya Rojviriyia for assistance and advice with protein crystallography software, and Hong-Hsiang Guan for assistance with diffraction and dynamic light-scattering data collection. We thank the protein crystallography beamline staff of the National Synchrotron Radiation Research Center (NSRRC), Hsinchu, Taiwan, for data collection and processing assistance. We acknowledge the Faculty of Science, Mahidol University for providing in-house X-ray diffraction facility. This work was supported by grants from the Thailand Research Fund (BRG4780020 and RTA4780006). W.C. is a Royal Golden Jubilee PhD Fellow and J.S. is a Senior Research Scholar of the Thailand Research Fund. R.C.R. was supported by A*STAR, Singapore. Portions of this research were carried out at the National Synchrotron Radiation Research Center, a national user facility supported by the National Science Council of Taiwan, ROC. The Synchrotron Radiation Protein Crystallography Facility is supported by the National Research Program for Genomic Medicine.

References

1. Béguin, P. (1990). Molecular biology of cellulose degradation. *Annu. Rev. Microbiol.* **44**, 219–245.
2. Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309–316.
3. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P. & 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.
4. Jenkins, J., Lo Leggio, L., Harris, G. & Pickersgill, R. (1995). β -glucosidase, β -galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes 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.
5. Coutinho, P. M. & Henrissat, B. (1999). Carbohydrate-active enzymes: an integrated database approach. In *Recent Advances in Carbohydrate Bioengineering* (Gilbert, H. J., Davies, G., Henrissat, B. & Svensson, B., eds), pp. 3–12, The Royal Society of Chemistry, Cambridge, UK.
 6. Isorna, P., Polaina, J., Latorre-Garcia, L., Canada, F. J., Gonzalez, B. & Sanz-Aparicio, J. (2007). Crystal structures of *Paenibacillus polymyxa* β -glucosidase B complexes reveal the molecular basis of substrate specificity and give new insights into the catalytic machinery of family I glycosidases. *J. Mol. Biol.* **371**, 1204–1218.
 7. 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.
 8. Opassiri, R., Ketudat, J., Cairns, R., Akiyama, T., Wara-Aswapati, O., Savasti, J. & Esen, A. (2003). Characterization of a rice β -glucosidase highly expressed in flower and germinating shoot. *Plant Sci.* **165**, 627–638.
 9. Opassiri, R., Hua, Y., Wara-Aswapati, O., Akiyama, T., Svasti, J., Esen, A. & Ketudat Cairns, J. R. (2004). β -Glucosidase, exo- β -glucanase and pyridoxine transglucosylase activities of rice BGLu1. *Biochem. J.* **379**, 125–131.
 10. Opassiri, R., Pomthong, B., Onkoksoong, T., Akiyama, T., Esen, A. & Ketudat Cairns, J. (2006). Analysis of rice glycosyl hydrolase family I and expression of Os4bglu12 β -glucosidase. *BMC Plant Biol.* **6**, 33.
 11. Leah, R., Kigel, J., Svendsen, I. & Mundy, J. (1995). Biochemical and molecular characterization of a barley seed β -glucosidase. *J. Biol. Chem.* **270**, 15789–15797.
 12. Hrmova, M., MacGregor, E. A., Biely, P., Stewart, R. J. & Fincher, G. B. (1998). Substrate binding and catalytic mechanism of a barley β -D-glucosidase/(1,4)- β -D-glucan exohydrolase. *J. Biol. Chem.* **273**, 11134–11143.
 13. Hommalai, G., Withers, S. G., Chuenchor, W., Ketudat Cairns, J. & Svasti, J. (2007). Enzymatic synthesis of cello-oligosaccharides by rice BGLu1 β -glucosidase glycosynthase mutants. *Glycobiology*, **17**, 744–753.
 14. Chuenchor, W., Pengthaisong, S., Yavaniyama, J., Opassiri, R., Svasti, J. & Ketudat Cairns, J. (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**, 798–801.
 15. Barrett, T., Suresh, C. G., Tolley, S. P., Dodson, E. J. & Hughes, M. A. (1995). The crystal structure of a cyanogenic β -glucosidase from white clover, a family 1 glycosyl hydrolase. *Structure*, **3**, 951–960.
 16. Burmeister, W. P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S. & 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**, 663–676.
 17. Burmeister, W. P., 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.
 18. Czjzek, M., Cicek, M., Zamboni, V., Bevan, D. R., Henrissat, B. & Esen, A. (2000). The mechanism of substrate (aglycone) specificity in β -glucosidases is revealed by crystal structures of mutant maize β -glucosidase-DIMBOA, -DIMBOAGlc, and -dhurrin complexes. *Proc. Natl Acad. Sci. USA*, **97**, 13555–13560.
 19. Czjzek, M., Cicek, M., Zamboni, V., Burmeister, W. P., Bevan, D. R., Henrissat, B. & Esen, A. (2001). Crystal structure of a monocotyledon (maize ZMGlu1) β -glucosidase and a model of its complex with *p*-nitrophenyl β -D-thioglucoside. *Biochem. J.* **354**, 37–46.
 20. Verdoucq, L., Moriniere, J., Bevan, D. R., 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.
 21. 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.
 22. Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797.
 23. Sanz-Aparicio, J., Hermoso, J. A., Martinez-Ripoll, M., Lequerica, J. L. & 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.
 24. Zechel, D., Boraston, A. B., Gloster, T., Boraston, C. M., Macdonald, J. M., Tibbrook, D. M. G. *et al.* (2003). Iminosugar glycosidase inhibitors: structural and thermodynamic dissection of the binding of Isofagomine and 1-Deoxynojirimycin to β -glucosidases. *J. Am. Chem. Soc.* **125**, 14313–14323.
 25. Biarnes, X., Nieto, J., Planas, A. & Rovira, C. (2006). Substrate distortion in the Michaelis complex of *Bacillus* 1,3-1,4- β -glucanase: Insight from first principles molecular dynamics simulations. *J. Biol. Chem.* **281**, 1432–1441.
 26. Barleben, L., Panjikar, S., Ruppert, M., Koepke, J. & Stöckigt, J. (2007). Molecular architecture of strictosidine glucosidase: the gateway to the biosynthesis of the monoterpene indole alkaloid family. *Plant Cell*, **19**, 2886–2897.
 27. Hrmova, M., Harvey, A. J., Wang, J., Shirley, N. J., Jones, G. P., Stone, B. A. *et al.* (1996). Barley β -D-glucan exohydrolase with β -D-glucosidase activity: purification, characterization, and determination of preliminary structure from a cDNA clone. *J. Biol. Chem.* **271**, 5277–5286.
 28. Hrmova, M., Burton, R. A., Biely, P., Lahnstein, J. & Fincher, G. B. (2006). Hydrolysis of (1,4)- β -D-mannans in barley (*Hordeum vulgare* L.) is mediated by the concerted action of (1,4)- β -D-mannan endohydrolase and β -D-mannosidase. *Biochem. J.* **399**, 77–90.
 29. Mildvan, A. S., Weber, D. J. & Kuliopulos, A. (1992). Quantitative interpretations of double mutations of enzymes. *Arch. Biochem. Biophys.* **294**, 327–340.
 30. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology* (Carter, C. W. & Sweet, R. M., eds), vol. 276, pp. 307–326, Academic Press, New York.
 31. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A*, **50**, 157–163.
 32. Brunger, A. T., Adam, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system (CNS): a new

- software system for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.
33. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **47**, 110–119.
34. Murshudov, G. N., Lebedev, A., Vagin, A. A., Wilson, K. S. & Dodson, E. J. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. *Acta. Crystallogr. D*, **55**, 247–255.
35. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
36. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G. *et al.* (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765–784.
37. Parsiegla, G., Reverbel-Leroy, C., Tardif, C., Belaich, J. P., Driguez, H. & Haser, R. (2000). Crystal structures of the cellulase Cel48F in complex with inhibitors and substrates give insights into its processive action. *Biochemistry*, **39**, 11238–11246.
38. Gasteiger, J. & Marsili, M. (1980). Interactive partial equalization of orbital electronegativity - a rapid access to atomic charges. *Tetrahedron*, **36**, 3219–3228.
39. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K. & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19**, 1639–1662.
40. Leatherbarrow, R. J. (2001). *GraFit*, 5th edit., Erithacus Software Ltd, Horley, UK.
41. Gouet, P., Robert, X. & Courcelle, E. (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31**, 3320–3323.



Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/locate/yabbiRice family GH1 glycoside hydrolases with β -D-glucosidase and β -D-mannosidase activitiesTeerachai Kuntothom^a, Sukanya Luang^a, Andrew J. Harvey^b, Geoffrey B. Fincher^b, Rodjana Opasiri^a, Maria Hrmova^{b,*}, James R. Ketudat Cairns^{a,*}^a School of Biochemistry, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand^b School of Agriculture, Food and Wine, Australian Centre for Plant Functional Genomics, University of Adelaide, Waite Campus, Glenn Osmond, Australia

ARTICLE INFO

Article history:

Received 21 July 2009

and in revised form 10 September 2009

Available online 18 September 2009

Keywords:

*Hordeum vulgare**Oryza sativa*

Molecular modelling

Phylogenetic analysis

Subsite mapping

Substrate specificity

ABSTRACT

Plant β -D-mannosidases and a rice β -D-glucosidase, Os3BGlu7, with weak β -D-mannosidase activity, cluster together in phylogenetic analysis. To investigate the relationship between substrate specificity and amino acid sequence similarity in family GH1 glycoside hydrolases, Os3BGlu8 and Os7BGlu26, putative rice β -D-glucosidases from this cluster, and a β -D-mannosidase from barley (rHvBII), were expressed in *Escherichia coli* and characterized. Os3BGlu8, the amino acid sequence and molecular model of which are most similar to Os3BGlu7, hydrolysed 4-nitrophenyl- β -D-glucopyranoside (4NPGlc) faster than 4-nitrophenyl- β -D-mannopyranoside (4NPMn), while Os7BGlu26, which is most similar to rHvBII by these criteria, hydrolysed 4NPMn faster than 4NPGlc. All the enzymes hydrolyzed cellooligosaccharides with increased hydrolytic rates as the degree of polymerization increased from 3–6, but only rHvBII hydrolyzed cellobiose with a higher k_{cat}/K_m value than cellotriose. This was primarily due to strong binding of glucosyl residues at the +2 subsite for the rice enzymes, and unfavorable interactions at this subsite with rHvBII.

© 2009 Elsevier Inc. All rights reserved.

Introduction

Plant β -D-glucosidases (3.2.1.21) are enzymes that hydrolyse glycosidic linkages at the non-reducing ends of their substrates, which include β -D-glucooligosaccharides and β -D-glucosides. They are involved in several important processes in plants, including cell wall remodelling [1], seed germination [2], release of reactive metabolites from inactive precursors [3], and release of phytohormones from inactive glycoconjugates [4]. Plant β -D-glucosidases belong to glycoside hydrolase (GH) families GH1, GH3 and GH5 (CAZy: <http://www.cazy.org/>) [5]. The GH1 and GH5 enzymes belong to the GH-A clan and have typical $(\beta/\alpha)_8$ folds [6,7], while the GH3 enzymes have a multi-domain organization that folds into $(\alpha/\beta)_8$ and $(\alpha/\beta)_6$ architectures [8]. Most plant β -D-glucosidases fall in the GH1 family, which also contains plant β -D-mannosidases (3.2.1.25) and β -D-thioglucohydrolases (3.2.3.1).

In rice (*Oryza sativa* L.), 40 GH1 genes were identified, 34 of which at least are likely to encode functional rice β -D-glucosidases [9]. From this number of genes, it is highly likely that rice β -D-glucosidases have a range of substrate specificities. Several β -D-glucosidases have been characterized from rice seedlings [9–13].

Schliemann [10] described an enzyme in dwarf rice seedlings that hydrolysed gibberellin β -glucosides. A cell wall-associated β -D-glucosidase that hydrolysed glucose disaccharides and oligosaccharides, but not polysaccharides, has been described [11]. It was specific for cellooligosaccharides and laminarioligosaccharides, and preferred disaccharides to higher oligosaccharides. Rice BGlu1, hereafter designated Os3BGlu7 to be consistent with the names of other rice isoenzymes, is another rice β -D-glucosidase, which is highly expressed in germinating seed, shoot, and flower, and is also specific to (1,3)- and (1,4)- β -D-linked glucooligosaccharides [12]. Opasiri et al. [9] found that Os3BGlu7 and three other rice β -D-glucosidase isoenzymes (Os1BGlu1, Os3BGlu8, and Os7BGlu26) were grouped in the same phylogenetic cluster with barley (*Hordeum vulgare*, HvBII) [14], *Arabidopsis thaliana* [15] and tomato β -D-mannosidases [16] (Fig. 1), and it was previously speculated that one of these enzymes may actually be a β -D-mannosidase [15]. Barley HvBII also has β -D-glucosidase activity similar to Os3BGlu7, but Os3BGlu7 and HvBII differ in that Os3BGlu7 is primarily a β -D-glucosidase and prefers cellotriose to cellobiose, while HvBII is primarily a β -D-mannosidase and prefers cellobiose to cellotriose.

In this study, the two rice β -D-glucosidase isoenzymes closely related to Os3BGlu7 and plant β -D-mannosidases, Os3BGlu8 and Os7BGlu26, were expressed and their kinetic parameters for cellooligosaccharides, laminarioligosaccharides, and aryl β -D-glycosides

* Corresponding authors. Fax: +66 44 224185 (J.R. Ketudat Cairns).

E-mail addresses: maria.hrmova@adelaide.edu.au (M. Hrmova), cairns@sut.ac.th (J.R. Ketudat Cairns).

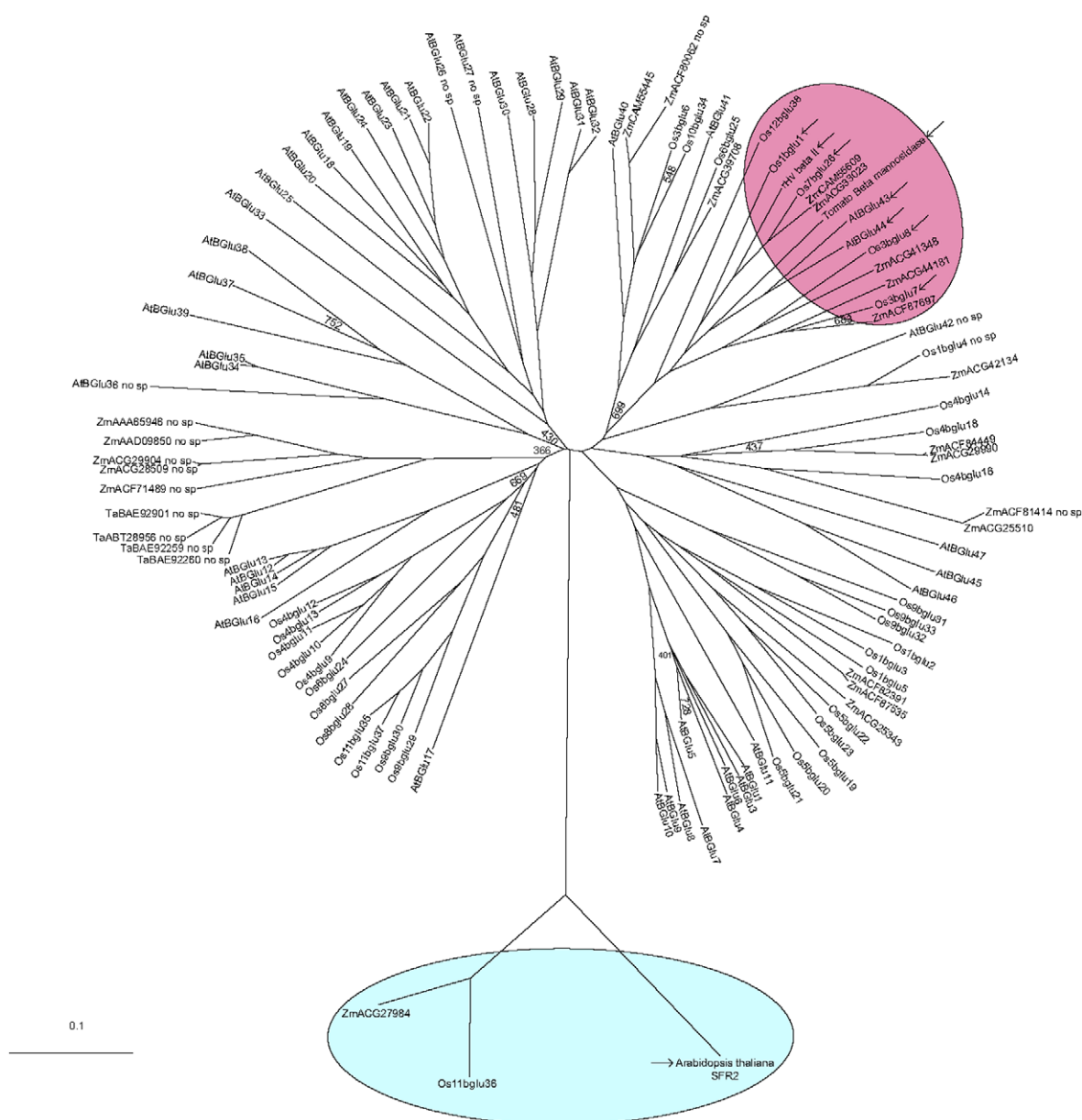


Fig. 1. An unrooted radial phylogenetic tree of 111 plant sequences of family GH1. Sequences from *Arabidopsis*, rice, and maize, as well as four wheat sequences, were taken from CAZy database [5] and aligned with the barley rHvBII (GenBank Accession Nos. EU807965) and tomato (AC AAL37714) [17] β -D-mannosidases to show the relation of this cluster to other plant family GH 1 enzymes. The sequences divided in subgroups 1 and 2 are indicated by dashed and full lines, respectively. Amino acid sequences were aligned with ClustalX and branch lengths are drawn to scale. Bootstrap values for reproducibility out of 1000 trials are shown at the confluence of clusters, for which these values were below 800. A complete set of bootstrap reproducibility values is shown in Supplementary Fig. 1. The eight sequences marked by arrows in the circle denote the plant β -D-mannosidases and β -D-glucosidases shown in the alignment in Fig. 2. The “no sp” indicates proteins with no secretory signal sequence, although the cluster of sequences ZmAAA65946 to TaBAE92260 represents chloroplast enzymes with chloroplast signal peptides. The cluster of the distantly related SFR2-like enzymes from *Arabidopsis*, rice and maize is circled at the base of the tree.

were determined. The relative β -D-mannosidase and β -D-glucosidase activities of, Os3Bglu8 and Os7Bglu26 were compared to those of Os3Bglu7 and HvBII that was also expressed in the same heterologous bacterial system (rHvBII) to see, how well the activities of all enzymes matched their sequence similarities and positions on a phylogenetic tree.

Materials and methods

Cloning and expression of rice and barley GH1 cDNAs

Total mRNA was extracted from 3-day-old germinated barley cv. Clipper seeds with TRIzol® reagent and reverse-transcribed

with polyT₁₇ primer and Superscript II reverse transcriptase, according to the manufacturer's instructions (Invitrogen, Carlsbad, CA, USA). The cDNA encoding the full-length barley rHvBII was PCR amplified from the first-strand cDNA pool with *Pfu* DNA polymerase (Promega, Madison, WI, USA) and the BGQ60_Ntermf (5'-GGAC ACGAGGATGAGGTCTC-3') and BGQ60_Ctermr (5'-CCGGATCCTG TTTGCGATCCTAG-3') primers, designed based on the BGQ60 cDNA sequence (GenBank Accession No. L41869) [14]. The cDNA encoding the mature barley β -D-glucosidase, designated rHvBII, was amplified with the sense primer BGQ60MatNcof (5'-CCGGATC CTGTTTTCGATCCTAG-3') and the antisense primer BGQ60CtermX-ho1 (5'-CTCGAGCTAGCTCTCTCTTTTCGGAGAG-3'). The mature BGQ60 cDNA was cloned into pBluescript SK II (Stratagene, La Jolla,

CA, USA) at the *EcoR* V site and sequenced in both directions (GenBank Accession No. EU807965). The cDNA encoding rHvBII was sub-cloned into the pET32a expression vector (Novagen, Madison, WI, USA) at the *NcoI* and *XhoI* sites.

The *Os7Bglu26* cDNA was amplified by RT-PCR from RNA isolated from 7-day-old rice (*O. sativa* L., cv. KDML105) seedlings with (dT)₂₀ primer and the Thermoscript RT-PCR system (Eppendorf, Hamburg, Germany). The full-length *Os7Bglu26* cDNA was amplified with the AK068499Nter (5'-ACATGAGGAAATTCATAGCAGC-3') and AK068499StopR (5'-TAGTCCCTTCTGTGCTGCTC-3') primers and *Pfu* DNA polymerase. A cDNA encoding the predicted mature *Os7Bglu26* isoenzyme was amplified from the initial PCR product with the AK068499MatNf (5'-CACCTACTGGCTCAACCCGGAG-3') and AK068499StopR primers and *Pfu* DNA polymerase, cloned into a Gateway entry clone with the pENTRTM-D-TOPO[®] Cloning Kit (Invitrogen), and sequenced in both directions. The cDNA encoding the mature *Os7Bglu26* (GenBank Accession No. EU835514) was recombined into the pET32a/DEST expression vector [9] with LR clonase (Invitrogen).

Plasmid constructs containing full-length cDNAs encoding the *Os1Bglu1* (GenBank Accession No. AK069177, clone ID 204872), *Os3Bglu8* (AK120790, clone ID 215509), *Os12Bglu38* (GeneBank Accession No. AK071058, clone ID 023711) rice β -D-glucosidase isoenzymes were acquired from the Rice Genome Resource Center, Tsukuba, Japan (<http://www.rgrc.dna.affrc.go.jp/>) [17]. The cDNA encoding the mature *Os1Bglu1* isoenzyme was amplified from the clone 204872 plasmid with *Pfu* DNA polymerase and the AK069177CostF2 (5'-CACCATGGTCACGGCGGGCTGAGC-3') and AK069177ctermR2 (5'-CCTCGAGTCAGTTTTGCTGCTG-3') primers, while that for the *Os3Bglu8* isoenzyme was amplified from the clone 215509 plasmid with the AK120790NcoIFwd (5'-CCATGGCGGCGCGCTTCCC-3') and AK120790stopXhoIR (5'-TAATCGAGCTAAATTGCTACTCTACAG-3') primers, and for *Os12Bglu38* was amplified from the clone 023711 plasmid with the primers AK071058NcoIFwd (5'-CACCATGGAGCAGACCGACCT-3') AND AK071058StopXhoIR (5'-GCCTCGAGTGATAATAATCAACTCAAACC-3'). The cloning, sequencing, and construction of a pET32/DEST expression construct for the cDNA encoding mature *Os1Bglu1* were carried out as described for *Os7Bglu26*. The cDNAs encoding the mature *Os3Bglu8* and *Os12Bglu38* proteins were cloned, sequenced and sub-cloned into pET32a, as described for rHvBII. The cDNA encoding the mature *Os1Bglu1*, *Os3Bglu8*, and *Os12Bglu38* proteins matched the AK069177, AK120790, and AK071058 sequences, respectively, over their entire lengths.

Phylogenetic analysis of family GH1 members

For the construction of the bootstrapped rooted and radial unrooted phylogenetic trees, 111 plant amino acid sequences of family GH1 glycoside hydrolases were taken from CAZY database (<http://www.cazy.org/>), and aligned with the sequences listed in Fig. 2. All sequences were scanned for the presence of signal peptides, using the program SignalP 3.0 [18], and secretory signal sequences, where present, were removed. The sequences were aligned with the program ClustalW [19], and a tree was generated with the internal neighbour joining algorithm with the default parameters, bootstrapped with 1000 trials, and viewed using Tree-View [20].

Molecular modelling

The rice Bglu1 (*Os3Bglu7*) β -D-glucosidase crystal structure (Protein Data Bank Accession No. 2RGL) was used as template for comparative homology modelling [21]. Protein sequence alignments between the 2RGL template and each β -D-glucosidase were

prepared with ClustalW [19]. The three-dimensional (3D¹) models of β -D-glucosidases were constructed with Modeller9v2 software [22], that uses satisfaction of spatial restraints and statistical analysis of the known secondary structures in the Modeller database. The five from 20 models generated by Modeller9v2 that showed the lowest objective function values were selected. The choices of best models were further refined with Procheck, by checking the stereochemical parameters and G-factors of the models [23], and by Prosa2003 [24], which determines if the model has a correct fold by energy comparison between template and model structures (z-scores) that specify these folds.

Protein expression

Protein expression was carried out after transformation of the pET32a expression constructs into *Escherichia coli* Origami(DE3) cells (Novagen) by the CaCl₂ method [25] and selection on Luria-Bertani (LB) agar containing 50 μ g/ml ampicillin, 15 μ g/ml kanamycin, and 12.5 μ g/ml tetracycline. The selected clones were grown and induced for expression as previously described for *Os3Bglu7* [12]. The optimum expression conditions were determined by varying the duration of expression from 3 to 16 h and the concentration of IPTG at 0.2, 0.4 and 0.6 mM, at temperatures of 20 and 25 °C.

Purification of recombinant enzymes

The cells were collected by centrifugation after overnight expression, frozen at -70 °C and lysed to release total soluble proteins, as previously described [12]. The heterologously expressed enzymes were first purified by IMAC resin (GE Healthcare, Buckinghamshire, United Kingdom) charged with Co²⁺ (rHvBII and *Os3Bglu8*) or with Ni²⁺ (*Os7Bglu26*), as with *Os3Bglu7* [12,21]. The IMAC column was washed twice with 10 column volumes (CV) of equilibration/wash buffer (100 mM NaCl, 50 mM Tris-HCl, pH 7.2) to remove unbound protein. Resin-bound fusion proteins were eluted with 2 CV of elution buffer (150 mM imidazole, 50 mM Tris-HCl, pH 7.2). The eluted fractions were analysed for hydrolysis of 4NPGlc before overnight dialysis in 20 mM Tris-HCl, pH 7.2, at 4 °C.

The *Os3Bglu8* and *Os7Bglu26* proteins were then purified by Q-Sepharose chromatography (GE Healthcare) in 50 mM sodium acetate, pH 5.0, with elution by a gradient from 0–0.5 M NaCl. The *Os3Bglu8* and *Os7Bglu26* β -D-glucosidases eluted at 0.18 M, and 0.22 M NaCl, respectively. The NaCl concentration of the *Os3Bglu8* β -D-glucosidase solution was adjusted to 3 M with solid NaCl and the enzyme was loaded on a Phenyl Sepharose (GE Healthcare) column, which was pre-equilibrated with 50 mM sodium acetate, pH 5.0, containing 3 M NaCl. The protein was eluted with a decreasing gradient of 3–0 M NaCl over 50 ml at a flow rate of 1 ml/min. The *Os3Bglu8* β -D-glucosidase eluted at 0.22 M NaCl. The fractions that were active toward 4NPGlc were pooled and dialysed in 20 mM Tris-HCl, pH 7.2, typically for 16 h at 4 °C. *Os7Bglu26* was further purified by size-exclusion chromatography on a Superdex S200 column (10 $\times$ 300 mm, GE Healthcare), in 50 mM Tris-HCl, pH 8.0, containing 150 mM NaCl, at a flow rate of 0.5 ml/min.

The NH₂-terminal thioredoxin-His tags of the *Os3Bglu8* and rHvBII proteins were removed by digestion with 2 μ g of enterokinase (New England Biolabs, Beverly, MA, USA) per 1 mg of total

¹ Abbreviations used: 3D, three-dimensional; 4NP, 4-nitrophenol; 4NPGlc, 4NP- β -D-glucosylpyranoside; 4NPMann, 4NP- β -D-mannopyranoside; CV, column volume(s); DP, degree(s) of polymerization; GH1, glycoside hydrolase family 1; HvBII, native barley (*Hordeum vulgare*) β -D-glucosidase isoenzyme β II; IPTG, isopropyl β -D-1-thiogalactopyranoside; IMAC, immobilized metal affinity column; MALDI-TOF, matrix-assisted laser desorption/ionization time of flight; MS, mass spectrometry; Q-Sepharose, quaternary amino strong anion exchanger Sepharose; rHvBII, recombinant barley β -D-glucosidase isoenzyme β II, TLC, thin layer chromatography.

protein according to the supplier's instructions. The released fusion tag was removed by adsorption to a second IMAC column. The unbound and wash fractions from IMAC with hydrolytic activity toward 4NPGlc were pooled and dialysed for 16 h against 50 mM sodium acetate, pH 5.0. The dialysed rHvBII was purified by cation-exchange chromatography on CM-Sepharose (Takara, Tokyo, Japan) with elution by a 0–0.5 M NaCl gradient in 50 mM sodium acetate, pH 5.0, at a flow rate of 1 ml/min. rHvBII eluted at 0.15 M NaCl. Because Os7BGlu26 appeared to be more stable as a fusion protein, it was not digested with enterokinase, and the fusion protein was investigated, as described for Os3BGlu7 [26].

Tryptic mapping by MALDI-TOF/TOF mass spectrometry (MS) of expressed rHvBII and Os3BGlu8 was done as described previously [27], and confirmed that the purified proteins corresponded to the expected enzymes.

Enzyme assays and determination of pH and temperature optima

Enzyme activities were assayed for 10–30 min, depending on the enzyme, in reaction mixtures with 1 mM 4NPGlc in 50 mM sodium acetate, pH 5.0, at 30 °C. Reactions were terminated with two volumes of 2 M Na₂CO₃ and absorbance at 405 nm was measured.

The optimum temperature was determined over the range of 0–90 °C in 10 °C increments, by incubating 4 pmol of enzyme at the varying temperatures for 15 min, after which 4NPGlc was added to 1 mM concentration and incubation proceeded for a further 30 min. The optimum pH was determined by incubating 4 pmol of an enzyme with 1 mM 4NPGlc in McIlvain buffers, with pH ranging from 3.0 to 12.0, for 30 min at 30 °C. The reactions were terminated and rates measured as specified above.

Determination of substrate specificity

The substrate specificities of Os3BGlu8, Os7BGlu26 and rHvBII were evaluated by incubating 4 pmol of enzyme in 100 µl reaction mixtures with 3 mM substrates at 30 °C for 1–16 h in 50 mM sodium acetate, at the optimum pH. Hydrolysis of aryl substrates was detected as described above for 4NPGlc. For cellooligosaccharides and laminarioligosaccharides, the reactions were terminated by heating at 95 °C for 5 min. The amount of glucose released was determined by peroxidase/glucose-oxidase coupled reactions (PGO assay, Sigma–Aldrich, St. Louis, MO, USA) as previously described [12]. Hydrolysis of mannanoligosaccharides and natural glucosides was detected by TLC, as previously described [28].

Hydrolytic rates with mannanoligosaccharides were determined for rHvBII and Os3BGlu7. The 200 µl reaction mixtures contained 0.45 nmol of enzyme and 1.6 nmol of mannanoligosaccharides in 10 mM sodium acetate, at pH 4.0 for rHvBII and pH 5.0 for Os3BGlu7. The reaction mixtures were incubated at 30 °C for 20 min, terminated by boiling for 5 min and the mixtures were dried in a DNA110 SpeedVac® vacuum centrifuge (Savant Instruments, Holbrook, NY, USA). Dry products were dissolved in 75% (v/v) acetonitrile in water. The released mannose was separated on an Alltech® Prevail carbohydrate ES column (Grace, Deerfield, IL, USA), which was connected to a model 1090 liquid chromatography system (Agilent Technology, Palo Alto, CA, USA), at a flow rate of 0.4 ml/min. The mannanoligosaccharides were vaporized and detected with an Alltech® 800 evaporative light scattering detector. The integrated area of a known mannose peak was used to calculate the amount of mannose released by comparison to a mannose standard curve.

Determination of kinetic parameters

Kinetic parameters were determined from triplicate reactions containing 2–10 pmol (20–100 nM) enzyme and substrate concen-

trations from 0.1 to 90 mM, depending on the specific activity of the enzyme and on the apparent K_m , in 50 mM sodium acetate at the optimum pH values. Time course assays were done for the highest and lowest substrate concentrations for each substrate and enzyme combination to ensure that the velocities measured were in the linear range (i.e. they were at the apparent V_0), and all absorbance readings were measured in the range of 0.1–0.9 absorbance units. Kinetic parameters were calculated by nonlinear regression of the Michaelis–Menten curves with Grafit 5.0 (Erithacus Software, Horley, Surrey, UK). Subsite affinities of β -D-glucosidase enzymes were estimated by the method derived from those of Hrmova et al. [29] and Hiromi et al. [30] that was previously described [26].

Results

Sequence analysis of rice Os1BGlu1, Os3BGlu8, Os7BGlu26, Os12BGlu38 enzymes

Protein sequence-based phylogenetic analysis showed that four putative rice β -D-glucosidases, Os1BGlu1, Os3BGlu8, Os7BGlu26, and Os12BGlu38 cluster together with Os3BGlu7, HvBII, and tomato [16] and *Arabidopsis* [15] β -D-mannosidases (Fig. 1). The amino acid sequence of Os3BGlu8 is most similar to Os3BGlu7 with 74% identity over the catalytic region, but Os3BGlu8 contains a C-terminal extension, including a hydrophobic motif that could act as a membrane anchor at the C-terminus. In contrast, the Os7BGlu26 amino acid sequence shared 82% identity with that of HvBII, but only 64% with that of Os3BGlu7, while that of Os1BGlu1 shared 73% identity to HvBII, but only 64% identity with Os3BGlu7. Os12BGlu38 was more distantly related and shared only 56–61% amino acid sequence identity with the other enzymes in this phylogenetic cluster. Since overall sequence similarity may not reflect the active site residues that interact with the substrate to determine an enzyme's specificity, the residues within the active site of the Os3BGlu7 crystal structure (Protein Data Bank entry 1RGM [21]) and the corresponding residues in all the rice isoenzymes and HvBII were identified (Fig. 2). All the residues in direct contact with the sugar at the –1 subsite were conserved in the five rice and barley enzymes and tomato and *Arabidopsis* β -D-mannosidases, but Val241 in Os3BGlu7, which is in contact with the side chain of the catalytic acid/base Glu176 corresponds to Leu in HvBII and its close relatives. Both conserved and variant residues were mapped onto the +1 and +2 subsites, with most sequence differences occurring between the subgroup 1 of Os3BGlu7 and Os3BGlu8 and the subgroup 2 of Os7BGlu26 and HvBII (Fig. 1 underlined sequences, and Fig. 2).

To further investigate the active site similarities and differences within subgroups 1 and 2 (Figs. 1 and 2), molecular models of HvBII and the rice isoenzymes were constructed using the Os3BGlu7 crystal structure as a template [21]. The overall depth and width of the active sites of the Os3BGlu8, Os7BGlu26, and HvBII active sites were similar to the template structure of Os3BGlu7 (Fig. 3). Further, all enzymes had negatively charged surface potentials around the catalytic residues and formed extended funnels leading into the active site regions, which bind cellooligosaccharides in HvBII [29] and Os3BGlu7 [26]. As for the active site amino acid residues, the HvBII, Os7BGlu26, and Os1BGlu1 enzymes were most similar to each other, while Os3BGlu8 was more similar to Os3BGlu7. The active site amino acid residues corresponding to the individual subsites (–1 and +1 to +4) are marked in the alignment in Fig. 2.

Cloning and characterization of barley rHvBII

To compare the rice GH1 enzymes to a plant β -D-mannosidase produced in the same bacterial expression system, a barley GH1

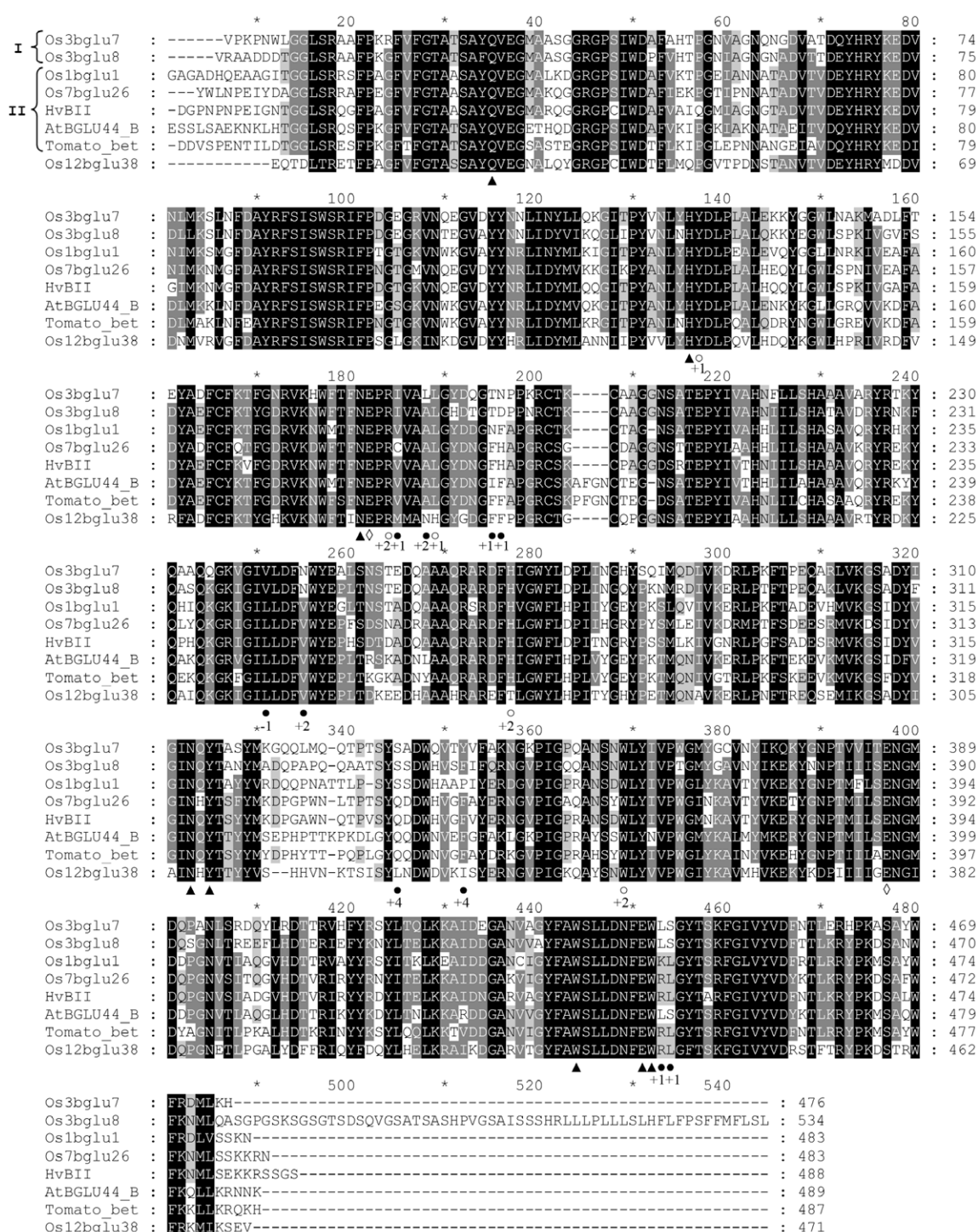


Fig. 2. Amino acid sequence alignment of rHvBII, Os1Bglu1, Os3Bglu7, Os3Bglu8, and Os7Bglu26 with tomato and *Arabidopsis* β-D-mannosidases. Diamonds indicate acid/base and nucleophilic residues, triangles represent amino acid residues contacting the glucosyl residue at the -1 subsite in Os3Bglu7, open circles denote amino acid residues predicted to be at the +1, +2, +3, and +4 subsites in Os3Bglu7, filled circles denote predicted active site amino acid residues that differ between the various β-D-glucosidases. The numbers beneath the circles represent the primary subsites at which the marked amino acid residues are predicted to contact a glucosyl residue, based on the Os3Bglu7 3D structure. Only those residues that differ are indicated for the predicted +3 and +4 subsites. The alignment was generated using ClustalW [18]. The subgroups 1 and 2 and indicated by Roman numerals (cf. Fig. 1).

glycoside hydrolase was cloned, based on the BGQ60 sequence [14], and expressed as a His₆-tagged thioredoxin fusion protein in *E. coli*. Its sequence matched that of BGQ60 β-D-glucosidase at all but four residues, two of which matched barley β-D-glucosidase

isoenzyme βII (HvBII) at the residues that were different between HvBII and BGQ60 in the first 110 residues [2] (Supplementary Fig. 2). Given that original sequences were generated from enzymes isolated from different barley cultivars, this clone was

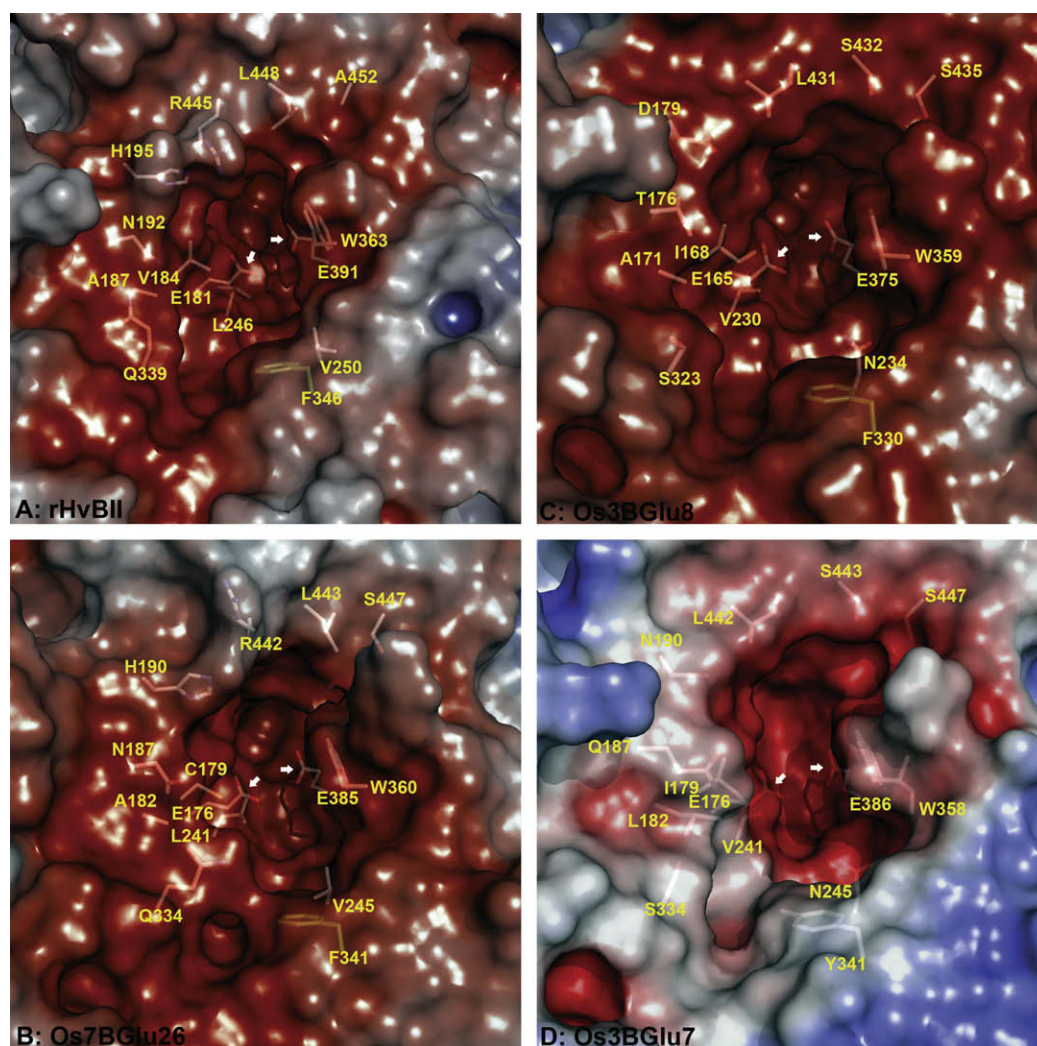


Fig. 3. Active site structures of the rice and barley β -D-glucosidase/ β -D-mannosidase cluster enzymes. The molecular surface morphologies illustrate the catalytic sites of the modelled barley rHvBII (A), and rice Os7BGlu26 (B), and Os3BGlu8 (C), and the crystal structure of Os3BGlu7 (D). The variant amino acid residues between β -D-glucosidases are shown as sticks under the projected molecular surfaces. The catalytic acid/base and nucleophile residues (Glu residues marked by white arrows) and a conserved aglycone-binding site tryptophan are also shown in all models. The surface electrostatic potentials were calculated with Adaptive Poisson–Boltzmann Solver [41]. Blue and red areas indicate electropositive and electronegative regions contoured at $\pm 5 \text{ kTe}^{-1}$.

considered to match the HvBII protein and the protein expressed from it was designated as rHvBII. Its activity was nearly identical to that of HvBII purified from germinating seed (Table 1) [2]. The pH and temperature optima of rHvBII were determined to be pH 4.0 and 30 °C.

The relative β -D-mannosidase activities of rHvBII toward mannoooligosaccharides of DP 2–6 were compared to those of Os3BGlu7. While the relative β -D-mannosidase activities of both rHvBII and Os3BGlu7 increased with increasing DP of mannoooligosaccharides from 2 to 6, rHvBII had higher β -D-mannosidase activity compared to Os3BGlu7 for every mannoooligosaccharide (Fig. 4).

Substrate specificity comparison among the clustered rice and barley enzymes

The preceding structural analysis suggested that Os1BGlu1, Os3BGlu8, and Os7BGlu26 should hydrolyse the same substrates as rHvBII and Os3BGlu7, and that Os1BGlu1 and Os7BGlu26 should show substrate specificity more similar to rHvBII, i.e. with preferences for 4NPMAN over 4NPGlc and cellobiose over cellotriose, while Os3BGlu8 should be more similar to Os3BGlu7. To test this hypothesis, the four rice isoenzymes Os1BGlu1, Os3BGlu8, Os7B-

Glu26, and Os12BGlu38 were expressed in the *E. coli* expression system that was used for expression of rHvBII and Os3BGlu7. The Os3BGlu8 and Os7BGlu26 were expressed as functional enzymes, while Os1BGlu1 and Os12BGlu38 could not be expressed in active forms in this heterologous host. The pH and temperature optima were at pH 5.0 and 30 °C for Os3BGlu8, and at pH 4.5 and 40 °C for Os7BGlu26 (data not shown). These values were comparable to the pH 5.0 and 30 °C optima for Os3BGlu7 [12] and pH 4.0 and 30 °C optima for rHvBII.

Measurements of hydrolytic activities on synthetic and natural substrates for the 3 enzymes indicated that, while they were similar in hydrolysing 4NP-glycosides, they showed marked differences in hydrolysis of 4NPMAN, mannoooligosaccharides and natural β -linked glucosides. Os3BGlu8 had little activity toward 4NPMAN and mannoooligosaccharides, while Os7BGlu26 and rHvBII hydrolysed 4NPMAN better than 4NPGlc, and also hydrolysed (1,4)- β -linked mannoooligosaccharides with DP of 2–6. Aside from 4NPGlc and 4NPMAN, all enzymes hydrolysed 4NP- β -D-fucoside slightly better than 4NPGlc, and 4NP- β -D-galactoside (4NPGal), 4NP- β -D-xyloside (4NPXyl), and 4NP- β -L-arabinoside at low rates (Table 1). All three enzymes could hydrolyse (1,4)- β -linked glucooligosaccharides of DP 2–6 and (1,3)- β -linked glucooligosaccharides

Table 1

Relative rates of hydrolysis of natural and synthetic substrates by rice Os3BGlu8, Os7BGlu26 and barley rHvBII. Hydrolysis of 4NP-glycosides was detected spectrophotometrically, while hydrolysis of gluco-oligosaccharides was detected by glucose-oxidase assay and that of manno-oligosaccharides and glucosides by TLC, as described in the Experimental section. Rates of hydrolysis are given relative to 4NPGlc (listed as 100%), which corresponds to $2.59 \times 10^{-4} \mu\text{mol/min}/\mu\text{g}$ enzyme for Os1BGlu1, $1.87 \times 10^{-4} \mu\text{mol/min}/\mu\text{g}$ for Os3BGlu8, $1 \times 10^{-6} \mu\text{mol/min}/\mu\text{g}$ for Os7BGlu26, and $1.6 \times 10^{-4} \mu\text{mol/min}/\mu\text{g}$ for HvBII. Characters + or – indicate that hydrolysis was or was not detected, respectively, while NT means not tested. The structures of the natural glucosides are shown in Supplementary Fig. 3.

Substrate	Os3BGlu8	Os7BGlu26	rHv BII
4NP- β -D-glucopyranoside	100 $\pm$ 3	100 $\pm$ 4	100 $\pm$ 3
4NP- β -D-mannopyranoside	4.6 $\pm$ 0.4	302 $\pm$ 4	238 $\pm$ 12
4NP- β -D-galactopyranoside	11 $\pm$ 3	12.6 $\pm$ 1.7	17 $\pm$ 3
4NP- β -D-xylopyranoside	2.9 $\pm$ 2.0	6.4 $\pm$ 1.5	3.1 $\pm$ 1.9
4NP- β -D-fucopyranoside	115 $\pm$ 7	128 $\pm$ 8	129 $\pm$ 10
4NP- β -L-arabinopyranoside	2.6 $\pm$ 1.1	9 $\pm$ 3	16 $\pm$ 3
<i>Oligosaccharides</i>			
Sophorose [β -(1,2)-linked]	57 $\pm$ 4	25 $\pm$ 4	15 $\pm$ 4
Laminaribiose [β -(1,3)]	199 $\pm$ 17	158 $\pm$ 13	13 $\pm$ 4
Laminaritriose	39 $\pm$ 10	23 $\pm$ 5	2.8 $\pm$ 1.1
Cellobiose [β -(1,4)]	17 $\pm$ 6	21 $\pm$ 5	33 $\pm$ 7
Cellotriose	83 $\pm$ 18	124 $\pm$ 18	19 $\pm$ 8
Cellotetraose	156 $\pm$ 9	194 $\pm$ 11	52 $\pm$ 9
Cellopentaose	147 $\pm$ 15	403 $\pm$ 17	225 $\pm$ 15
Cellohexaose	188 $\pm$ 16	472 $\pm$ 12	303 $\pm$ 12
Gentiobiose [β -(1,6)]	4.8 $\pm$ 1.0	–	–
Mannobiose [β -(1,4)]	–	+	+
Mannotriose	–	+	+
Mannotetraose	–	+	+
Mannopentaose	–	+	+
Mannohexose	–	+	+
<i>Natural glucosides</i>			
Salicin	–	–	+
Esculin	+	–	+
Epiheterodendrin	–	–	+
Dhurrin	NT	+	+
d-Amygdalin	–	+	+
Prunasin	NT	–	NT
p-Coumaric alcohol glucoside	–	+	+
Coniferin	–	–	+

No hydrolytic activities were detected against 4NP- β -L-arabinopyranoside, 4NP- β -D-glucopyranoside, laminaritetraose, laminaripentaose, laminarihexaose, linamarin and indoxyl- β -D-glucoside.

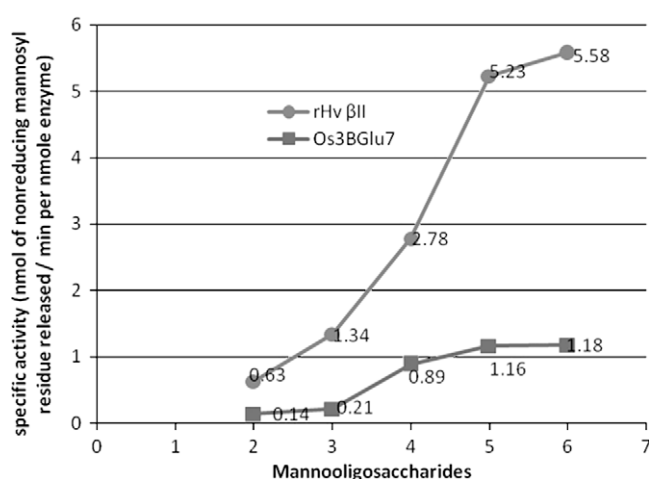


Fig. 4. Relative β -D-mannosidase activities of barley rHvBII and rice Os3BGlu7 toward (1,4)- β -mannooligosaccharides of DP 2–6. The numbers shown in the graph represent hydrolytic rates, expressed as nmol of mannosyl residues released per min per nmol enzyme. The values for mannobiose are divided by two, as both reducing and non-reducing end residues are released in a single cleavage event.

of DP 2–3, but not laminaritetraose and longer laminarioligosaccharides. All enzymes hydrolysed sophorose [(1,2)- β -linked gluco-

disaccharide], but only Os3BGlu8 significantly hydrolyzed gentiobiose [(1,6)- β -linked gluco-disaccharide]. Among the natural glucosides tested, rHvBII hydrolysed the broadest range of aglycones, while Os3BGlu8 could only hydrolyse esculin (Table 1 and Supplementary Fig. 3).

The activity of Os7BGlu26 and rHvBII toward β -(1,4)-linked manno-oligosaccharides were compared (Fig. 5). Os7BGlu26 had very low activity toward mannobiose, but had significantly higher activity toward mannotriose, mannotetraose, mannopentaose, and mannohexaose, resulting in a build-up of mannobiose. In comparison, rHvBII had much higher activity toward mannobiose than Os7BGlu26.

The apparent kinetic parameters for hydrolysis of 4NPGlc, 4NPMan, cellooligosaccharides and laminaribiose and laminaritriose by Os3BGlu8, Os7BGlu26, and rHvBII are summarised in Table 2. Os3BGlu8 hydrolysed 4NPGlc more efficiently than 4NPMan (80 times higher k_{cat}/K_m for 4NPGlc), as did Os3BGlu7, to which the Os3BGlu8 sequence was most similar, while Os7BGlu26 and rHvBII, which had sequences most similar to each other, both hydrolysed 4NPMan with apparent k_{cat}/K_m values that were higher than those for 4NPGlc. All three recombinant enzymes hydrolysed laminaribiose more efficiently than laminaritriose, and hydrolysed cellooligosaccharides with higher efficiency as the degree of polymerization (DP) increased from 3 to 6. The two rice isoenzymes hydrolysed cellobiose poorly, as did Os3BGlu7 [26], but rHvBII hydrolysed cellobiose with a higher k_{cat}/K_m value than cellotriose.

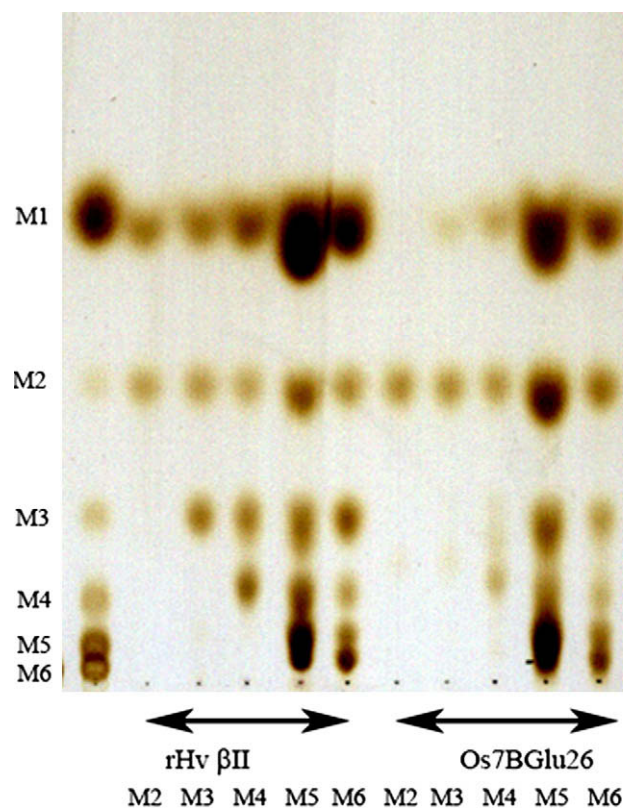


Fig. 5. Hydrolysis of manno-oligosaccharides by HvBII and Os7BGlu26. Hydrolytic reactions were catalysed by 0.27 nmol of rHvBII or 0.19 nmol Os7BGlu26 in the presence of 1 mM manno-oligosaccharides in 10 mM sodium acetate buffer, pH 4.5. The reactions proceeded for 30 min at 30 °C. Standards (lane 1) were mannose, (M1), and mannobiose to mannohexaose (M2–M6). M2 to M6 indicate reaction mixtures with individual manno-oligosaccharides with DP of 2 (M2) to 6 (M6) hydrolysed with either rHvBII or Os7BGlu26. The other conditions were as previously described [27].

Table 2
Apparent kinetic parameters of Os1BGlu1, Os3BGlu8, Os7BGlu26, rHvBII, HvBII, and Os3BGlu7 toward 4NPGlc, 4NPMAN, celooligosaccharides with DP of 2–6 and laminaribiose and laminarotriose.

Substrate ^a	Apparent kinetic parameters	Os3BGlu8	Os7BGlu26	rHvBII	HvBII ^b	Os3BGlu7 ^c
4NPGlc	K_m (mM)	0.27 ± 0.04	0.27 ± 0.02	0.50 ± 0.03	0.50	0.23 ± 0.02
	k_{cat} (s ⁻¹)	3.3 ± 0.4	0.16 ± 0.003	0.50 ± 0.07	0.50	7.9 ± 0.4
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	13 ± 3	0.63 ± 0.003	1.00 ± 0.05	1.00	35 ± 1
4NPMAN	K_m (mM)	1.6 ± 0.3	0.52 ± 0.04	0.25 ± 0.01	ND	1.3 ± 0.1
	k_{cat} (s ⁻¹)	0.24 ± 0.03	1.10 ± 0.03	3.06 ± 0.02	ND	1.32 ± 0.05
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	0.16 ± 0.05	2.11 ± 0.20	12.7 ± 0.2	ND	1.01 ± 0.02
Cellobiose	K_m (mM)	26 ± 2	19.6 ± 1.9	2.76 ± 0.10	2.7 ± 0.2	32 ± 2
	k_{cat} (s ⁻¹)	0.98 ± 0.08	0.52 ± 0.03	16.1 ± 0.2	11.6 ± 0.6	1.5 ± 0.1
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	0.04 ± 0.01	0.026 ± 0.001	5.8 ± 0.2	4.34 ± 0.07	0.050 ± 0.002
Cellotriose	K_m (mM)	0.56 ± 0.01	0.52 ± 0.05	0.74 ± 0.06	0.97 ± 0.06	0.72 ± 0.02
	k_{cat} (s ⁻¹)	6.08 ± 0.08	0.67 ± 0.03	3.0 ± 0.4	2.0 ± 0.1	18.1 ± 0.4
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	11 ± 2	1.27 ± 0.03	3.45 ± 0.08	2.01 ± 0.01	25.4 ± 0.04
Cellotetraose	K_m (mM)	0.25 ± 0.01	0.09 ± 0.01	1.03 ± 0.02	0.89 ± 0.05	0.28 ± 0.01
	k_{cat} (s ⁻¹)	16.1 ± 0.1	0.97 ± 0.02	9.6 ± 0.4	8.9 ± 0.6	17.3 ± 0.6
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	65 ± 2	10.9 ± 0.4	9.3 ± 0.5	9.98 ± 0.08	61.1 ± 0.4
Cellopentaose	K_m (mM)	0.15 ± 0.01	0.06 ± 0.01	0.33 ± 0.02	0.41 ± 0.02	0.24 ± 0.01
	k_{cat} (s ⁻¹)	15.9 ± 0.1	1.07 ± 0.03	12.9 ± 0.4	11.7 ± 0.8	16.9 ± 0.1
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	109 ± 3	18.5 ± 0.4	40.2 ± 0.8	28.4 ± 0.4	72 ± 2
Cellohexaose	K_m (mM)	0.12 ± 0.01	0.05 ± 0.01	0.23 ± 0.01	0.29 ± 0.02	0.22 ± 0.01
	k_{cat} (s ⁻¹)	212 ± 2	1.31 ± 0.04	13.10 ± 0.08	11.8 ± 0.8	16.9 ± 0.3
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	175 ± 5	29 ± 3	54 ± 2	40.7 ± 0.1	153 ± 0.5
Laminaribiose	K_m (mM)	0.32 ± 0.07	0.86 ± 0.07	4.7 ± 0.4	5.4 ± 0.4	2.05 ± 0.01
	k_{cat} (s ⁻¹)	11.1 ± 0.3	0.61 ± 0.02	11.3 ± 0.03	14.1 ± 1.0	32 ± 3
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	36 ± 8	0.70 ± 0.06	2.30 ± 0.06	2.63 ± 0.01	16 ± 2
Laminarotriose	K_m (mM)	6.0 ± 0.6	8.7 ± 0.9	2.8 ± 0.2	2.8 ± 0.2	1.92 ± 0.04
	k_{cat} (s ⁻¹)	13 ± 1	2.4 ± 0.1	2.3 ± 0.2	1.44 ± 0.10	21.2 ± 0.2
	k_{cat}/K_m (s ⁻¹ mM ⁻¹)	2.2 ± 0.3	0.28 ± 0.006	0.84 ± 0.01	0.52 ± 0.01	11.0 ± 0.2

^a Hydrolysis of 4NP-glycosides was detected as described in the Materials and methods.

^b Kinetics constants were taken from [29]. ND, means not determined.

^c Kinetics constants were taken from [26].

Subsite mapping

To understand the differences in celooligosaccharide preferences between the rice and barley β -D-glucosidases and β -D-mannosidases, their subsite-binding affinities for (1,4)- β -linked glucooligosaccharides were estimated, with the assumption of an intrinsic k_{cat} constant [30]. Fig. 6 indicates that all the enzymes have six subsites for (1,4)- β -linked glucosyl residues, as predicted from the similarities of their active site residues to Os3BGlu7 and rHvBII and by molecular modelling. The main difference between rHvBII and the rice isoenzymes was the apparent negative affinity at the +2 subsite in rHvBII and the apparent relatively high positive affinity of all the rice isoenzymes at this site. The subsite affinities of Os7BGlu26 are the most similar to those of the barley enzyme HvBII, with the interaction at the +1 subsite being stronger than at the +2 subsite. The Os3BGlu8 isoenzyme was more similar to Os3BGlu7 [26], with stronger interaction at the +2 subsite, but still maintained relatively strong binding at the +1 subsite compared to Os3BGlu7.

Discussion

From the GH1 β -mannosidase-like enzyme cluster, the coding regions of six enzymes were sub-cloned in the pET32(+) vector and expressed in *E. coli* as thioredoxin fusion proteins. The pET-32a(+) expression system was chosen, because it facilitates disulfide bond formation in the reducing environment of *E. coli* cytoplasm [31], and alleviates toxic effects of recombinant protein on host cells. It is of note that all the studied enzymes possessed one disulfide bond at the end of catalytic regions. While Os3BGlu7 [12], Os3BGlu8, and Os7BGlu26 and barley HvBII enzymes were produced in active forms, Os1BGlu1 and Os12BGlu38 could not be produced in active forms. The catalytically competent Os3BGlu7 [12], Os3BGlu8, and Os7BGlu26 and barley HvBII enzymes had two, two, five and two putative *N*-glycosylation sites, respectively,

whereas Os1BGlu1 and Os12BGlu38 had four and five respective *N*-glycosylation sites. Although production of active Os7BGlu26 suggests that the absolute number of *N*-glycosylation site is not directly related to expression of active GH1 β -mannosidase-like enzymes, lack of glycosylation and other factors are likely to contribute to production of active forms.

The substrate specificities and kinetic parameters of rHvBII and HvBII isolated from germinated seeds [2,29] were nearly identical, suggesting that the substrate specificities determined for the rice isoenzymes that were expressed in the *E. coli* expression system, are likely to match those of the corresponding native enzymes functioning in rice. Rice Os3BGlu7 had a k_{cat}/K_m value for 4NPGlc over 30-fold higher than that for 4NPMAN [12], but rHvBII hydrolysed 4NPMAN better than 4NPGlc. While Os3BGlu7 activity against celooligosaccharides increased with DP and cellobiose served as a very poor substrate, the activity of rHvBII was higher for cellobiose than cellotriose, and increased with the DP from cellotriose to cellohexaose. Thus, it might be expected the rice isoenzymes with higher sequence similarity to rHvBII would be primarily β -D-mannosidases with high cellobiase activities, while isoenzymes more similar to Os3BGlu7, would be primarily β -D-glucosidases with low cellobiase activities.

Investigation of substrate binding of maize β -D-glucosidase ZmGlu1 by X-ray crystallography and mutagenesis has documented the importance of Phe198, Phe205, Trp378 and Phe466 and Ala467 to aglycone-binding and positioning for hydrolysis of a glycosidic bond [32–34]. Despite sharing 72% amino acid sequence identity with the maize β -D-glucosidase, the substrate specificity of *Sorghum bicolor* dhurrinase Dhr1 appeared to be guided by a different set of aglycone-binding residues, Asn259, Phe261 and Ser262, and a differently shaped pocket compared to ZmGlu1 [35]. These studies also showed that the orientation of the glucosyl residue in the -1 subsite differed between the maize and sorghum enzymes, despite the absolute conservation of glycone-binding residues, and that the binding of the aglycone

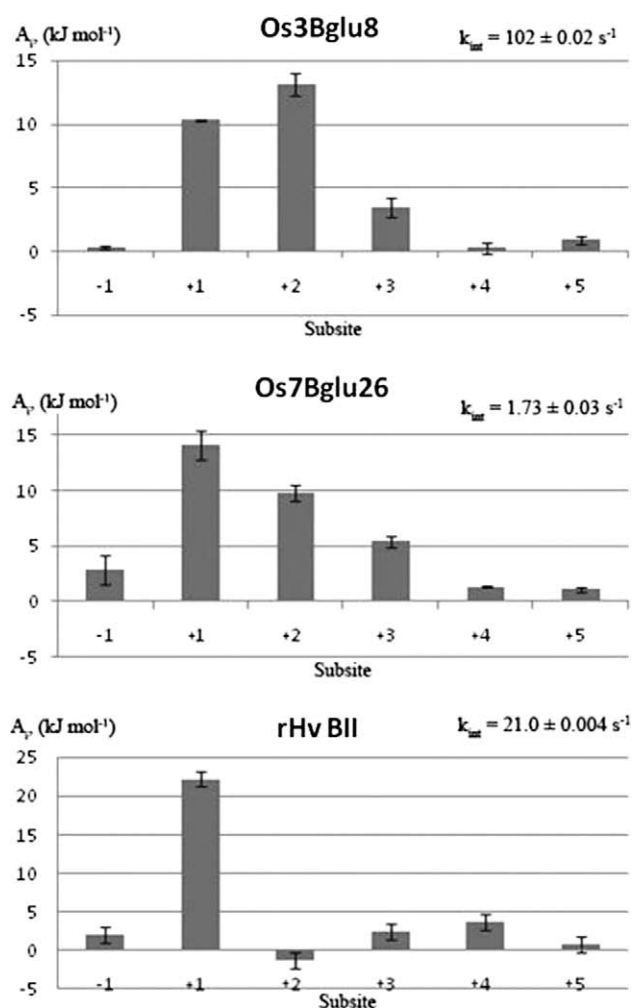


Fig. 6. Apparent subsite-binding affinities of rice Os3BGlu8, and Os7BGlu26 and barley rHvBII for cellooligosaccharides. Subsite affinity and the intrinsic k_{cat} (k_{int}) value estimates were calculated from the kinetic data in Table 2 as previously described [23]. The apparent intrinsic k_{cat} (k_{int}) for each isoenzyme is shown in the upper right hand corner of the graph with standard deviations.

can affect the binding position of the glucosyl residue at the -1 subsite [32,35,36].

Protein sequence alignment of Os3BGlu8, Os7BGlu26, rHvBII, and Os3BGlu7, (Fig. 2) demonstrated that these enzymes have a range of identical amino acid residues that control glycone and aglycone-binding in plant β -D-glucosidases [36,37]. Comparative molecular modelling of Os3BGlu8, Os7BGlu26, and rHvBII with the Os3BGlu7 crystal structure as a template, yielded reliable three-dimensional models (Fig. 3). The predicted catalytic residues, corresponding to Glu181 and Glu391 in rHvBII, were found to be located at the base of the pocket near the centre of the GH1 (β/α)₈ barrel fold in all four models, as expected. The similar shapes of the pockets with the long grooves leading to the active site were consistent with the fact that all four enzymes hydrolysed the linear cellooligosaccharides with efficiencies that increased with DP from 2 up to 6, but did not tolerate the more curved geometries of longer (1,3)- β -glucosaccharides. Thus, laminaritrise was hydrolysed at a lower rate than laminaribiose. The similar size and negative surface potential of the active site funnel is consistent with the need to stabilize the oxocarbenium cation-like transition state [37]. Based on the structure of Os3BGlu7 [21] and its mutant bound to cellopentaose (PDB entry 3F5K, W. Chuenchor, S. Pengthaisong and J.R. Ketudat Cairns, unpublished), the rice β -D-glucosidases and barley rHvBII possessed few differences in identities of

amino acid residues that formed the -1 to +3 subsites. Based on the similarity of their amino acid sequences and molecular models, Os7BGlu26 was projected to have activities similar to those of rHvBII, while Os3BGlu8 was expected to have activities similar to those of Os3BGlu7. However, a mutation of the ZmGlu1 aglycone-binding residue Phe198 to Val was previously shown to lead to a series of unpredictable changes in the positions of surrounding amino acid residues in the active site that resulted in poor binding and hydrolysis of 4NPGlc [32]. This observation suggested that it might be difficult to predict the relationships between the identities of the active site residues and substrate specificities.

Comparison of the apparent k_{cat}/K_m values of Os3BGlu8, Os7BGlu26, rHvBII, and Os3BGlu7, toward 4NPMann and 4NPGlc showed that rHvBII and Os7BGlu26 had the highest apparent k_{cat}/K_m values for 4NPMann hydrolysis and the lowest apparent k_{cat}/K_m values for 4NPGlc hydrolysis. In Os7BGlu26 and rHvBII, the respective apparent k_{cat}/K_m values for 4NPGlc were approximately 3-fold and 12-fold lower than their apparent k_{cat}/K_m values for 4NPMann. The Os3BGlu7 and Os3BGlu8 isoenzymes displayed lower apparent k_{cat}/K_m values for 4NPMann (Table 2). Therefore, it would be more appropriate to designate rHvBII and Os7BGlu26 as β -D-mannosidases, while Os3BGlu7 and Os3BGlu8 need to be denoted as β -D-glucosidases, because their apparent k_{cat}/K_m values were at least 30 times higher for 4NPGlc than for 4NPMann.

Since the difference between mannosides and glucosides is in the axial versus equatorial position of the C2 hydroxyl group of the sugar residue at the non-reducing end of the substrates bound in the active site, it is expected that the active sites should exhibit differences near their catalytic acid/base and nucleophile residues. However, all of the amino acid residues directly contacting the non-reducing end sugar residue are conserved in the plant GH1 β -mannosidase-like enzymes, so the differences that define the substrate specificities of these enzymes are likely to be caused by the surrounding residues. One notable difference is the sequence of LSENG around the catalytic nucleophile region in rHvBII and ITENG in Os3BGlu7. It was previously suggested that the LS/AENG sequence could be diagnostic of β -D-mannosidases [15]. However, despite the presence of the intermediate ISENG sequence at the putative nucleophile in Os3BGlu8, it had the greatest preference for glucoside over mannoside with a k_{cat}/K_m for 4NPGlc 81-fold higher than its k_{cat}/K_m for 4NPMann compared to 35-fold higher in Os3BGlu7, which has the ITENG sequence. The only other difference we noted in the -1 subsite between the Os3BGlu7 crystal structure and the molecular model of rHvBII, was the replacement of Val241, which lies against the catalytic acid/base sidechain in the Os3BGlu7 crystal structure, with Leu246. Verdoucq et al. [35] noted that differences in aglycone-binding could affect sugar position, which could, in principle, change the sugar specificity. It is interesting that another residue that is distinct between the β -D-glucosidases and β -D-mannosidases is Asn245 in Os3BGlu7, which is replaced by Val in β -D-mannosidases. Although this residue appears to be at the +2 subsite, mutation to Val in Os3BGlu7 resulted in a 10-fold increase in K_m for 4NPGlc to give an apparent k_{cat}/K_m of $2.2 \text{ s}^{-1} \text{ mM}^{-1}$, similar to that of rHvBII ($1.0 \text{ s}^{-1} \text{ mM}^{-1}$) [21]. The Asn245 O δ is hydrogen-bonded to the amino group of Tyr315, the side chain of which hydrogen bonds with the glycone O5 and also interacts with the indole ring of Trp358, the conserved platform for positioning the aglycone in GH1 β -D-glucosidases. However, the N245V mutation alone could not change the preference of Os3BGlu7 to that of a β -D-mannosidase (W. Chuenchor and J.R. Ketudat Cairns, unpublished data). Since β -D-mannosidases and β -D-glucosidases have been noted to use the same conserved amino acid residues to bind differently shaped transition states [37], it seems likely that subtle differences in the positioning of these residues or of the glycone residue, due to the interactions of several residues, may be responsible for differences in substrate specificities.

Barley rHvBII hydrolysed cellobiose at least 60-fold more efficiently than the rice β -D-glucosidase isoenzymes, which have sequences that are closely related to rHvBII. As pointed out by Hrmova et al. [29], the ability of the barley β -D-glucosidase to hydrolyse cellobiose reflects its poor binding at the subsite +2, thus preventing unproductive binding of cellobiose between subsites +1 and +2. The assumption of an intrinsic reaction rate constant (k_{int}) used subsite-binding affinity estimation [30] does not reflect the conversion of binding energy into formation of the transition state or cooperative interactions between sites. However, it provides a convenient way to estimate the number and relative strengths of subsites from steady-state kinetic data, without the need for pre-steady-state kinetics-based estimates of binding parameters. This approach has proven to be reliable for estimating the number of binding subsites and relative strengths of binding in many instances [26,29], although substantial errors were observed in some cases [38]. It was at least unambiguous that all isozymes tested possessed 6 subsites for binding (1,4)- β -linked glucosyl residues. The relative hydrolytic rates of HvBII and Os3BGlu7 with (1,4)- β -linked mannooligosaccharides suggested that (1,4)- β -linked mannosyl residues could also bind up to 6 subsites. However the complexity of the mannose assay prevented a full kinetic analysis of Os3BGlu7 and characterization of the other rice isoenzymes. The rice isoenzymes that bind (1,4)- β -linked glucosyl residues well at the +2 subsite, are predicted to bind cellobiose best in an unproductive mode that overlaps this site, whereas the strongest binding of cellobiose to a barley enzyme is at the productive position covering the –1 and +1 subsites. Notably, most of the rice isoenzymes are predicted to bind cellobiose best at the +1 and +2 subsites, while Os3BGlu7 would bind most favourably at the +2 and +3 subsites, due to an unusually low affinity at the +1 subsite [26]. During cellotriose hydrolysis, the rice β -D-glucosidases showed higher $k_{\text{cat}}/K_{\text{m}}$ values compared to those for cellobiose hydrolysis, since binding to the –1 to +2 subsites was more favourable.

One residue that might be expected to account for the more favourable binding of the glucosyl residue in the +2 subsite to Os3BGlu7 is Asn245, which is in a position to make a hydrogen bond to this glucosyl residue, and is replaced by Val250 in rHvBII. Indeed, the Met251 in this position in rice Os3BGlu6 β -glucosidase extends into the active site in a position that appears to hinder binding of long oligosaccharides [13]. Mutation of Os3BGlu7 Asn245 to Val resulted in a 2–3-fold increase in the apparent $k_{\text{cat}}/K_{\text{m}}$ for cellobiose and a 15-fold decrease in the apparent $k_{\text{cat}}/K_{\text{m}}$ for cellotriose [21]. However, this change was relatively small and Os7BGlu26 has Val in this position, and had relatively high affinity at the +2 subsite. These observations suggested that the hydrogen bond between Asn245 and the glucosyl residue in the +2 subsite might not be absolutely critical. Other residues that are positioned at the interface between the +1 and +2 subsites may also influence the relative affinities of these two subsites. For instance, the modelling suggested that the replacement of Gln187, Thr189 and Asn190 in Os3BGlu7 with Asn192, Phe194 and His195 may weaken a strong polar interaction from Gln187 in the +2 subsite, while creating an interaction between His195 and the glucosyl residue in the +1 subsite, which could contribute to higher affinity at the +1 subsite than at the +2 subsite in rHvBII and Os7BGlu26. The single change of Os3BGlu7 N190H caused a slight increase in the apparent $k_{\text{cat}}/K_{\text{m}}$ for cellobiose, but the cumulative effect of these changes has yet to be assessed. It must also be remembered that the differences in preference for oligosaccharide substrates between the rice β -D-glucosidases and rHvBII may not be dictated simply by differences in a few amino acid residues on the surface of the active sites, but are also likely to be determined by subtly different shapes and chemical environments generated by the interactions of those residues with neighbouring amino acids.

Although the rice β -D-glucosidases and barley HvBII have similar amino acid sequences and substrate specificities with cello- and laminarioligosaccharides, they may fulfil different roles in plants. Barley HvBII appears to act in remodelling of cell walls during germination by breaking-down oligosaccharides that are released from β -D-mannans and β -D-glucans [2,14,28,29]. The rice β -D-glucosidase isoenzymes may have similar roles during germination, since mannose-containing polysaccharides make up to 17% of rice endosperm based on dry weight [39]. However, Os3BGlu7 is the only enzyme thus far reported to be expressed in germinating rice seed, while transcripts of Os7BGlu26, which showed the highest β -D-mannosidase activity and could efficiently hydrolyse (1,4)- β -D-mannooligosaccharides, are found in seedling shoots, leaves, panicle [9], and fertilised embryos of developing seeds (<http://www.ncbi.nlm.nih.gov/UniGene/clust.cgi?ORG=Os&CID=20617>). The β -D-mannosidases rHvBII and Os7BGlu26 have significant β -D-glucosidase activities against cellooligosaccharides and natural glucosides, while the rice Os3BGlu7 β -D-glucosidase has significant β -D-mannosidase activity [12]. Therefore, they may have overlapping roles in cell wall hydrolysis and recycling, though their efficiencies for hydrolysis of specific oligosaccharides *in vivo* may differ.

Conclusion

In summary, we expressed and characterized two rice enzymes, Os3BGlu8 and Os7BGlu26 that are closely related to Os3BGlu7 and plant β -D-mannosidases, including the barley β -D-mannosidase, rHvBII. The two newly characterized rice β -D-glucosidase isoenzymes, hydrolysed short laminarioligosaccharides (DP of 2–3) and cellooligosaccharides with DP of up to 6, 4NPGlc, and 4NPMann, in line with predictions from sequence analyses and comparative molecular modelling. However, none of the rice isoenzymes had high cellobiose activity, despite the similarity of the Os7BGlu26 sequence and active site model to those of rHvBII, which has a relatively high cellobiose activity. This appeared to be due to differences in the binding of β -1,4-linked glucosyl residues at the +2 subsite, which could not be predicted from the model. Thus, this work points to the importance of verifying substrate specificities and biological functions of proteins identified by genomic studies [40].

Acknowledgments

Dr. Chantragan Srisomsap is thanked for mass spectrometry of rice Os3BGlu8. This work was supported by the Thailand Research Fund (TRF) [Grant BRG5080007], the Thailand National Center for Genetic Engineering and Biotechnology, and the Australian Research Council and the South Australian State Government. TK was sponsored by The Institute for the Promotion of Teaching Science and Technology of Thailand, and SL was supported by the TRF Royal Golden Jubilee PhD Fellowship Program.

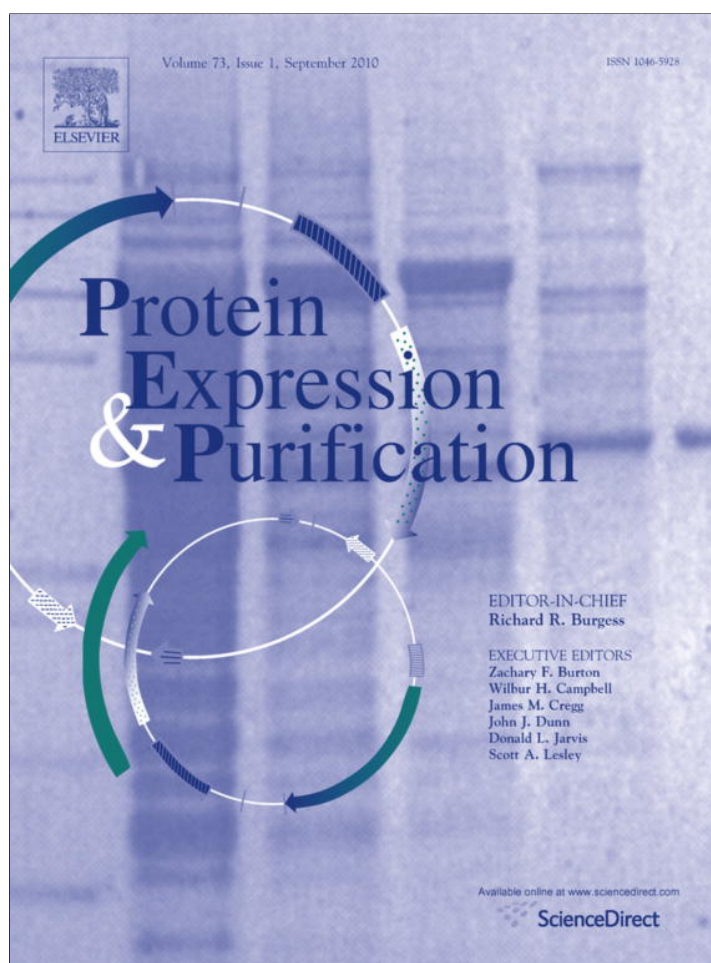
Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.abb.2009.09.004](https://doi.org/10.1016/j.abb.2009.09.004).

References

- [1] C. Gerardi, F. Blando, A. Santino, G. Zacheo, *Plant Sci.* 160 (1996) 795–805.
- [2] M. Hrmova, A.J. Harvey, J. Wang, N.J. Shirley, G.P. Jones, B.A. Stone, P.B. Høj, G.B. Fincher, *J. Biol. Chem.* 271 (1996) 5277–5286.
- [3] L. Barleben, S. Panjikar, M. Ruppert, J. Koepke, J. Stöckigt, *Plant Cell* 19 (2007) 2886–2897.
- [4] K.-L. Dietz, A. Sauter, K. Wichert, D. Messdaghi, W. Hartung, *J. Exp. Biol.* 51 (2000) 937–944.

- [5] B.L. Cantarel, P.M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B. Henrissat, *Nucleic Acids Res.* 37 (2009) D233–D238.
- [6] B. Henrissat, I. Callebaut, S. Fabrega, P. Lehn, J.P. Mornon, G. Davies, *Proc. Natl. Acad. Sci. USA* 92 (1995) 7090–7094.
- [7] J. Jenkins, L. Lo Leggio, G. Harris, R. Pickersgill, *FEBS Lett.* 362 (1995) 281–285.
- [8] M. Hrmova, G.B. Fincher, *Carbohydrate Res.* 342 (2007) 1613–1623.
- [9] R. Opasiri, B. Pomthong, T. Onkoksoong, T. Akiyama, A. Esen, J.R. Ketudat Cairns, *BMC Plant Biochem.* 6 (2006) 33.
- [10] J. Schlieman, *J. Plant, Physiology* 116 (1984) 123–132.
- [11] T. Akiyama, H. Kaka, N. Shibuya, *Phytochemistry* 48 (1998) 49–54.
- [12] R. Opasiri, J.R. Ketudat Cairns, T. Akiyama, O. Wara-aswapati, J. Svasti, A. Esen, *Plant Sci.* 165 (2003) 627–638.
- [13] S. Seshadri, T. Akiyama, R. Opasiri, B. Kuaprasert, J.R. Ketudat Cairns, *Plant Physiol.* 151 (2009) 47–58.
- [14] R. Leah, J. Kigel, I. Svendsen, J. Mundy, *J. Biol. Chem.* 270 (1995) 15789–15797.
- [15] Z. Xu, L.L. Escamilla-Trevino, L. Zeng, M. Lalgordar, D.R. Bevan, S.J. Winkel, A. Mohamed, C.-L. Ceng, M.-C. Shih, J.E. Poulton, A. Esen, *Plant Mol. Biol.* 55 (2004) 343–367.
- [16] B. Mo, J.D. Bewley, *Planta* 215 (2002) 141–152.
- [17] S. Kikuchi, K. Satoh, T. Nagata, N. Kawagashira, K. Doi, N. Kishimoto, J. Yazaki, M. Ishikawa, H. Yamada, *Science* 301 (2003) 376–379.
- [18] J.D. Bendtsen, H. Nielsen, G.V. Heijne, S. Brunak, *J. Mol. Biol.* 340 (2004) 783–795.
- [19] J.D. Thompson, D.G. Higgin, T.J. Gibson, *Nucleic Acids Res.* 22 (1994) 4680.
- [20] R.D.M. Page, *Comp. Appl. Biosci.* 12 (1996) 357–358.
- [21] W. Chuenchor, S. Pengthaisong, R.C. Robinson, Y. Jirundon, A. Esen, C.-J. Chen, R. Opasiri, J. Svasti, J.R. Ketudat Cairns, *J. Mol. Biol.* 377 (2008) 1200–1215.
- [22] A. Sali, T.L. Blundell, *J. Mol. Biol.* 234 (1993) 779–815.
- [23] R.A. Laskowski, M.W. MacArthur, D.S. Mss, J.M. Thornton, *J. Appl. Crystallogr.* 26 (1993) 283–291.
- [24] P. Güntert, V. Dötsch, G. Wider, K. Wüthrich, *J. Biomol. NMR* 2 (1992) 619–629.
- [25] J. Sambrook, P. MacCallum, D. Russel, *Molecular Cloning: A Laboratory Manual*, third ed., Cold Spring Harbor Laboratory Press, 2001.
- [26] R. Opasiri, Y. Hua, O. Wara-Aswapati, T. Akiyama, J. Svasti, A. Esen, J.R. Ketudat Cairns, *Biochem. J.* 379 (2004) 125–131.
- [27] M. Hrmova, V. Farkas, A.J. Harvey, J. Lahnstein, B. Wischmann, N. Kaewthai, I. Ezcurra, T.T. Terri, G.B. Fincher, *FEBS J.* 276 (2009) 437–456.
- [28] M. Hrmova, R.A. Burton, P. Biely, J. Lahnstein, G.B. Fincher, *Biochem. J.* 399 (2006) 77–90.
- [29] M. Hrmova, E.A. MacGregor, P. Biely, R.J. Stewart, G.B. Fincher, *J. Biol. Chem.* 273 (1998) 11134–11143.
- [30] K. Hiromi, Y. Nitta, C. Namura, S. Ono, *Biochim. Biophys. Acta* 302 (1973) 362–375.
- [31] E.J. Stewart, F. Aslund, J. Beckwith, *EMBO J.* 17 (1998) 5543–5550.
- [32] L. Verdoucq, M. Czjzek, J. Morinieré, D.R. Bevan, A. Esen, *J. Biol. Chem.* 278 (2003) 25055–25062.
- [33] J. Zouhar, J. Vévodová, J. Marek, J. Damborský, X.-D. Su, B. Brzobohatý, *Plant Physiol.* 127 (2001) 973–985.
- [34] R. Dopitová, P. Mazura, L. Janda, R. Chaloupková, P. Jerábek, J. Damborský, F. Tomas, N.S. Kiran, B. Brzobohatý, *FEBS J.* 275 (2008) 6123–6135.
- [35] L. Verdoucq, J. Morinieré, D.R. Bevan, A. Esen, A. Vasella, B. Henrissat, M. Czjzek, *J. Biol. Chem.* (2004) 31796–31803.
- [36] M. Czjzek, M. Cicek, V. Zamboni, D.R. Bevan, B. Henrissat, A. Esen, *Proc. Natl. Acad. Sci.* 97 (2000) 13555–13560.
- [37] G.J. Davies, V.M.A. Ducros, A. Varrot, D.L. Zechel, *Biochem. Soc. Trans.* 31 (2003) 523–527.
- [38] M.M. Meagher, B.Y. Tao, J.M. Chow, P.J. Reilly, *Carbohydr. Res.* 173 (1988) 273–283.
- [39] N. Shibuya, A. Misaki, *Agric. Biol. Chem.* 42 (1978) 2267–2274.
- [40] M. Hrmova, G.B. Fincher, in: D. Somers, P. Langridge, P. Fustafson (Eds.), *Methods in Molecular Biology: Plant Genomics*, Humana Press, 2008.
- [41] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, *Proc. Natl. Acad. Sci.* 98 (2001) 10037–10041.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

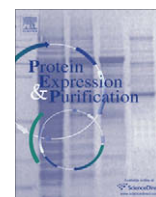
<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep



High-level expression of barley β -D-glucan exohydrolase HvExol from a codon-optimized cDNA in *Pichia pastoris*

Sukanya Luang^a, Maria Hrmova^{b,*}, James R. Ketudat Cairns^{a,**}

^a School of Biochemistry, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand

^b Australian Centre for Plant Functional Genomics, The University of Adelaide, Waite Campus, Glen Osmond, SA 5064, Australia

ARTICLE INFO

Article history:

Received 13 March 2010
and in revised form 12 April 2010
Available online 18 April 2010

Keywords:

Catalytic properties
Hordeum vulgare
N-deglycosylation
Substrate specificity
Transglycosylation

ABSTRACT

The native β -D-glucan exohydrolase isoenzyme Exol from barley seedlings, designated HvExol, was the first GH3 glycoside hydrolase, for which a crystal structure was determined. A precise understanding of relationships between structure and function in this enzyme has been gained by structural and enzymatic studies. To allow testing of hypotheses gained from these studies, an efficient system for expression of HvExol in *Pichia pastoris* was developed using a codon-optimized cDNA. Protein expression at a temperature of 20 °C yielded a recombinant enzyme, designated rHvExol, which had molecular masses of 70–110 kDa due to heavy glycosylation at Asn221, Asn498 and Asn600, the three sites of *N*-glycosylation in native HvExol. Most of the *N*-linked carbohydrate could be removed from rHvExol, resulting in *N*-deglycosylated rHvExol with a substantially decreased molecular mass of 67 kDa. rHvExol was able to hydrolyse barley (1,3;1,4)- β -D-glucan, laminarin and lichenans. The catalytic efficiency value k_{cat}/K_M of rHvExol with barley (1,3;1,4)- β -D-glucan was similar to that reported for native HvExol. Further, laminaribiose, cellobiose and gentiobiose were formed through transglycosylation reactions with 4-nitrophenyl β -D-glucoside and barley (1,3;1,4)- β -D-glucan. Overall, the biochemical properties of rHvExol were similar to those reported for native HvExol, although differences were seen in thermostabilities and hydrolytic rates of certain β -linked glucosides.

© 2010 Elsevier Inc. All rights reserved.

Introduction

Glycoside hydrolases (EC 3.2.1.) are widely distributed in living organisms. These enzymes hydrolyse glycosidic linkages between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety. Based on amino acid sequence similarities, catalytic mechanisms and structural features, the glycoside hydrolase family GH3 is one of 115 glycoside hydrolase families currently listed in the CAZy database (<http://www.cazy.org/>) [1]. The GH3 enzymes are more frequently represented in bacteria, plants and fungi, than in archaea and mammals. The GH3 family includes catalytic proteins with β -D-glucosidase (EC 3.2.1.21), xylan 1,4- β -D-xylosidase (EC 3.2.1.37), β -N-acetylhexosaminidase (EC 3.2.1.52), exo- β -D-glucanase (EC 3.2.1.-) and α -L-arabinofuranosidase (EC 3.2.1.55) activities. The predicted functions of the GH3 enzymes involve: (i) the biodegradation and assimilation of oligo- and polysaccharides [2–5], (ii) modification of bacterial macrolide antibiotics and other toxic plant compounds [6,7], and (iii) turnover of cell wall components [8–15]. The biochemical and bio-

physical properties of various GH3 enzymes have previously been described [16–29]. Some GH3 enzymes exhibit broad substrate specificity, such as the β -D-glucosidases BGL1 from *Pichia etchellsii* and *Saccharomycopsis fibuligera* [30,31], Gbg1 from *Agrobacterium tumefaciens* [32], and bglB from *Thermotoga neapolitana* [33] and enzymes with α -L-arabinofuranosidase and β -D-xylosidase activities, such as ARA-I/XYL from barley (*Hordeum vulgare*), XarB from *Thermoanaerobacter ethanolicus* and MsXyl1 from alfalfa roots [34–36]. Therefore, the identification of natural substrates of GH3 enzymes based on their hydrolytic reactions, and assignments of the biological functions of these enzymes are often highly conjectural.

One of the most intensely studied enzymes in the GH3 family is a barley β -D-glucan exohydrolase, isoform Exol [37], here designated as HvExol. Based on substrate specificity and gene expression studies, it has been suggested that this enzyme might be involved in the turnover or modification of cell walls during the elongation of coleoptiles [9,15]. The enzyme is able to hydrolyse a variety β -D-glucosidic linkages [37,38]. HvExol was the first GH3 enzyme for which a crystal structure was determined [39]. The 3D structure consists of an NH₂-terminal (β/α)₈ barrel domain and a COOH-terminal (α/β)₆ sandwich domain [39]. Two catalytic amino acid residues were identified in the active site and these are the catalytic nucleophile Asp285, located in the conserved

* Corresponding author. Fax: +61 883037102.

** Corresponding author. Fax: +66 44224185.

E-mail addresses: maria.hrmova@adelaide.edu.au (M. Hrmova), cairns@sut.ac.th (J.R. Ketudat Cairns).

SDW motif of the first domain, and the acid/base Glu491, positioned in the second domain [40,41]. The active site of HvExol is located at the interface between the two domains and a glucose molecule was found at the –1 subsite in the native crystal structure, bound predominantly through hydrogen bonds to charged amino acid residues [39].

The catalytic mechanism and structural basis of substrate specificity of HvExol have been investigated by kinetic and crystallographic studies with substrate analogues and inhibitors [41–44]. Based on the premise that enzymes in the same family possess similar structures and catalytic mechanisms, HvExol was also used as a structural model to predict the architectures of other GH3 members [45]. The catalytic mechanism of other GH3 enzymes has also been studied by site-directed mutagenesis with enzymes for which a crystal structure has yet to be determined, such as a β -D-glucosidase from *Flavobacterium meningosepticum* [46,47]. However, similar mutagenesis studies with HvExol have yet to be conducted due to lack of an appropriate recombinant expression system. It is expected that the catalytic mechanism of HvExol may be more precisely understood, if the roles of active site residues suggested from the structural studies could be confirmed by a site-directed mutagenesis approach.

Although many plant glycoside hydrolases have been expressed in bacteria, not every glycoside hydrolase could be expressed in a prokaryotic system in an active form. The yeast *Pichia pastoris* has proven to be an effective eukaryotic host in many other instances (e.g. [48,49]). *P. pastoris* is a methylotrophic yeast, which can produce large amounts of recombinant proteins by methanol induction of the alcohol oxidase 1 (AOX) promoter, and has been shown to effectively synthesize eukaryotic post-translationally modified proteins [50,51]. *P. pastoris* has similar molecular genetics to *Saccharomyces cerevisiae*, but researchers have achieved much higher yields of plant glycoside hydrolases, such as barley α -amylase [48] and Thai rosewood β -D-glucosidase [52,53], in *P. pastoris* than *S. cerevisiae*.

Here, we report the recombinant expression of the HvExol iso-enzyme (rHvExol) from barley in various expression systems. This is the first successful recombinant expression of an active form of a plant enzyme from the GH3 family that as of April 2010 contains nearly 3000 entries. We further describe the substrate specificity and biochemical properties of rHvExol expressed in *P. pastoris*, and compare these characteristics with those of the native HvExol enzyme.

Materials and methods

Chemicals, reagents and expression plasmids

The substrates 4-nitrophenyl- β -D-glucopyranoside (4NPGlc)¹, 4-nitrophenyl- β -D-galactopyranoside, 4-nitrophenyl- β -D-xylopyranoside, 4-nitrophenyl- β -D-fucopyranoside, 4-nitrophenyl- α -L-arabinopyranoside, laminarin (*Laminaria digitata*), orcinol, bovine serum albumin (BSA) and the glucose diagnostic kit were purchased from Sigma (St. Louis, MO, USA). The sources of other substrates have been described previously [38]. For recombinant plasmid construction, the previously described bacterial pET32a/DEST [54] and yeast pPICZ α BNH₈ [53] expression vectors were used, while the pPICZ α BNH₈/DEST vector was created by inserting the Gateway conversion cassette C from Invitrogen (Carlsbad, CA, USA) into the SnaBI site in pPICZ α BNH₈.

¹ Abbreviations used: 4NP, 4-nitrophenyl; 4NPGlc, 4-nitrophenyl- β -D-glucopyranoside; BSA, bovine serum albumin; MALDI-ToF, matrix-assisted laser-desorption ionization-time of flight; IMAC, immobilized metal affinity chromatography; SP-Sepharose, sulphopropyl-Sepharose; TLC, thin-layer chromatography; GH3, glycoside hydrolase family 3; HvExol, native barley (*Hordeum vulgare*) β -D-glucan exohydrolase isoform Exol; rHvExol, recombinant barley β -D-glucan exohydrolase isoform Exol.

Recombinant plasmid construction

A full-length HvExol cDNA from barley seedlings [55] was used as a template to amplify the native cDNA region encoding the mature HvExol (GenBank Accession No. AF102868) with *Pfu* DNA polymerase and the ExoIMatF (5'-CACCGACTACGTGCTCTACAAGGA-3') and ExoIstopR (5'-CTAGTACTTCTTCGTCGCGTTGGT-3') primers. The exoglucanase I fragment was cloned into the pENTR™/D-TOPO Gateway® system entry vector (Invitrogen) according to the manufacturer's instructions. The cDNA of HvExol was transferred to the pET32a/DEST and pPICZ α BNH₈/DEST vectors by LR clonase recombination reactions (Invitrogen).

The HvExol cDNA was codon-optimized for expression in *P. pastoris* and synthesized by GenScript (Piscataway, NJ, USA). The optimized cDNA (GenBank Accession No. GU441535) was cloned into the pPICZ α BNH₈ expression vector [53], between the PstI and EcoRI restriction sites by standard methods [56]. The recombinant bacterial clones were selected on a 25 μ g/mL zeocin Lennox broth (LB) plate, the plasmid DNA isolated and the DNA insert corresponding to HvExol were sequenced at MacroGen (Seoul, Korea).

Expression of rHvExol in *Escherichia coli*

The pET32/DEST expression vector containing the cDNA of HvExol was transformed into *E. coli* strain Origami(DE3) (Novagen, Madison, WI, USA). Protein was induced with 0.4 mM isopropyl thiogalactoside (IPTG) at 20 °C for 16–18 h, and extracted as previously described [57]. Protein expression was detected by measuring release of 4-nitrophenol from 4NPGlc. Refolding of rHvExol from the insoluble fraction was performed with the iFOLD protein refolding system 2 kit (Novagen), following the protocol provided by the manufacturer. In addition, inclusion bodies were washed with wash buffer containing 0.5% (v/v) TritonX-100, 100 mM NaCl and 0.1% (w/v) sodium azide in 50 mM Tris-HCl, pH 7.4. Inclusion bodies were solubilized with the denaturing buffer (6 M guanidine-HCl, 50 mM Tris-HCl, pH 7.4, 100 mM NaCl, 10 mM EDTA and 10 mM DTT) and rHvExol was refolded by dialysis into the refolding buffer with gradually decreasing concentrations of guanidine-HCl in 10 mM Tris-HCl, pH 7.4, 10% (v/v) glycerol, 1 mM reduced glutathione and 0.1 mM oxidized glutathione.

Expression of rHvExol in *P. pastoris*

The pPICZ α -HvExol plasmids were linearized with PmeI and transformed into *P. pastoris* strains Y11430 and SMD11680H (Invitrogen) by electroporation. Colonies were grown on YPDS agar plates containing 100 μ g/mL zeocin, selected again on plates containing 500 μ g/mL zeocin and screened for protein production, as suggested by the manufacturer (Invitrogen).

A single transformed colony was inoculated into 500 mL BMGY medium (Invitrogen) containing 100 μ g/mL zeocin and grown at 28 °C with shaking (160–200 rpm) until the culture reached an OD₆₀₀ of 2–3. Cells were harvested by centrifugation (3000g, 5 min, 20 °C) and resuspended in the BMMY medium (Invitrogen) to reach OD₆₀₀ of 1. Expression of rHvExol was induced with 1% (v/v) methanol for 4 days at 20 °C.

Purification of rHvExol

The culture broth with secreted rHvExol was supplemented with phenyl methyl sulfonyl fluoride (PMSF) to 1 mM and the pH was adjusted to 4.7 on ice with concentrated acetic acid. The protein solution was loaded onto an SP-Sepharose cation-exchange column at a flow rate of 1.8 mL/min. The column was washed with 50 mM sodium acetate, pH 4.7, at a flow rate of 0.5 mL/min, and protein was eluted with a linear gradient of 0–2 M NaCl in

50 mM sodium acetate, pH 4.7, at a flow rate of 1 mL/min. The active fractions were concentrated and resuspended in 50 mM sodium phosphate, pH 7.8, containing 300 mM NaCl (IMAC buffer). IMAC purification of rHvExol was performed by mixing rHvExol with Talon Co²⁺-bound IMAC resin (Clontech, Mountain View, CA, USA), and then eluting the enzyme with a 0–0.5 M imidazole gradient in the IMAC buffer at a flow rate of 0.5 mL/min. The active fractions were reconstituted in 20 mM sodium acetate buffer, pH 5.0, by centrifugal filtration (Vivaspin, 10 kDa exclusion limit). Typically, the yield of nearly homogenous rHvExol (as judged by SDS-PAGE and immunoblot analyses) was around 4 mg per liter of culture.

N-deglycosylation of rHvExol

Purified rHvExol (60–100 µg) was deglycosylated by 500 U endoglycosidase H (New England BioLabs, Ipswich, MA, USA) in 50 mM sodium citrate buffer, pH 5.5. The mixture was incubated at 4 °C for 3–4 days with gentle shaking. Finally, deglycosylated rHvExol was purified through a 2nd IMAC column, as described above, to remove endoglycosidase H. After IMAC, imidazole was removed by centrifugal concentration, as described above and a nearly homogenous rHvExol was suspended in 20 mM sodium acetate, pH 5.25.

Tryptic mapping of rHvExol by MALDI-ToF/ToF spectrometry

About 10 µg of rHvExol were S-amidomethylated, digested with 100 ng sequencing grade trypsin (Promega, Madison, WI, USA) in 5 mM ammonium bicarbonate and concentrated to 5 µL. A 0.5 µL aliquot of the digest was applied to a 600 µm AnchorChip (Bruker Daltonik GmbH, Bremen, Germany) according to the α -cyano-4-hydroxycinnamic acid (HCCA, Bruker Daltonik) thin-layer method. MALDI-ToF mass spectra were acquired with a Bruker Ultraflex III MALDI-ToF/ToF mass spectrometer (Bruker Daltonik GmbH) operating in reflectron mode under the control of FlexControl, version 3.0 (Bruker Daltonik GmbH). All spectrometry data were processed with FlexAnalysis, version 3.1 (Bruker Daltonik GmbH), and the spectra and mass lists were exported to BioTools (Bruker Daltonik GmbH). The MS and corresponding MS/MS spectra were combined and submitted to a Mascot database-search.

Molecular mass determination of rHvExol via MALDI spectrometry

Samples of 1–3 µL of glycosylated and N-deglycosylated rHvExol were mixed with 1 µL matrix solution (10 mg/mL sinipinic acid in 90% acetonitrile/0.1% trifluoroacetic acid) and applied to a 600 µm AnchorChip target plate (Bruker Daltonik GmbH, Bremen, Germany). MALDI-ToF mass spectra were acquired on a Bruker Ultraflex III MALDI-ToF/ToF mass spectrometer (Bruker Daltonik GmbH) operating in linear mode under the control of the FlexControl software (Version 3.0, Bruker Daltonik GmbH). External calibration was performed with a mix of ClinProt Protein Calibration Standard and Protein Calibration Standards 2 (Bruker Daltonik GmbH), over a range of 1–25 kDa that were analysed under the same conditions. Spectra were obtained at various locations over the surface of the matrix spot.

rHvExol characterization

Enzyme assays and analyses of hydrolytic and transglycosylation activities, substrate specificity, kinetic properties, pH optimum and thermostability were performed as described previously [38]. The kinetic parameters were determined by a proportional weighted fit, with a nonlinear regression analysis program, based on Michaelis–Menten kinetics [58].

Protein determination, NH₂-terminal sequencing, SDS-PAGE and immunoblot analyses

Protein determination, SDS-PAGE and protein detection on SDS-PAGE gels were performed as described [34]. Immunoblot analyses were performed with 0.22 µm nitrocellulose blotting membranes (Millipore, Billerica, MA, USA), a mouse monoclonal anti-poly-Histidine-alkaline phosphatase IgG2a isotype antibody and the BCIP/NBT-purple liquid reagent for membranes, as suggested by the manufacturer (Sigma).

Results

Expression of rHvExol

A construct in which the HvExol cDNA was inserted in pET32a/DEST was first used to express an NH₂-terminal thioredoxin-His₆-tagged fusion protein in *E. coli* strain Origami(DE3) cells, as this system has been successfully used for expression of other plant glycoside hydrolases [54,57,59–61]. Under the conditions tested, the rHvExol protein was observed only in the insoluble fraction upon cell extraction and no enzyme activity was detected (data not shown). Therefore, refolding of rHvExol was attempted after solubilization in guanidine-HCl, but again the enzyme was found to be inactive under any refolding conditions tested.

Expression of rHvExol in yeast was tested by inserting native cDNA into the pPICZ α BNH₈/DEST expression vector, to produce a protein fused to the α -factor prepropeptide for secretion. *P. pastoris* (Y11430) was tested as a host to produce rHvExol. No increase in β -D-glucosidase (with 4-nitrophenyl β -D-glucopyranoside, 4NPGlc) or exoglucanase (with barley (1,3;1,4)- β -D-glucan) activities were observed and the protein could not be detected on Coomassie-stained SDS-PAGE gels in the predicted mass range (data not shown). Since this construct contains an eight-histidine tag at the NH₂-terminus of rHvExol (AHHHHHHHHAA) after the secretory peptide cleavage site, the rHvExol could be detected by immunoblot analysis with anti-polyhistidine antibody. Here, a 43 kDa band was detected, which could be correlated to the size of domain 1 of rHvExol (data not shown). This analysis indicated that either rHvExol was produced as a full-length protein and then proteolytically degraded by an unspecified protease from the host, or that premature termination of translation occurred near the end of domain 1, possibly because the native barley cDNA contained codons of low usage in *P. pastoris*. To avoid the problems of proteolysis and poor codon usage for protein synthesis in *P. pastoris*, the HvExol cDNA was codon-optimized and expressed in the protease-deficient *P. pastoris* strain SMD1168H.

The native and codon-optimized rHvExol cDNA fusions were transformed into *P. pastoris* strain SMD1168H and protein expression was induced at a temperature of 20 °C. High levels of HvExol activity, measured with 4NPGlc, were detected from the codon-optimized rHvExol cDNA fusion and a series of bands from 75 to 85 kDa were detected by immunoblot analysis with anti-polyhistidine antibody (Fig. 1A). The activity in the media slowly increased until 4 days of induction, after which it did not change. However, little increase in rHvExol activity and protein production (as determined by SDS-PAGE and Coomassie-staining) were observed from the cDNA fusions containing codon-optimized HvExol that were induced at a temperature of 28 °C, or from a construct containing the native HvExol cDNA induced at any temperature (data not shown).

rHvExol purification

After rHvExol was expressed in *P. pastoris* from the codon-optimized cDNA at 20 °C for 4 days, a set of major protein bands, with

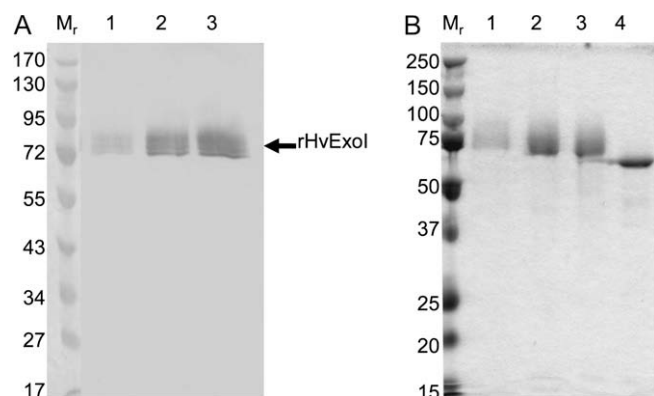


Fig. 1. Protein production of rHvExol detected by immunoblot analysis (A) and SDS-PAGE (B). (A) Expression of rHvExol in *P. pastoris* was detected via anti-His-tag antibodies by immunoblot analysis. *M_r* is a prestained protein marker, lanes 1–3 indicate levels of rHvExol production after 1, 2 and 3 days, respectively. (B) SDS-PAGE of rHvExol from crude broth (lane 1), and purified via SP-Sepharose chromatography (lane 2), the 1st IMAC step (lane 3), and the 2nd IMAC step, after deglycosylation with endoglycosidase H (lane 4).

apparent masses corresponding to the molecular mass of rHvExol and higher, was detected by immunoblot analysis and SDS-PAGE with Coomassie staining (Fig. 1B). This protein was purified from the culture media by SP-Sepharose chromatography (Table 1). Protein was eluted with 1.2 M NaCl in 50 mM NaOAc, pH 4.7. The fractions containing narrow ranges of activity near the Gaussian peak (based on activity, absorbance at 280 and SDS-PAGE profiles) were concentrated and purified by Immobilized Metal Affinity Chromatography (IMAC), which bound rHvExol containing the His₈-tag. A small amount of exoglucanase activity, which might have originated from the presence of rHvExol without a His₈-tag, passed through the column. A portion of the activity that passed through the column may also be attributed to a native *P. pastoris* (1,3)- β -D-glucan exohydrolase [62]. Nevertheless, most of the expressed rHvExol was eluted from the column with 250 mM imidazole and the specific activity of rHvExol was increased by about 4-fold after IMAC purification (Table 1). It has been reported that native HvExol has three *N*-glycosylation sites at Asn221, Asn498 and Asn600 [39]. To test whether the bands detected in the range between 80 and 100 kDa (Fig. 1B) were due to the presence of over-*N*-glycosylated forms of rHvExol, and whether these *N*-glycosylation sites could affect enzyme properties, the *N*-linked carbohydrate was removed by endoglycosidase H. After *N*-deglycosylated rHvExol was purified by a second round of IMAC, the specific activity remained more-or-less unchanged (Table 1). The amino acid sequence of *N*-deglycosylated rHvExol was confirmed by mass spectrometry (Table 2; *vide infra*).

Table 1
Enzyme yields during purification of rHvExol.

Purification step	Yield ^a		Specific activity ^a (units mg ⁻¹)	Recovery ^{a,c} (%)	Purification factor ^{a,d} (fold)
	Protein (mg)	Activity ^b (units)			
Crude protein	119	98.7	0.8	100	1
SP-Sepharose	18.9	41.3	2.2	41.9	2.6
1st IMAC	3.3	11.7	3.5	11.8	4.2
2st IMAC	2.9	9.0	3.1	9.1	3.7

^a The numbers were rounded to the nearest decimal place.

^b Recovered enzyme units assayed on 4NPGlc.

^c Expressed in percent of enzyme activity in the crude extract.

^d Calculated on the basis of specific activity (units/mg protein).

Table 2

Amino acid sequences of tryptic fragments of rHvExol identified by MALDI-ToF/ToF.

Tryptic fragments ^a	
1	MTLAEKIGQMTQIERLVATPDVLRDNFISLLSGGGSVPR
2	GATAKEWQDMVDGFQK
3	RIGETALEVRATGIQYAFAPCIAVCR
4	RIVQSMTELIPGLQGDVPKDFTS GMPFVAGK
5	HFVGDGGTVDGINENNTIINREGLMNIHMPAYKNAMDKGVSTVMISYS SWNGVKMHANQDLVTGYLKDITLKFQGVISDWEGIDRITTPAGSDYSYS VKASILAGLDMIMVPMK
6	RVKFTMGLFENPYADPAMAEQLGK
7	NGKTSTDAPLLPLPK
8	TTVGTITILEAVKAAVDPTSTVVFAENPDA EFVKSGGFSYAIVAVGEHPYTETKGDNLNLTIPPEGLSTVQAVCGGVR
9	SVDQLPMNVGDAHYDPLFRLGYGLTTNATK

^a The underlined letters indicate *N*-glycosylated Asn (N) residues.

Tryptic mapping of rHvExol

Purified *N*-deglycosylated rHvExol (Fig. 1B) was digested with trypsin, and the molecular masses of individual tryptic fragments were determined by MALDI-ToF/ToF spectrometry, which provided a coverage of 65% of the ions with expected *m/z* values. Nine peptide sequences (Table 2) deduced from the nucleotide sequences of the pPICZ α -HvExol DNA fusion had predicted *m/z* values that exactly matched those ions that were observed during mass spectrometry analyses. The three peptides containing the putative *N*-linked glycosylation sites had masses that were 203.2 Da higher than those of the predicted peptides. The additional mass of 203.2 Da corresponded to a single *N*-acetyl β -D-glucosaminyl residue, which remained attached to the asparaginyl residues after the remainder of the sugar had been cleaved off by endoglycosidase H.

Molecular mass and NH₂-terminal sequencing of rHvExol

The full-length, His-tagged rHvExol of 67.1 kDa was observed in the zip-tip purified *N*-deglycosylated sample by MALDI-ToF mass spectrometry (Fig. 2A). A doubly charged form of rHvExol was also observed at 33.6 kDa and a very low abundance form of 44.7 kDa was also seen, along with a likely doubly charged form of this species at 22.5 kDa. However, the MALDI-ToF spectra obtained from the protein sample that had not been *N*-deglycosylated were undefined and did not provide clear mass indications. This *N*-glycosylated protein sample, which was expected to contain multiple *N*-glycosylated forms of rHvExol with different molecular masses, resulted in the very low intensity spectra shown in Fig. 2B. Subsequent analyses of these two samples by an electrospray ionisation ion trap mass spectrometry failed to detect any masses above the background ions, therefore NH₂-terminal protein sequencing of *N*-deglycosylated rHvExol was undertaken. The data revealed that the AHHH amino acid sequence was indeed present in rHvExol and that the EAEA signal cleavage sequences had been removed from the NH₂-terminal region of the otherwise full-length rHvExol enzyme [63].

Effect of pH and temperature on enzyme activity

The pH optima of the glycosylated and *N*-deglycosylated forms of rHvExol were determined in McIlvaine buffer over the pH range of 3.5–8.5 at 30 °C (Fig. 3). The pH profiles were bell-shaped, and the highest activities of glycosylated and *N*-deglycosylated rHvExol were detected at pH 5.0, although the pH profile suggested the optimum for the *N*-deglycosylated rHvExol should be 5.25 (Fig. 3). As for the temperature stability of rHvExol, the activities were assayed after incubation at temperatures over the range between 0 and 80 °C for 15 min (Fig. 4). While glycosylated rHvExol

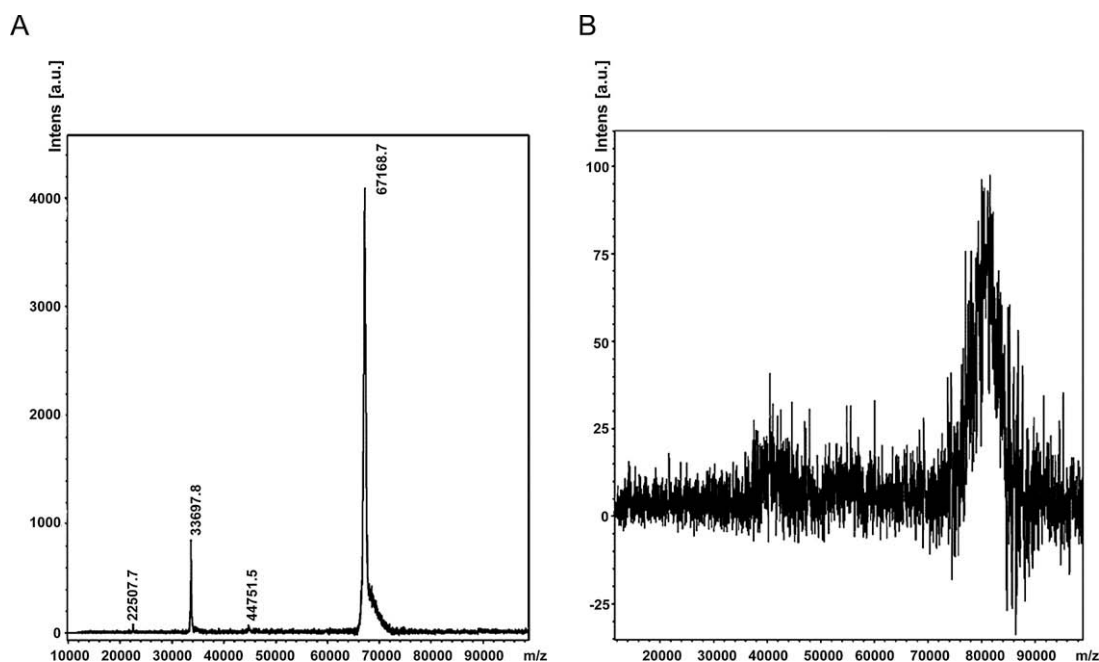


Fig. 2. MALDI-ToF spectra of *N*-deglycosylated (A) and glycosylated (B) rHvExol.

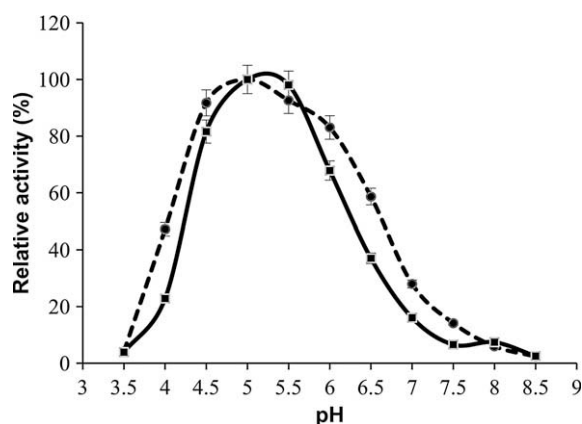


Fig. 3. pH-activity profile of glycosylated (---) and *N*-deglycosylated (—) rHvExol. Activity was assayed with 0.2% (w/v) 4NPGlc, 160 μ g/mL BSA in 0.1 M citric acid–0.2 M disodium hydrogen phosphate (McIlvaine buffers) over the pH range of 3.5–8.5.

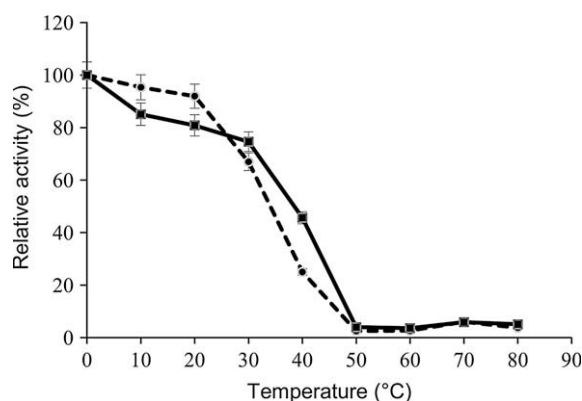


Fig. 4. Thermostability of glycosylated (---) and *N*-deglycosylated (—) rHvExol. rHvExol was assayed with 4NPGlc for 15 min at 30 °C after being incubated at the indicated temperatures between 0 and 80 °C for 15 min.

was stable at the temperature range between 10 and 20 °C, *N*-deglycosylated rHvExol was stable in the range between 10 and 30 °C. Above these temperature ranges, the activities of both rHvExol forms decreased substantially (Fig. 4). The temperatures at which 50% inactivation of the glycosylated and *N*-deglycosylated rHvExol forms were observed, were 34 and 39 °C, respectively. Both rHvExol forms were denatured at 50 °C and no activity was detected at higher temperatures (Fig. 4). The addition of BSA at 160 mg/L could marginally stabilise and protect the rHvExol protein against heat inactivation (data not shown).

Substrate specificity

The substrate specificities of the glycosylated and *N*-deglycosylated rHvExol enzymes were determined towards the polysaccharides laminarin, barley (1,3;1,4)- β -D-glucan and lichenan and the synthetic glycoside 4NPGlc at 0.2% (w/v) concentrations. The data indicated that rHvExol was similar to native HvExol, in that it hydrolysed polysaccharides, including barley (1,3;1,4)- β -D-glucan and laminarin and lichenans from Icelandic moss (*Cetraria islandica*), as well as the aryl- β -D-glucoside substrate 4NPGlc (Table 3). The rHvExol enzyme was specific for glucosides and could not hydrolyse α -L-arabinoside, β -D-galactoside, β -D-xyloside, and β -D-fucoside (data not shown). Comparative analyses of the activities of glycosylated and *N*-deglycosylated rHvExol indicated that both enzymes behaved similarly with laminarin and barley (1,3;1,4)- β -D-glucan, although glycosylated rHvExol had higher activity with 4NPGlc than the *N*-deglycosylated form of rHvExol (Table 3).

Kinetic and inhibition parameters

The kinetic parameters for hydrolysis of laminarin, barley (1,3;1,4)- β -D-glucan, and 4NPGlc were determined for both glycosylated and *N*-deglycosylated rHvExol (Table 4). Both forms were highly active on polysaccharides and the most efficient polymeric substrate was barley (1,3;1,4)- β -D-glucan, which was hydrolysed approximately 1.5- to 2-fold more efficiently, in terms of the k_{cat}/K_M catalytic efficiency values than laminarin. The hydrolysis of barley (1,3;1,4)- β -D-glucan proceeded about 1.3-fold faster with

Table 3

Relative activities of rHvExol on polysaccharides, oligosaccharides and synthetic substrates.

Substrate	Relative activity (%) ^a		
	Native HvExol ^b	Glycosylated rHvExol	N-deglycosylated rHvExol
<i>Polysaccharides</i>			
Laminarin (<i>L. digitata</i>)	100 ^c	98.9 ± 1.1 ^d	98.8 ± 1.2 ^d
Barley (1,3;1,4)-β-D-glucan	10	10.0 ± 0.5	10.0 ± 0.8
Lichenan (<i>C. islandica</i>)	nm ^e	1.37 ± 0.09	nm
<i>Synthetic substrate</i>			
4-NPGlc	10	22.8 ± 1.4	16.8 ± 0.8

^a The numbers were rounded to three significant figures.

^b The data are from [38].

^c The relative activity of 100% equals to 63 U/mg.

^d The relative activity of laminarin equals to 54.4 and 65.3 U/mg of glycosylated and N-deglycosylated forms of rHvExol, respectively.

^e 'nm' indicates 'not measured'.

Table 4

Kinetic parameters of rHvExol on laminarin, barley (1,3;1,4)-β-D-glucan, and 4NPGlc.

Substrate	Native HvExol ^a	Glycosylated rHvExol ^a	N-deglycosylated rHvExol ^a
<i>Laminarin</i>			
K_M (mM)	0.098 ^b	0.18 ± 0.01	0.22 ± 0.02
k_{cat} (s ⁻¹)	73 ^b	42.1 ± 1.6	52 ± 3
k_{cat}/K_M (mM ⁻¹ s ⁻¹)	740 ^b	234 ± 14	240 ± 20
<i>Barley (1,3;1,4)-β-D-glucan</i>			
K_M (mM)	0.012 ^b	0.044 ± 0.004	0.020 ± 0.002
k_{cat} (s ⁻¹)	4 ^b	14.1 ± 0.7	9.7 ± 1.1
k_{cat}/K_M (mM ⁻¹ s ⁻¹)	330 ^b	365 ± 57	478 ± 50
<i>4NPGlc</i>			
K_M (mM)	1.4 ^b	1.99 ± 0.15	2.0 ± 0.12
k_{cat} (s ⁻¹)	5 ^b	27.7 ± 0.9	25.4 ± 0.7
k_{cat}/K_M (mM ⁻¹ s ⁻¹)	3 ^b	13.9 ± 0.6	12.8 ± 0.8

^a The values were rounded to three significant figures, or to the precision of the error.

^b Kinetic parameters of HvExol with laminarin, barley (1,3;1,4)-β-D-glucan and 4NPGlc are from [38].

the N-deglycosylated rHvExol form, although with the other two substrates, the hydrolytic rates were similar (Table 4). The k_{cat}/K_M catalytic efficiency values of rHvExol were higher than those of native HvExol for barley (1,3;1,4)-β-D-glucan and 4NPGlc, while native HvExol was more efficient at hydrolysing laminarin (Table 4).

Transglycosylation

Both the glycosylated and N-deglycosylated forms of rHvExol possessed both hydrolytic and glucosyltransferase activities toward 4NPGlc at 20 mM concentration. Here, the formation of transglycosylation products was detectable within 3 min of reaction (0.05 h; Fig. 5). The transglycosylation product patterns of N-deglycosylated rHvExol were very similar to those of glycosylated rHvExol shown in Fig. 5. During the initial stages of the reaction at 3 min, a small amount of glucose was observed, in addition to 4NP-gentiobioside and 4NP-laminaribioside. The 4NP-oligosaccharide products were identified by comparison with standard compounds of known R_f values [64]. As the reaction progressed, after 4 h reaction, the transient transglycosylation products laminaribiose, cellobiose, 4NP-laminaribioside, 4NP-cellobioside and an unknown disaccharide (P₁) and trisaccharides (P₂–P₄) were detected. By 18 h, the 4NP-glycosides, laminaribiose and cellobiose were hydrolysed, and only the major hydrolytic product glucose and traces of gentiobiose were observable (Fig. 5). Transglycosylation activity

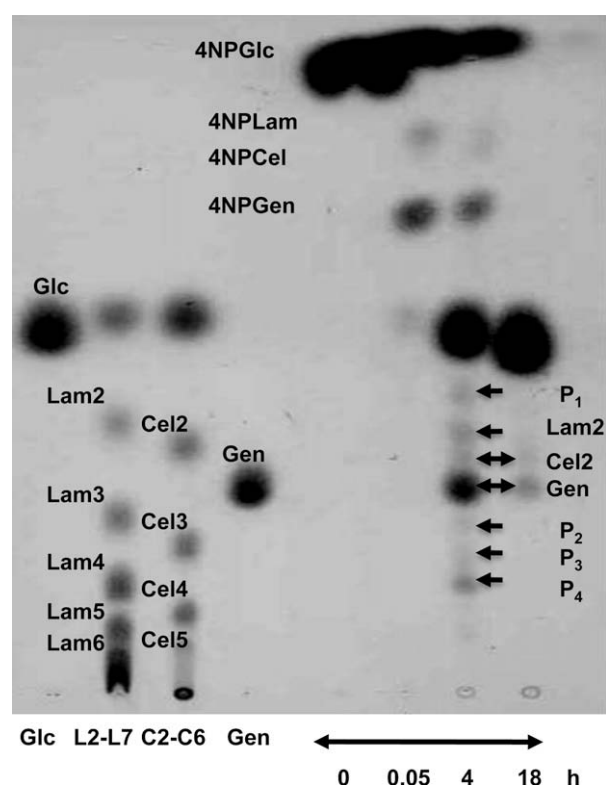


Fig. 5. TLC chromatogram of hydrolysis and transglycosylation products formed by glycosylated rHvExol. The enzyme was incubated in the presence of 20 mM 4NPGlc at 30 °C for 0, 3 min, 4 h and 18 h. Standards are glucose (Glc), laminari-oligosaccharides (L2–L7), cello-oligosaccharides (C2–C6), and gentiobiose (Gen), 4NP-β-laminaribiose (4NPLam), and 4NP-β-cellobiose (4NPCel). The presence of oligosaccharide products with unknown structures are indicated as P₁–P₄. The reaction times are indicated in hours below the lanes.

also occurred with barley (1,3;1,4)-β-D-glucan, from which only the formation of the disaccharide gentiobiose was detected (data not shown). Similar observations were reported for native HvExol [38].

Discussion

Development of an expression system for active rHvExol

Bacteria and yeast are common hosts for production of target proteins. Although, bacterial expression is convenient, many eukaryotic proteins are produced in low protein yields [65,66] or are produced as insoluble aggregates [67–69]. Although HvExol could be expressed in bacteria, it was not able to fold properly, and refolding experiments using various refolding formulations failed to recover active rHvExol. The reasons why rHvExol could not be refolded successfully might include its two-domain organisation and the fact that the enzyme contains 10 cysteine residues. From these residues, two disulfide linkages are formed, between Cys151 and C159 in domain 1, and between Cys513 and Cys518 in domain 2 [39]. Hence, the probability that the correct disulfide linkages will be formed is low. Therefore, a yeast expression system was introduced that was reported to be able to produce plant enzymes in active forms (e.g. [63]).

Prokaryotic and eukaryotic cells have their own species-specific codon usage patterns. Many target genes from mammals and plants are expressed at low levels in bacteria or yeast, in part because the rate of protein translation is not well correlated to codon usage and tRNA bias [70]. The HvExol cDNA amplified from barley

seedlings produced low protein yields in *P. pastoris*. Thus, a codon-optimized cDNA was synthesised. Expression of rHvExol was only successful from the codon-optimized cDNA in a protease-deficient strain of *P. pastoris* and at a low temperature of 20 °C. The expressed rHvExol possessed three post-translationally modified asparaginy residues, similar to native HvExol. rHvExol could be purified by a two-step procedure using a strong cation-exchanger and IMAC with a final yield of 12%, based on the total protein used for purification. The use of selective pooling of the chromatographic fractions, as well as initial contamination with *P. pastoris* exoglucanase [62], which was eliminated during the purification, and loss of a small amount of HvExol that lacked its N-terminal histidine tag may account for the relatively lower apparent final yield of purified enzyme that was around 10–12%. Upon *N*-deglycosylation with endoglycosidase H and purification with a second IMAC step, *N*-deglycosylated rHvExol was produced with a yield of 9%, and the specific activity of rHvExol was nearly the same as before *N*-deglycosylation. Here, the endoglycosidase H was chosen, because it cleaves the chitobiose core of high mannose *N*-linked glycoproteins, which are commonly observed in yeast, while retaining the first *N*-acetylglucosamine residue linked to Asn [71]. A His-tagged form of rHvExol of 67.2 kDa lacking an EAEA secretion motif was obtained after expression in *P. pastoris* and *N*-deglycosylation. Thus, secreted rHvExol was properly processed by the proteases in *Pichia* cells, contrary to previous observations with xyloglucan xyloglucosyl transferase enzymes [63].

Catalytic properties of rHvExol

The recombinant glycosylated and *N*-deglycosylated rHvExol forms were unique in terms of their pH-activity profiles in the range of pH 3.5–8.5 and thermostabilities. The pH optimum and thermostability of *N*-deglycosylated rHvExol were similar to native HvExol, although the same parameters of glycosylated rHvExol were slightly lower. Koseki et al. [72] reported that the *N*-glycosylation of oligosaccharide chain (2.5 kDa) of asparagines at the catalytic domain increased the thermostability of *Aspergillus kawachii* α -L-arabinosidase, whereas glycosylated rHvExol did not exhibit increased stability at elevated temperatures. The molecular mass of glycosylated rHvExol was higher by approximately 7.7–17.7 kDa than that of *N*-deglycosylated enzyme due to occupation of three *N*-glycosylation sites. The calculated molecular mass of HvExol was 66.8 kDa [37]. Based on the sizes of *N*-linked carbohydrates at the *N*-glycosylation sites, we would expect that the presence of these carbohydrates could affect certain enzyme properties, such as pH optimum and thermostability. However, the glycosylated and *N*-deglycosylated rHvExol enzymes had pH optima and thermostabilities quite similar to native HvExol [38].

The polysaccharides laminarin and barley (1,3;1,4)- β -D-glucan and the aryl glycoside 4NPGlc were hydrolysed by rHvExol with similar hydrolytic rates, compared to native HvExol. It has well been documented that native HvExol prefers hydrolysing substrates containing (1,3)- β -linked glucoside residues [37,38,42]. As for the catalytic properties, we found that rHvExol had k_{cat}/K_M values very similar to native HvExol with barley (1,3;1,4)- β -D-glucan, although the k_{cat}/K_M value of rHvExol was higher with 4NPGlc, and about 3-fold lower with laminarin (Table 4). The only obvious differences between the rHvExol and native HvExol enzymes is the presence of the extension at the N-terminus by 11 residues (AHHHHHHHHAA) in rHvExol, and differences in *N*-glycosylation of both forms produced in *P. pastoris*. The 11-residue extension and the Asn600 glycosylation site lie at the N- and C-termini of HvExol, respectively, distant from the active site, however, the Asn221 and Asn498 *N*-linked glycosylation sites lie in the proximity of the active site of HvExol [39,40]. It is possible that the native barley carbohydrate might help in interaction with a polymeric

substrate, such as laminarin, resulting in the much tighter binding seen in the native enzyme. However, it appears the high mannose and *N*-acetyl glucosamine containing carbohydrates affixed to these sites by *P. pastoris* makes no significant net interaction with substrates. It is also possible that the differences in N-terminal sequence and posttranslational modification result in differences in protein flexibility or other protein physical properties that might explain the differences in hydrolysis of these substrates.

Conclusions

Recombinant rHvExol was expressed at high levels from a codon-optimized HvExol cDNA in protease-deficient *P. pastoris*, under low temperature conditions, while the expressions from a native HvExol cDNA in *E. coli* or *P. pastoris* were unsuccessful. rHvExol exhibited hydrolase activities with β -linked polysaccharide and aryl glycoside substrates and produced a variety β -linked oligosaccharide products through transglucosylation activities, similar to the native HvExol enzyme. Thus, the recombinant rHvExol is an appropriate model enzyme to study the roles of amino acid residues in catalysis and substrate specificity, as revealed by numerous crystal structures of HvExol, by site-directed mutagenesis.

Acknowledgments

We are grateful to Dr. Kris Ford (University of Melbourne, Australia), Mark Condina and Alexander Colella (Adelaide Proteomics Centre, the University of Adelaide, Australia) for mass spectrometry, molecular mass and sequencing analyses. Tassanee Onksoong is gratefully acknowledged for generation of the pPICZ α BNH₈/DEST plasmid. This work was supported by grants from the Australian Research Council to MH, the Thailand Research Fund (TRF) Grant BRG5080007 to J.R.K.C. and a TRF Royal Golden Jubilee Ph.D. Fellowship to S.L.

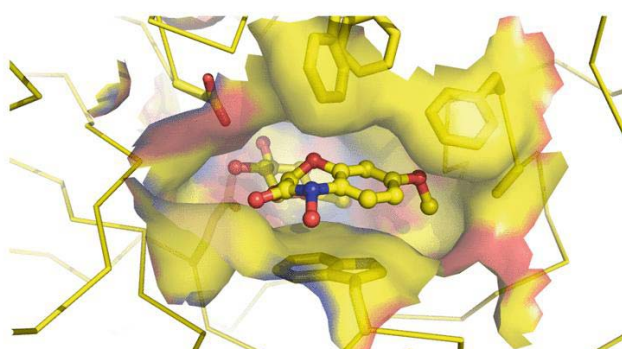
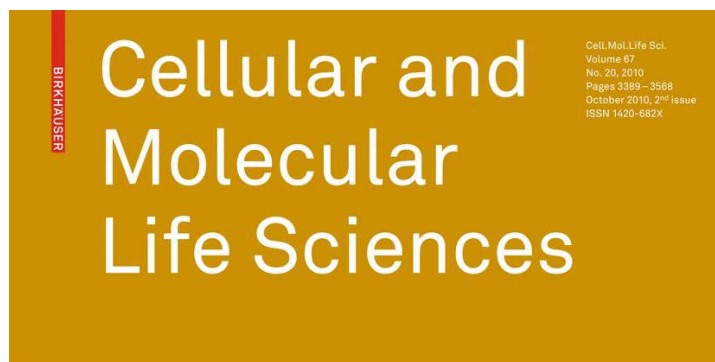
References

- [1] B.L. Cantarel, P.M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B. Henrissat, The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics, *Nucleic Acids Res.* 37 (2009) D233–D238.
- [2] D. Faure, J. Desair, V. Keijers, M.A. Bekri, P. Proost, B. Henrissat, J. Vanderleyden, Growth of *Azospirillum irakense* KBC1 on the aryl β -glucoside salicin requires either *salA* or *salB*, *J. Bacteriol.* 181 (1999) 3003–3009.
- [3] H. Tsujibo, C. Takada, A. Tsuji, M. Kosaka, Cloning, sequencing and expression of the gene encoding an intracellular β -D-xylosidase from *Streptomyces thermoviolaceus* OPC-520, *Biosci. Biotechnol. Biochem.* 65 (2001) 1824–1831.
- [4] S. Shipkowski, J.E. Brenchley, Characterization of an unusual cold-active β -glucosidase belonging to family 3 of the glycoside hydrolases from the psychrophilic isolate *Paenibacillus* sp. strain C7, *Appl. Environ. Microbiol.* 71 (2005) 4225–4232.
- [5] D. Faure, B. Henrissat, D. Ptacek, M.A. Bekri, J. Vanderleyden, The *celA* gene, encoding a glycosyl hydrolase family 3 β -glucosidase in *Azospirillum irakense*, is required for optimal growth on cellobiosides, *Appl. Environ. Microbiol.* 67 (2001) 2380–2383.
- [6] L.M. Quirós, I. Aguirrezabalaga, C. Olano, C. Méndez, J.A. Salas, Two glycosyltransferases and a glycosidase are involved in oleandomycin modification during its biosynthesis by *Streptomyces antibioticus*, *Mol. Microbiol.* 28 (1998) 1177–1185.
- [7] C.R. Wulff-Strobel, D.B. Wilson, Cloning, sequencing, and characterization of a membrane-associated *Prevotella ruminicola* B₄ β -glucosidase with cellodextrinase and cyanoglycosidase activities, *J. Bacteriol.* 177 (1995) 5884–5890.
- [8] Q. Cheng, H. Li, K. Merdek, J.T. Park, Molecular characterization of the β -N-acetylglucosaminidase of *Escherichia coli* and its role in cell wall recycling, *J. Bacteriol.* 182 (2000) 4836–4840.
- [9] M. Hrmova, G.B. Fincher, Structure–function relationships of β -D-glucan endo- and exohydrolases from higher plants, *Plant Mol. Biol.* 47 (2001) 73–91.
- [10] C.Y. Hung, J.J. Yu, P.F. Lehmann, G.T. Cole, Cloning and expression of the gene which encodes a tube precipitin antigen and wall-associated β -glucosidase of *Coccidioides immitis*, *Infect. Immun.* 69 (2001) 2211–2222.
- [11] J.B. Kim, A.T. Olek, N.C. Carpita, Cell wall and membrane-associated exo- β -D-glucanases from developing maize seedlings, *Plant Physiol.* 123 (2000) 471–485.

- [12] H. Crombie, S. Chengappa, A. Hellyer, J.S.G. Reid, A xyloglucan oligosaccharide-active, transglycosylating β -D-glucosidase from the cotyledons of nasturtium (*Tropaeolum majus* L.) seedlings-purification, properties and characterization of a cDNA clone, *Plant J.* 15 (1989) 27–38.
- [13] T. Kotake, N. Nagawa, K. Takeda, N. Sakuria, Purification and characterization of wall-bound exo-1,3- β -D-glucanase from barley (*Hordeum vulgare* L.) seedlings, *Plant Cell Physiol.* 38 (1997) 194–200.
- [14] E. Labrador, D.J. Nevins, An exo- β -D-glucanase derived from *Zea mays* coleoptile walls with a capacity to elicit cell elongation, *Physiol. Plantarum* 77 (1989) 479–486.
- [15] M. Hrmova, G.B. Fincher, Plant and microbial enzymes involved in the depolymerization of (1,3)- β -D-glucans and related polysaccharides, in: A. Bacic, G.B. Fincher, B.A. Stone (Eds.), *Chemistry, Biochemistry and Biology of (1 → 3)- β -Glucans and Related Polysaccharides*, Academic Press, Elsevier Inc., San Diego, USA, 2009, pp. 119–170.
- [16] R. Kawai, K. Igarashi, M. Kitaoka, T. Ishii, M. Samejima, Kinetics of substrate transglycosylation by glycoside hydrolase family 3 glucan (1 → 3)- β -glucosidase from the white-rot fungus *Phanerochaete chrysosporium*, *Carbohydr. Res.* 339 (2004) 2851–2857.
- [17] C. Mayer, D.J. Vocadlo, M. Mah, K. Rupitz, D. Stoll, R.A.J. Warren, S.G. Withers, Characterization of a β -N-acetylhexosaminidase and a β -N-acetylglucosaminidase/ β -glucosidase from *Cellulomonas fimi*, *FEBS J.* 273 (2006) 2929–2941.
- [18] H. Adelsberger, C. Hertel, E. Glawisching, V.V. Zverlov, W.H. Schwarz, Enzyme system of *Clostridium stercorarium* for hydrolysis of arabinoxylan: reconstitution of the *in vivo* system from recombinant enzymes, *Microbiology* 150 (2004) 2257–2266.
- [19] H. Li, K. Morimoto, N. Katagiri, T. Kimura, K. Sakka, S. Lun, K. Ohmi, A novel β -N-acetylglucosaminidase of *Clostridium paraputrificum* M-21 with high activity on chitobiose, *Appl. Microbiol. Biotechnol.* 60 (2002) 420–427.
- [20] D. Dodd, S.A. Kocherginskaya, M.A. Spies, K.E. Beery, C.A. Abbas, R.I. Mackie, I.K.O. Cann, Biochemical analysis of a β -D-xylosidase and a bifunctional xylanase-ferulic acid esterase from a xylanolytic gene cluster in *Prevotella ruminicola* 23, *J. Bacteriol.* 191 (2009) 3328–3338.
- [21] A.K. Goyal, D.E. Eveleigh, Cloning, sequencing and analysis of the ggh-A gene encoding a 1,4- β -D-glucan glucosylase from *Microbispora bisporea*, *Gene* 172 (1996) 93–98.
- [22] R. Breves, K. Bronnenmeier, N. Wild, F. Lottspeich, W.L. Staudenbauer, J. Hofemeister, Gene encoding two different β -glucosidases of *Thermoanaerobacter brockii* are clustered in a common operon, *Appl. Environ. Microbiol.* 63 (1997) 3902–3910.
- [23] P. Turner, D. Svensson, P. Adlercreutz, E.N. Karlsson, A novel variant of *Thermotoga neapolitana* β -glucosidase B is an efficient catalyst for the synthesis of alkyl glucosides by transglycosylation, *J. Biotechnol.* 130 (2007) 67–74.
- [24] M. Kurakake, T. Fujii, M. Yata, T. Okazaki, T. Komaki, Characteristics of transxylosylation by β -xylosidase from *Aspergillus awamori* K4, *Biochim. Biophys. Acta* 1726 (2005) 272–279.
- [25] D.C. Grange, I.S. Pretorius, M. Claeysens, W.H. van Zyl, Degradation of xylan to D-xylose by recombinant *Saccharomyces cerevisiae* coexpressing the *Aspergillus niger* β -xylosidase (*xlnD*) and the *Trichoderma reesei* xylanase II (*xyn2*) genes, *Appl. Environ. Microbiol.* 67 (2001) 5512–5519.
- [26] M. Wakiyama, K. Yoshihara, S. Hayashi, K. Ohta, Purification and properties of an extracellular β -xylosidase from *Aspergillus japonicus* and sequence analysis of the encoding gene, *J. Biosci. Bioeng.* 106 (2008) 398–404.
- [27] P. Harnpicharnchai, V. Champreda, W. Sornlake, L. Eurwilaichitr, A thermotolerant β -glucosidase isolated from an endophytic fungi, *Periconia* sp., with a possible use for biomass conversion to sugars, *Protein Expr. Purif.* 67 (2009) 61–69.
- [28] B. Li, V. Renganathan, Gene cloning and characterization of a novel cellulose-binding β -glucosidase from *Phanerochaete chrysosporium*, *Appl. Environ. Microbiol.* 64 (1998) 2748–2754.
- [29] J. Hong, H. Tamaki, H. Kumagai, Cloning and functional expression of thermotable β -glucosidase gene from *Thermoascus aurantiacus*, *Appl. Microbiol. Biotechnol.* 73 (2007) 1331–1339.
- [30] A. Wallecha, S. Mishra, Purification and characterization of two β -glucosidases from a thermo-tolerant yeast *Pichia etchellsii*, *Arch. Biochem. Biophys.* 1649 (2003) 74–84.
- [31] M. Machida, I. Ohtsuki, S. Fukui, I. Yamashita, Nucleotide sequences of *Saccharomycopsis fibuligera* genes for extracellular β -glucosidases as expressed in *Saccharomyces cerevisiae*, *Appl. Environ. Microbiol.* 54 (1998) 3147–3155.
- [32] D.K. Watt, H. Ono, K. Hayashi, *Agrobacterium tumefaciens* β -glucosidase is also an effective β -xylosidase, and has a high transglycosylation activity in the presence of alcohols, *Biochim. Biophys. Acta* 1385 (1998) 78–88.
- [33] V.V. Zverlov, I.Y. Volkov, T.V. Velikodvorskaya, W.H. Schwarz, *Thermotoga neapolitana* *bglB* gene, upstream of *lamA*, encodes a highly thermostable β -glucosidase that is a laminaribiose, *Microbiology* 143 (1997) 3537–3542.
- [34] R.C. Lee, M. Hrmova, R.A. Burton, J. Lahnstein, G.B. Fincher, Bifunctional family 3 glycoside hydrolases from barley with α -L-arabinofuranosidase and β -D-xylosidase activity, *J. Biol. Chem.* 278 (2003) 5377–5387.
- [35] V. Mai, J. Wiegler, W.W. Lorenz, Cloning, sequencing, and characterization of the bifunctional xylosidase-arabinosidase from the anaerobic thermophile *Thermoanaerobacter ethanolicus*, *Gene* 247 (2000) 137–143.
- [36] J.S. Xiong, M.B. Vanney, Z.P. Xie, M. Schultze, A. Kondorosi, E. Kondorosi, C. Staehelin, Molecular cloning of a bifunctional β -xylosidase/ α -L-arabinosidase from alfalfa roots: heterologous expression in *Medicago truncatula* and substrate specificity of the purified enzyme, *J. Exp. Bot.* 58 (2007) 2799–2810.
- [37] M. Hrmova, A.J. Harvey, J. Wang, N.J. Shirley, G.P. Jones, B.A. Stone, P.B. Høj, G.B. Fincher, Barley β -D-glucan exohydrolases with β -D-glucosidase activity: purification, characterization, and determination of primary structure from a cDNA clone, *J. Biol. Chem.* 271 (1996) 5277–5286.
- [38] M. Hrmova, G.B. Fincher, Barley β -D-glucan exohydrolases. Substrate specificity and kinetic properties, *Carbohydr. Res.* 305 (1998) 209–221.
- [39] J.N. Varghese, M. Hrmova, G.B. Fincher, Three-dimensional structure of a barley β -D-glucan exohydrolase, a family 3 glycosyl hydrolase, *Structure* 7 (1999) 179–190.
- [40] M. Hrmova, J.N. Varghese, R.D. Gori, B.J. Smith, H. Driguez, G.B. Fincher, Catalytic mechanisms and reaction intermediates along the hydrolytic pathway of a plant β -D-glucan glucosylase, *Structure* 9 (2001) 1005–1016.
- [41] M. Hrmova, R.D. Gori, B.J. Smith, A. Vasela, J.N. Varghese, G.B. Fincher, Three-dimensional structure of the barley β -D-glucan glucosylase in complex with a transition state mimic, *J. Biol. Chem.* 279 (2004) 4970–4980.
- [42] M. Hrmova, R.D. Gori, B.J. Smith, J.K. Fairweather, H. Driguez, J.N. Varghese, G.B. Fincher, Structural basis for broad substrate specificity in higher plant β -D-glucan glucosylases, *Plant Cell* 14 (2002) 1033–1052.
- [43] M. Hrmova, V.A. Streltsov, B.J. Smith, A. Vasella, J.N. Varghese, G.B. Fincher, Structural rationale for low-nanomolar binding of transition state mimics to a family GH3 β -D-glucan glucosylase from barley, *Biochemistry* 44 (2005) 16529–16539.
- [44] M. Hrmova, G.B. Fincher, Dissecting the catalytic mechanism of a plant β -D-glucan glucosylase through structural biology using inhibitors and substrate analogues, *Carbohydr. Res.* 342 (2007) 1613–1623.
- [45] P.J.M. Steenbakkers, H.R. Harhangi, M.W. Bosscher, M.M.C. van der Hooft, J.T. Keltjens, C. van der Drift, G.D. Vogels, H.J.M. Op den Camp, β -Glucosidase in cellulosome of the anaerobic fungus *Pirromyces* sp. strain E2 is a family 3 glycoside hydrolase, *Biochem. J.* 370 (2003) 963–970.
- [46] Y.K. Li, J. Chir, F.Y. Chen, Catalytic mechanism of a family 3 β -glucosidase and mutagenesis study on residue Asp-247, *Biochem. J.* 355 (2001) 835–840.
- [47] Y.K. Li, J. Chir, S. Tanaka, F.Y. Chen, Identification of the general acid/base catalyst of a family 3 β -glucosidase from *Flavobacterium meningosepticum*, *Biochemistry* 41 (2002) 2751–2759.
- [48] N. Juge, J.S. Andersen, D. Tull, P. Roepstorff, B. Svensson, Overexpression, purification, and characterization of recombinant barley α -amylases 1 and 2 secreted by the methylotrophic yeast *Pichia pastoris*, *Protein Expr. Purif.* 8 (1996) 204–214.
- [49] L. Ferrares, L. Trainotti, S. Gattolin, G. Casadoro, Secretion, purification and activity of two recombinant pepper endo- β -1,4-glucanases expressed in the yeast *Pichia pastoris*, *FEBS Lett.* 23 (1998) 23–26.
- [50] J.L. Cereghino, J.M. Cregg, Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*, *FEMS Microbiol. Rev.* 24 (2000) 45–66.
- [51] J.M. Cregg, I. Tolstorukov, A. Kusari, J. Sunga, K. Madden, T. Chappell, Expression in the yeast *Pichia pastoris*, *Methods Enzymol.* 463 (2009) 169–189.
- [52] J.R. Ketudat-Cairns, V. Champattanachai, C. Srisomsap, B. Wittman-Liebold, B. Thiede, J. Svasti, Sequence and expression of Thai Rosewood β -glucosidase/ β -fucosidase, a family 1 glycosyl hydrolase glycoprotein, *J. Biochem.* 128 (2000) 999–1008.
- [53] P. Toonkool, P. Metheenukul, P. Sujwattananat, P. Paiboon, N. Tongtubtim, J.R. Ketudat-Cairns, J. Svasti, Expression and purification of dalcocinase, a β -glucosidase from *Dalbergia cochinchinensis* Pierre, in yeast and bacterial hosts, *Protein Expr. Purif.* 48 (2006) 195–204.
- [54] R. Opasiri, B. Pomthong, T. Onksoong, T. Akiyama, A. Esen, J.R. Ketudat-Cairns, Analysis of rice glycosyl hydrolase family 1 and expression of Os4bglu12 β -glucosidase, *BMC Plant Biol.* 6 (2006) 33.
- [55] A.J. Harvey, M. Hrmova, G.B. Fincher, Regulation of genes encoding β -D-glucan glucosylases in barley (*Hordeum vulgare* L.), *Physiol. Plant* 113 (2001) 108–120.
- [56] J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*, third ed., Cold Spring Harbor Press, Cold Spring Harbor, New York, 2001.
- [57] R. Opasiri, J.R. Ketudat-Cairns, T. Akiyama, O. Wara-Asawapati, J. Svasti, A. Esen, Characterization of a rice β -glucosidase highly expressed in flower and germinating shoot, *Plant Sci.* 165 (2003) 627–638.
- [58] F. Perella, EZ-FIT: a practical curve-fitting microcomputer program for the analysis of the enzyme kinetic data on IBM-PC compatible computers, *Anal. Biochem.* 174 (1988) 437–447.
- [59] M. Chantarangsee, W. Tanthanuch, T. Fujimura, S.C. Fry, J.R. Ketudat-Cairns, Molecular characterization of β -galactosidases from germinating rice (*Oryza sativa*), *Plant Sci.* 173 (2007) 118–134.
- [60] T. Kuntothom, S. Luang, A.J. Harvey, G.B. Fincher, R. Opasiri, M. Hrmova, J.R. Ketudat-Cairns, Rice family GH1 glycoside hydrolases with β -D-glucosidase and β -D-mannosidase activities, *Arch. Biochem. Biophys.* 491 (2009) 85–95.
- [61] S. Seshadri, T. Akiyama, R. Opasiri, B. Kuapaser, J.R. Ketudat-Cairns, Structural and enzymatic characterization of Os3Bglu6, a rice β -glucosidase hydrolyzing hydrophobic glycosides and (1 → 3)- and (1 → 2)-linked disaccharides, *Plant Physiol.* 151 (2009) 47–58.
- [62] Z. Xu, M.C. Shih, J.E. Poulton, An extracellular exo- β -(1,3)-glucanase from *Pichia pastoris*: purification, characterization, molecular cloning, and functional expression, *Protein Expr. Purif.* 47 (2006) 118–127.
- [63] M. Hrmova, V. Farkas, A.J. Harvey, J. Lahnstein, B. Wischmann, N. Kaewthai, I. Ezcurra, T.T. Teeri, G.B. Fincher, Substrate specificity and catalytic mechanism of a xyloglucan xyloglucosyl transferase HvXET6 from barley (*Hordeum vulgare* L.), *FEBS J.* 276 (2009) 437–456.

- [64] M. Hrmova, E.A. MacGregor, P. Biely, R.J. Stewart, G.B. Fincher, Substrate binding and catalytic mechanism of a barley β -D-glucosidase/(1,4)- β -D-glucan exohydrolase, *J. Biol. Chem.* 273 (1998) 11134–11143.
- [65] P. Ruile, C. Winterhalter, W. Liebl, Isolation and analysis of a gene encoding α -glucuronidase, an enzyme with a novel primary structure involved in the breakdown of xylan, *Mol. Microbiol.* 23 (1997) 267–279.
- [66] R.A. Reeves, M.D. Gibbs, D.D. Morris, K.R. Griffiths, D.J. Saul, P.L. Bergquist, Sequencing and expression of additional xylanase genes from the hyperthermophile *Thermotoga maritima* FjSS3B.1, *Appl. Environ. Microbiol.* 66 (2000) 1532–1537.
- [67] N. Oganessian, S.H. Kim, R. Kim, On-column protein refolding for crystallization, *J. Struct. Funct. Genomics* 6 (2005) 177–182.
- [68] B. Batas, C. Schiraldi, J.B. Chaudhuri, Inclusion body purification and protein refolding using microfiltration and size exclusion chromatography, *J. Biotechnol.* 68 (1999) 149–158.
- [69] A.K. Patra, R. Mukhopadhyay, R. Mukhija, A. Krishnan, Optimization of inclusion body solubilization and renaturation of recombinant human growth hormone from *Escherichia coli*, *Protein Expr. Purif.* 18 (2000) 182–192.
- [70] R.G. Brankamp, K. Sreekrishna, P.L. Smith, D.T. Blankenship, A.D. Cardin, Expression of a synthetic gene encoding the anticoagulant–antimetastatic protein ghilanten by the methylotropic yeast *Pichia pastoris*, *Protein Expr. Purif.* 6 (1995) 813–820.
- [71] F. Maley, R.B. Trimble, A.L. Tarentino, T.H. Plummer, Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases, *Anal. Biochem.* 180 (1989) 195–204.
- [72] T. Koseki, Y. Miwa, Y. Mese, A. Miyanaga, S. Fushinobu, T. Wakagi, H. Shoun, H. Matsuzawa, K. Hashizume, Mutational analysis of *N*-glycosylation recognition sites on the biochemical properties of *Aspergillus kawachii* α -L-arabinofuranosidase 54, *Biochim. Biophys. Acta* 1760 (2006) 1458–1464.

ISSN 1420-682X, Volume 67, Number 20



**This article was published in the above mentioned Springer issue.
The material, including all portions thereof, is protected by copyright;
all rights are held exclusively by Springer Science + Business Media.**

**The material is for personal use only;
commercial use is not permitted.**

**Unauthorized reproduction, transfer and/or use
may be a violation of criminal as well as civil law.**

REVIEW

β -Glucosidases

James R. Ketudat Cairns · Asim Esen

Received: 3 February 2010 / Revised: 13 April 2010 / Accepted: 30 April 2010 / Published online: 20 May 2010
© Springer Basel AG 2010

Abstract β -Glucosidases (3.2.1.21) are found in all domains of living organisms, where they play essential roles in the removal of nonreducing terminal glucosyl residues from saccharides and glycosides. β -Glucosidases function in glycolipid and exogenous glycoside metabolism in animals, defense, cell wall lignification, cell wall β -glucan turnover, phytohormone activation, and release of aromatic compounds in plants, and biomass conversion in microorganisms. These functions lead to many agricultural and industrial applications. β -Glucosidases have been classified into glycoside hydrolase (GH) families GH1, GH3, GH5, GH9, and GH30, based on their amino acid sequences, while other β -glucosidases remain to be classified. The GH1, GH5, and GH30 β -glucosidases fall in GH Clan A, which consists of proteins with $(\beta/\alpha)_8$ -barrel structures. In contrast, the active site of GH3 enzymes comprises two domains, while GH9 enzymes have $(\alpha/\alpha)_6$ barrel structures. The mechanism by which GH1 enzymes recognize and hydrolyze substrates with different specificities remains an area of intense study.

Keywords Biological function · Structure · Substrate-specificity · Glycoside hydrolase · Glycosides · Structure–function relationships

Introduction

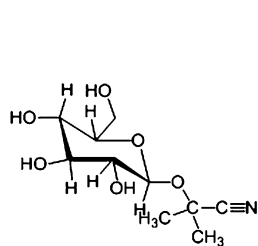
Beta-glucosidases (β -D-glucopyranoside glucohydrolases, E.C. 3.2.1.21) are enzymes that hydrolyze glycosidic bonds to release nonreducing terminal glucosyl residues from glycosides and oligosaccharides. These enzymes are found universally, in all domains of living organisms, Archaea, Eubacteria, and Eukaryotes, in which they play a variety of functions. These include biomass conversion in microorganisms, breakdown of glycolipids and exogenous glucosides in animals, and lignification, catabolism of cell wall oligosaccharides, defense, phytohormone conjugate activation, and scent release in plants, as well as both sides of plant–microbe and plant–insect interactions.

Although the definition of β -glucosidases is straightforward, the abundance of nonreducing terminal β -linked D-glucosyl residues in nature, some examples of which are shown in Fig. 1, has led to the assignment of many E.C. numbers for enzymes that hydrolyze their glycosidic bond. Among these enzymes are glucosyl ceramidases or glucocerebrosidases (3.2.1.45), glucan 1,4- β -glucosidases (3.2.1.58), glucan 1,3- β -glucosidases (3.2.1.74), steryl- β -glucosidase (3.2.1.104), strictosidine β -glucosidase (3.2.1.105), amygdalin hydrolase (3.2.1.117), prunasin hydrolase (3.2.1.118), vicianin β -glucosidase (3.2.1.119), raucassicine β -glucosidase (3.2.1.125), and coniferin β -glucosidase (3.2.1.126). In addition, β -glucosidases often exhibit additional activities, such as β -D-fucosidase (3.2.1.38), β -D-galactosidase (3.2.1.23), β -D-mannosidase (3.2.1.25), and β -disaccharidase activities, such as β -apio-syl- β -D-glucosidase (3.2.1.161).

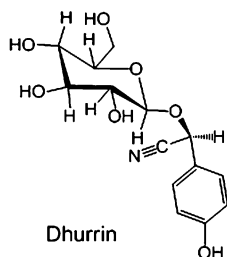
Since the name and the corresponding E.C. number(s) tell us little about mechanism of action, structure, and relationship to other glycoside hydrolases, and one enzyme may catalyze hydrolysis of several related substrates,

J. R. Ketudat Cairns (✉)
Schools of Biochemistry and Chemistry, Institute of Science,
Suranaree University of Technology, 111 University Avenue,
Muang District, Nakhon Ratchasima 30000, Thailand
e-mail: cairns@sut.ac.th

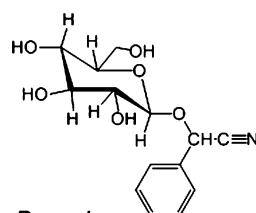
A. Esen
Department of Biological Sciences, Virginia Polytechnic
Institute and State University, Blacksburg, VA, USA



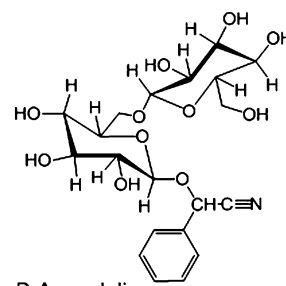
Linamarin



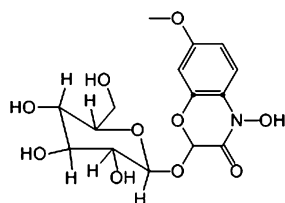
Dhurrin



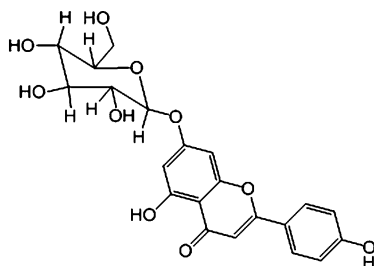
Prunasin



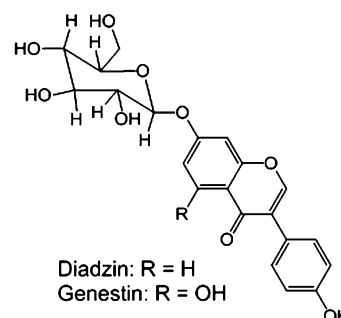
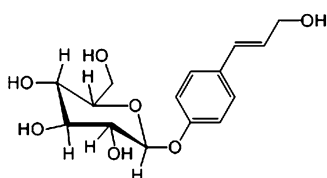
D-Amygdalin



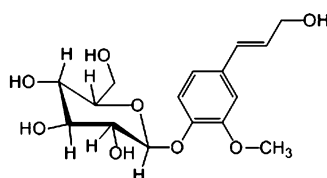
DIMBOA β-glucoside



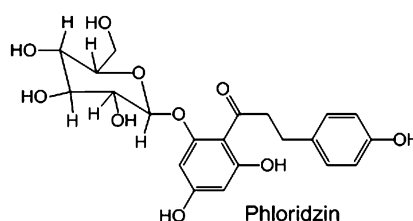
Apigenin 7-O-β-glucoside

Diadzin: R = H
Genestin: R = OH

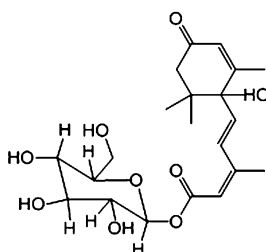
Coumaryl alcohol 4-β-glucoside



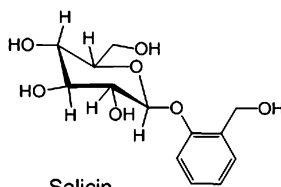
Coniferin



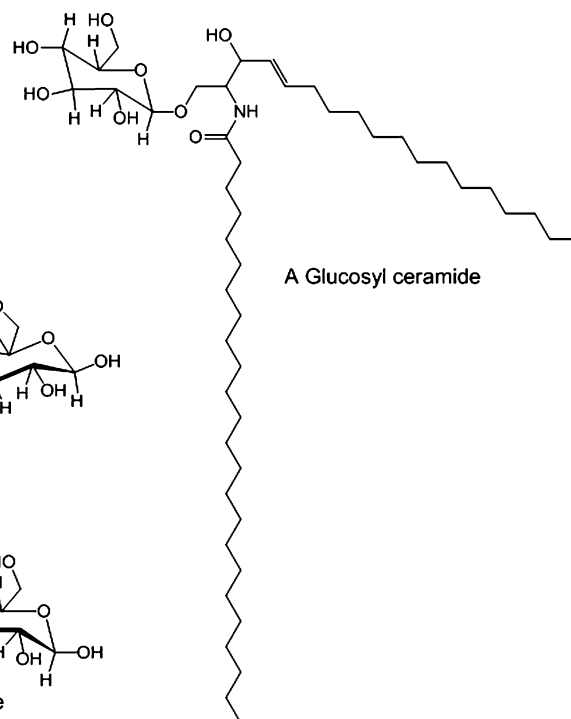
Phloridzin



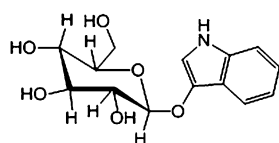
Absciscic acid glucosyl ester



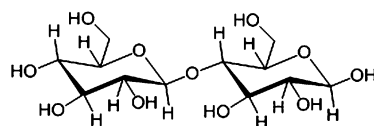
Salicin



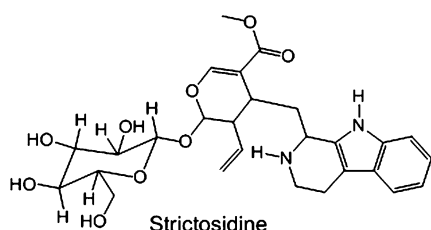
A Glucosyl ceramide



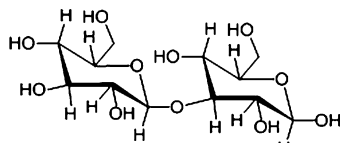
Indoxyl β-glucoside



Cellobiose



Strictosidine



Laminaribiose

◀ **Fig. 1** Structures of example β -glucosidase substrates. The plant cyanogenic glucosides linamarin, dhurrin, prunasin, and its precursor amygdalin. Other defense-related glycosides include 2-*O*- β -D-glucopyranosyl-4-hydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOAGlc) and the flavonoids apigenin 7-*O*- β -D-glucoside, the isoflavonoids diadzin and genistin, and phloridzin. Coniferin and coumaryl alcohol represent monolignol β -glucosides, while abscissic acid glucosyl ester is a phytohormone glucoconjugate and salicin and indoxyl β -glucoside are other plant glycosides with similarity to phytohormones. Strictosidine is the metabolic precursor to a wide array of monoterpene alkaloids. Cellobiose and laminaribiose represent plant cell-wall-derived oligosaccharides and can be extended with the same linkage to give the corresponding triose, tetraose, etc. In the lower right is an example of a glucosyl ceramide, one of the substrates for human acid β -glucosidase (GBA1) and other mammalian β -glucosidases

Henrissat developed an alternative classification system for glycoside hydrolases based on amino acid sequence and structural similarity [1–3]. In this system, those enzymes with overall amino acid sequence similarity and well-conserved sequence motifs are grouped into the same family. At this writing, 115 glycoside hydrolase families are listed in the frequently updated Carbohydrate Active enZYme (CAZY) Web site (<http://www.cazy.org>) [3]. The β -glucosidases that have been described in the literature fall in glycoside hydrolase families GH1, GH3, GH5, GH9, and GH30, [1, 3–5]. In addition, the human bile acid β -glucosidase/GBA2 glucocerebrosidase and its relatives are yet to be assigned to a family. The families that have similar catalytic domain structures and conserved catalytic amino acids, suggestive of a common ancestry and catalytic mechanism, are grouped into clans [2, 3]. Of these, clan GH-A has the largest number of families, and it includes the β -glucosidase-containing families GH1, GH5, and GH30.

GH1 includes the largest number of characterized β -glucosidases; therefore, it will be the emphasis of this review, with the other β -glucosidase families receiving brief mention. We will consider the roles of β -glucosidases and related enzymes in animals, in plants, and in microorganisms, followed by the application of these enzymes. Then, we will describe the structure, mechanism, and the resulting general properties of β -glucosidases that have been purified, and end with a perspective on what is known and needs for further study.

β -Glucosidases and their functional roles

Roles of β -glucosidases and their relatives in mammals

Mammals contain several β -glucosidases, including the family GH1 lactase-phloridzin hydrolase and cytoplasmic β -glucosidase, the GH30 human acid β -glucosidase (GBA1) and the bile acid β -glucosidase or GBA2. These

enzymes are thought to play roles in metabolism of glycolipids and dietary glucosides. In addition, a group of related family GH1 proteins is thought to play signaling functions.

Perhaps the best-studied mammalian β -glucosidase is the human acid β -glucosidase, which is generally considered a glucosyl ceramidase. Defects in the function of this enzyme and its transport to the lysosome lead to Gaucher disease, in which glycosphingolipids accumulate in the lysosomes of tissue leukocytes leading to their engorgement and buildup in the tissues [6]. Since enzyme replacement therapy is one way of alleviating the symptoms for this disease, the enzyme has been produced in recombinant mammalian and insect cells, and gene-activated human cells, and the structures of the enzymes determined [7–9]. Other means of treatment for Gaucher disease include use of imino sugars like deoxynojirimycin and isofagomine and their hydrophobic derivatives, which may inhibit synthesis of glycosphingolipids, but also bind to mutant GBA1 in the ER and help it to fold properly for transport to the lysosome [6]. However, these inhibitors also inhibit other β -glucosidases, such as GBA2 [10, 11].

A human bile acid β -glucosidase (GBA2) was found associated with liver microsomes [12]. Immunofluorescence showed a perinuclear reticulated localization [10], consistent with its earlier assignment to the endoplasmic reticulum (ER), while over-expressed green fluorescent protein tagged enzyme was localized near the plasma membrane [11]. When the bile acid β -glucosidase gene was knocked out in mice, little effect was seen on bile acid metabolism, but there was an accumulation of glucosphingolipids in the Sertoli cells of the testes, leading to round-headed sperm and decreased fertility [10]. Cells transfected with the gene showed an increased ability to hydrolyze fluorescent glycosphingolipids, confirming the protein's identity as GBA2, the nonlysosomal glycosphingolipidase [10, 11].

Currently, five GH1 proteins are known in humans: lactase-phloridzin hydrolase (LPH), cytoplasmic β -glucosidase, Klotho (α -Klotho, KL) β -Klotho (β -KL), and Klotho-LPH-related protein (KLPH) [13]. LPH, an intestinal hydrolase involved in food digestion, has both β -glucosidase activity toward exogenous glucosides, such as phloridzin, and β -galactosidase activity toward lactose. The precursor protein consists of four GH1 domains, the first two of which are removed during maturation, leaving a type I membrane protein with the LPH3 and LPH4 domains bound to the intestinal epithelial cells by a C-terminal transmembrane segment [14]. Flavonoid glucosides appear to be hydrolyzed at both the LPH3 active site and the lactase active site in the LPH4 domain [15, 16]. Deficiency of this enzyme leads to lactose intolerance, one of the most common enzyme deficiencies in humans.

The broad specificity cytoplasmic β -glucosidase has been studied for 30 years [17], and it has recently been given the additional name Klotho-related protein [18]. The cytoplasmic β -glucosidase is found in high levels in hepatocytes and brush border epithelial cells, and it has been shown to hydrolyze plant-derived flavonoid glucosides with high efficiency [19, 20], as shown in Table 1. Recently, the cytoplasmic β -glucosidase was shown to partially account for residual hydrolysis of glucoceramides and galactoceramides in fibroblasts treated with conduritol β -epoxide (CBE), a potent inhibitor of human GBA1 [18]. Although the recombinant enzyme produced in *Escherichia coli* showed extremely slow hydrolysis of natural glucoceramides (Table 1), its structure was solved and found to include lipids in positions suggestive of a glycosceramide binding site and its k_{cat}/K_m values for fluorescently labeled glycosceramides were comparable to those of flavonoid glycosides. The structure of the human cytoplasmic β -glucosidase expressed in *Pichia pastoris* was solved independently, and the residues likely to be involved in its binding of quercetin 4'-O-glucoside were identified by molecular docking and site-directed mutagenesis [13]. No lipids were observed in the active site of the enzyme produced in *P. pastoris*.

The Klotho subfamily of mammalian GH1 proteins (i.e., KL, β -KL, and KLPH) lack a complete active site with both the catalytic acid/base and nucleophile and thus have no β -glucosidase activity [12]. However, KL has been shown to have weak glucuronidase activity and to modify glycosylation of the transient receptor potential ion channel

TRPV5, suggesting they may act as glycoside hydrolases [21]. KL was identified by its induction of rapid aging-like symptoms in knock-out mice [22], and it plays regulatory roles in calcium and phosphate homeostasis [21, 23]. All Klotho subfamily members have N-terminal secretory signal sequences and C-terminal transmembrane domains, and KL has a secretory form as well [24, 25]. The KLPH has a single GH1 domain, while KL and β -KL have two GH1 domains, all of which lack essential catalytic amino acids [24].

Insect β -glucosidases and myrosinases

Although the *Drosophila melanogaster* genome contains only one GH1 gene, suggesting that insects may not have expanded this gene family at an early stage, other insects have adapted glycosides and glycoside hydrolases from the plants on which they feed for protection and digestive purposes [26]. Digestive β -glycosidases from GH1 have been isolated from insect larvae that feed on plants [27, 28]. Similarly, myrosinases have been isolated from specialist insects that feed on crucifers, such as the cabbage aphid, *Brevicoryne brassicae* [29]. The larval β -glycosidases mentioned above can hydrolyze glucoligosaccharides and plant glycosides, such as cellobiose, gentiobiose, and amygdalin [27], in line with their digestive functions. These insect β -glycosidases and myrosinases have sequences most similar to each other, then to vertebrate LPH, suggesting they diverged from the same animal GH1 gene ancestor, after its divergence from plants.

Table 1 Substrate specificity of human cytoplasmic β -glucosidase: substrate K_m and apparent k_{cat} values

Substrate	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\text{mM}^{-1} \text{s}^{-1}$)	Reference
Artificial aryl glycosides				
pNP- β -D-Glc	1,800	12	6.9	Berrin [20]
pNP- β -D-Fuc	370	11	29	Berrin [20]
pNP- α -L-Ara	570	6.0	10	Berrin [20]
pNP- β -D-Gal	3,100	18	5.6	Berrin [20]
Flavonoids and isoflavonoids				
Quercetin 4'-glucoside	34	ND		Day [19]
Quercetin 4'-glucoside	32	1.1	34	Berrin [20]
Quercetin 7-glucoside	42	0.7	16	Berrin [20]
Apigenin 7-glucoside	22	1.5	71	Berrin [20]
Luteolin 4'-glucoside	10	1.2	117	Berrin [20]
Luteolin 7-glucoside	50	3.0	61	Berrin [20]
Genistin	13	ND		Day [19]
Genistin	35	1.5	44	Berrin [20]
Glycosphingolipids				
C6-NBD-GlcCer	4.6	0.121	26	Hayashi [18]
C6-NBD-Gal-Cer	2.0	0.255	128	Hayashi [18]
C18-Glc-Cer	14	0.0072	0.51	Hayashi [18]
C18-Gal-Cer	9.2	<0.0002	<0.02	Hayashi [18]

Berrin et al. [20], recombinant human from *P. pastoris*; Day et al. [19], human from small intestine and liver; Hayashi et al. [18], human expressed in *E. coli*

These genes have since evolved to meet the unique needs of the herbivorous insects in their battle with plant defenses to exploit the plant nutrients.

Roles of GH1 β -glucosidases in plants

Functional diversity and multiplicity

It is in plants that β -glucosidases have been found to play the widest array of biological functions, which include roles in defense, symbiosis, cell wall catabolism and lignification, signalling, and plant secondary metabolism. Several putative β -glucosidase genes have been shown either to be induced by biotic or abiotic stress or to be necessary for successful response to the stress [30–34]. With the advent of genomics, it became clear that about 40 GH1 β -glucosidases are expressed in a typical plant, many in the same tissues [35, 36]. The roles of these enzymes are presumed to be determined by their substrate-specificities, their tissue and subcellular localization, and the conditions under which they come into contact with their physiological substrates.

To match this enormous functional diversity, plants have the largest number of GH1 family proteins. For example, 48 GH1 genes for putative β -glucosidases and thioglucosidases are found in *Arabidopsis thaliana* [35] and 40 GH1 genes are found in rice genome sequences [36]. A number of these represent pseudogenes, and, in the case of rice, two appear to be endophyte genes, but nonetheless both plants appear to express over 30 putative GH1 β -glucosidases. Sequence-based phylogenetic analysis grouped these proteins into eight clusters that include both rice and *Arabidopsis* representatives and two clusters found only in *Arabidopsis* and other plants of the family Caprales, including a cluster of classical thioglucosidases (myrosinases) and a cluster of ER and peroxisomal β -glucosidases and myrosinases. In addition, several groups of enzymes from other plants do not fall into the *Arabidopsis* and rice phylogenetic clusters, including the well-studied chloroplastic β -glucosidases of maize, sorghum, wheat and other cereals, which are not found in rice. Most of these plant GH1 enzymes are closely related to one another, but the lineage of SFR2 [33] shows higher similarity to enzymes from thermophilic bacteria and Archaea than other plant enzymes, and this lineage is thought to be distinct within GH1 [37]. A few of the *Arabidopsis* and rice enzymes have been shown to be primarily β -D-mannosidases [35, 38], so it is possible that some of the others will have different glycone specificities as well, but most are likely to be β -D-glucosidases. Given that plants also contain GH3 and GH5 β -glycosidases with β -glucosidase activity [5, 39], the precise number of β -glucosidase

isoenzymes in a particular plant species has yet to be determined.

Defense and microbial interaction

Plants have long been known to contain glycosides that release toxic compounds, such as cyanide and hydroxamic acids (4-hydroxy-1,4-benzoxazin-3-ones) [40, 41], and the use of β -glucosidases as “detonators” of these chemical “bombs” has recently been reviewed [42]. In general, the defense glycosides are stored in a different cell or a different cellular compartment from the β -glucosidases that hydrolyze them to release toxic compounds. The defense compounds tend to be stored in the vacuole, while their corresponding β -glucosidases are often found in the apoplast or plastid. Both β -glucosidases and thioglucosidases have been found to play these roles, and specialist insects that feed on these plants have adapted these enzymes to diffuse the glycoside bombs or use them for their own defense [26]. β -Glucosidase-mediated defenses are also required for endophytic fungi to develop symbiotic relationships with plants, evidently by modulating the growth of these microorganisms [43].

Plants have developed a wide range of compounds for defense, examples of which can be seen in Fig. 1. Cyanogenic β -glucosides, including linamarin from clover, cassava and various other plants, dhurrin from sorghum, and prunasin from cherry and other stone fruits, are hydrolyzed to release an α -hydroxynitrile, which then breaks down either enzymatically or spontaneously to release cyanide and an aldehyde [41, 42]. Noncyanogenic defense compounds, such as γ - and β -hydroxynitriles and isoflavones in legumes, other flavonoids, coumarins, hydroxaminic acids, such as 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA) in maize and wheat, and saponins are also stored as β -D-glucosides, which are hydrolyzed by specific β -glucosidases [44–54].

Aside from sequestration of the enzyme in the chloroplast or apoplast, several GH1 hydrolases are found in other compartments. The AtBGLU26 (PEN2) myrosinase is found in the pyroxisome [34], while AtBGLU23 (PYK10) is the most abundant protein in an ER-derived compartment called the ER body, which is only found in crucifers [55, 56]. AtBGLU26 has been shown to be critical to the *Arabidopsis* defense against nonspecialist fungi [34, 57]. AtBGLU23 is a β -glucosidase that has been found to be critical for establishment of symbiosis with the endophytic fungus *Piriformospora indica* by preventing it from overgrowing the roots and triggering a defense response [43, 58]. AtBGLU23 and the closely related isoenzymes AtBGLU21 and AtBGLU22 have recently been shown to be specific for scopolin, the most abundant

coumarin glucoside in *Arabidopsis* roots, thereby explaining the antifungal role of AtBGLU23 [54].

Upon cell disruption, plant defensive β -glucosidases and β -thioglucosidases often bind to cytoplasmic aggregating factors, which are thought to help localize the otherwise soluble β -glucosidases and β -thioglucosidases at the site of injury, ensuring a maximal release of defense compounds [53, 59–63]. The functional significance of the interactions of the various defensive β -glucosidases from different cellular compartments and their multiple aggregating factors is an area of active investigation.

Cell wall metabolism

The cell wall of plants is the largest repository of carbohydrates in nature, much of which are β -linked glucosyl residues, so it is not surprising that β -glucosidases should play important roles in cell wall development. β -Glucosidases, in fact, appear to play roles in both the degradation of oligosaccharides generated in cell wall turnover and release of monolignols from their glycosides to allow lignification to stabilize secondary cell walls.

Several β -glucosidases that hydrolyze cell-wall-derived oligosaccharides have been identified over the years and have been studied primarily in monocots. For example, a β -glucosidase in germinating barley seedlings showed activity toward β -1,3- and β -1,4-linked oligosaccharides [64–66]. More recently, it has been shown that this enzyme displays greater preference for mannooligosaccharides, which are also found in barley endosperm cell walls [67]. Rice seedling β -glucosidases have also been shown to hydrolyze oligosaccharides, with varying preferences [36, 38, 68–71]. Rice BGlu1 (Os3BGlu7), Os3BGlu8, and Os7BGlu26 hydrolyzed cellooligosaccharides with increasing efficiency as the degree of polymerization (DP) increased from 2 to 6, while Os4BGlu12 showed little increase in activity with DP and Os3BGlu6 hydrolyzed disaccharides best. The Os3BGlu7 and Os3BGlu8 isoenzymes are widely expressed in rice tissues, so they may be needed for release of glucose from oligosaccharides generated in cell wall remodeling at various stages of plant development. Since the rice β -glucosidase isoenzymes mentioned above also hydrolyze plant-derived glycosides, they may play other roles in the plant as well.

The lignification of secondary cell walls is thought to involve the activation of monolignols by removal of β -glucosyl residues from monolignol glycosides, like cinnamyl alcohol β -glucosides [72]. A coniferin β -glucosidase was identified from lodgepole pine tree xylem [73]. Immunological analysis indicated that this protein was localized to the differentiating region of the xylem, consistent with a role in lignification. More recently, two *Arabidopsis* β -glucosidases (AtBGLU45 and AtBGLU46)

that cluster with lodgepole pine coniferin β -glucosidase in phylogenetic analysis were shown to hydrolyze coniferin, syringin and coumaryl alcohol glucoside [74].

Phytohormone activation

Many phytohormone glucosides are found in plants, and there has been some debate as to whether these are terminal inactivation products or storage forms that can be readily activated by specific β -glucosidases. Partially purified rice β -glucosidases were shown to hydrolyze gibberellin glucosides [75], while maize β -glucosidase (Zm-p60.1, an isoform of ZmGlu1, which hydrolyzes the defense compound DIMBOAGlc) was shown to hydrolyze and activate cytokinin β -glucosides in vitro, as well as in vivo after infusion of the exogenous substrate [76]. Abscissic acid (ABA) glucosyl ester (ABA-GE) was shown to be transported from roots to leaves and be hydrolyzed by extracellular β -glucosidase in the leaves, although free ABA is transported in larger amounts [77]. An enzyme hydrolyzing the auxin glucosyl ester 6-*O*-(4-*O*)-indole-3-ylacetyl- β -D-glucose has also been identified, though the nature of this enzyme remains to be determined [78].

Recently, it has been shown that a specific *Arabidopsis* ER β -glucosidase (AtBGLU18, AtBG1) is activated to hydrolyze ABA-GE in response to drought stress [79]. Mutation of the gene for this enzyme caused early germination and defective stomata closing, which could be rescued by transgenic expression of the gene, but not by a gene encoding an inactive mutant, thereby verifying its role in increasing ABA levels. This is perhaps the clearest demonstration of a physiological role for a β -glucosidase in phytohormone activation and suggests that other phytohormone glucoconjugates also serve as β -glucosidase-activated storage forms.

Secondary metabolism

The monoterpene alkaloid intermediate strictosidine is hydrolyzed by a specific cytoplasmic β -glucosidase to allow metabolism to various monoterpene alkaloids, depending on the plant [80]. This enzyme has been characterized from several plants, and that of *Rauvolfia serpentina* has been structurally characterized [81]. One of the downstream products from strictosidine is raucaffricine, a glucoside that can be deglycosylated by raucaffricine β -glucosidase for further metabolism to ajmaline [82]. Recently, another β -glucosidase hydrolyzing alkaloid glucosides was isolated from *Psychotria ipecacuanha*, further expanding on this theme [83]. As such, β -glucosidases can play metabolic roles to release glucosyl blocking groups from metabolic intermediates and allow

their metabolism to various natural products, many of which are of medicinal importance.

Other functions

Plant β -glucosidases may play a variety of other roles. For instance, they appear to play roles in releasing volatiles from glycoside storage forms. This includes flower scents [84] as well as attractants for parasitic wasps that can attack herbivores feeding on the plant [85]. With the wide variety of glucosides found in plants, it is likely many roles remain to be discovered.

Roles of GH1 β -glucosidases in microorganisms

Although much research has been done on β -glucosidases from microorganisms, most of it has focused on their application rather than their endogenous function. As such, most of the enzymes that have been studied in the context of their natural function are those involved in bioconversion to produce glucose from plant biomass, or in breaking through plant cell walls to establish pathogenic or symbiotic relationships [86]. Bacterial β -glucosidases are often components of large complexes called cellosomes, contain polysaccharide degrading endoglucanases and carbohydrate binding proteins to localize the complex and to the cellulose surface and the cell membrane [87, 88]. Alternatively, some microorganisms secrete soluble endoglycosidases and exoglycosidases for this function, including exoglucanases/ β -glucosidases. Fungi, such as the white rot fungus *Phanerochaete chrysosporium*, may contain cytoplasmic β -glucosidases and extracellular exoglucanases, some of which may act in metabolism of the organism's own cell wall, in addition to plant cell wall biomass metabolism [89–91].

Applications of β -glucosidases

As noted above, β -glucosidases are of interest for biomass conversion, since conversion of β -glucans, the largest source of biomass in the world, is generally accomplished by three enzymes, endo- β -glucanases (e.g., cellulases), exo- β -glucanases (e.g., cellobiosidases) and β -glucosidases [86]. Limiting factors in conversion of cellulose to glucose for fermentation to alcohol include the inhibition of cellulases by oligosaccharides and the lack of adequate β -glucosidase production by certain microorganisms used for biomass breakdown. Thus, the identification and production of β -glucosidases, especially those with high glucose tolerance, has been of interest, and applicable β -glucosidases have been isolated from bacteria and fungi [86, 89–91].

There are hundreds of different β -glucosidic flavor precursors in plants, and their hydrolysis often enhances

the quality of the beverages and foods produced from them [92, 93]. Generally, there are native β -glucosidases in source-plant tissues that hydrolyze these flavor precursors to produce the desired aglycone moiety. Enzymes from the source plants or other sources may be added to foods and beverages before, during, or after processing to enhance food quality. Enzymes with desirable properties may be targeted for breeding programs to increase their abundance in the plants or for overproduction in transgenic microbial or plant hosts, and for engineering to improve their catalytic properties for flavor enhancement and stability.

Aside from flavor enhancement, foods, feeds, and beverages may be improved nutritionally by release of vitamins, antioxidants, and other beneficial compounds from their glycosides. For instance, vitamin B₆ (pyridoxine) can be released from its glucoside by enzymes in rice, in which pyridoxine glucoside is its predominant form [70, 94]. Other vitamins are also found as glucosides in plant sources, and release of their aglycones may improve their nutritional availability, despite the presence of animal and microbial β -glucosidases in the small intestine to aid in this process. Therefore, animal feeds are often treated with crude β -glucosidases. Much work has been done to identify β -glucosidases that can hydrolyze soy isoflavone glycosides, the aglycones of which have antioxidant properties [45–47, 95, 96]. Similarly, the pungent taste of food made from cruciferous vegetables (e.g., broccoli, cabbage, cauliflower, horseradish, kale, mustard, watercress, etc.) is derived from the products of the myrosinase/glucosinolate system, which may also have anticarcinogenic effects, although they may cause endemic goiter in large amounts [93, 97].

The compartmentalized β -glucosidase- β -glucoside defense systems found in such food and feed-plant tissues as cassava roots and leaves, lima beans, flax seeds, and clover leaves produce HCN when the tissue is macerated during preparation or by chewing [93]. The bitterness in almonds is caused by the presence of cyanogenic glucosides [98]. Cassava is highly consumed in parts of Asia, Africa, and South America, and contains the cyanogenic β -glucoside linamarin and its β -glucosidase linamarase. When consumed raw, cyanide poisoning can occur with symptoms of difficulty in breathing, paralysis, convulsion, coma, and even death. Similar symptoms can arise when bitter almonds are eaten raw. Processing of cassava roots by maceration results in the enzyme releasing the HCN, and subsequent aeration and washing removes the products of cyanogenesis. Alternatively, thorough cooking destroys the linamarase enzyme, preventing cyanide release.

In addition to their catalysis of hydrolysis, β -glucosidases also catalyze reverse hydrolysis and transglycosylation reactions, which can be used to synthesize oligosaccharides and glycosides of interest [99]. Various mutations have been

developed to maximize the products of these transglycosylation reactions by manipulating the catalytic mechanism, as described below.

Biochemistry of β -glucosidases

Structures of β -glucosidases

β -Glucosidases have various structures, but the overall fold of the catalytic domain is similar in each GH family. The families GH1, GH5, and GH30 belong to the Clan GH-A, and they all have similar $(\beta/\alpha)_8$ -barrel domains that contain their active site. In contrast, GH3 enzymes have two domains contributing to their active site. GH9 enzymes have $(\alpha/\alpha)_6$ -barrel structures, while the GBA2 family shows weak homology to proteins with this $(\alpha/\alpha)_6$ structure as well (Fig. 2). We will consider these in turn, followed by a more in-depth look at GH1 enzymes, which serve as an excellent model for studying the structural basis for diverse substrate specificities.

The clan GH-A enzymes of families GH1, GH5, and GH30 all have a common $(\beta/\alpha)_8$ -barrel structure and their

active sites contain two conserved carboxylic acid residues on β -strands 4 and 7, serving as the catalytic acid/base and nucleophile, respectively [100, 101]. Although structures are available for all three of these families, our focus will be on GH1 here, since the relatively closely related GH1 plant β -glycosidases show a high diversity of substrate specificities, the basis of which will be considered later. The lengths and subunit masses of these GH1 enzymes vary considerably, depending on the presence of auxiliary domains and redundant GH1 domains (as in human LPH), but the catalytic domain itself ranges from around 440 to 550 residues, depending on the lengths of the variable loops at the C-terminal ends of the β -strands of the $(\beta/\alpha)_8$ -barrel [102]. These monomers form a wide range of quaternary structures, including monomeric enzymes, dimers, tetramers, hexamers, octamers, and large aggregates.

The GH3 β -glucosidases and exoglucanases have a two-domain structure, a $(\beta/\alpha)_8$ -barrel followed by an α/β sandwich comprising a 6-stranded β -sheet sandwiched between three α -helices on either side [103]. The active site of GH3 enzymes is situated between the $(\beta/\alpha)_8$ and $(\alpha/\beta)_6$ sandwich domains, each of which contributes one catalytic carboxylate residue (Fig. 2). The catalytic nucleophile for

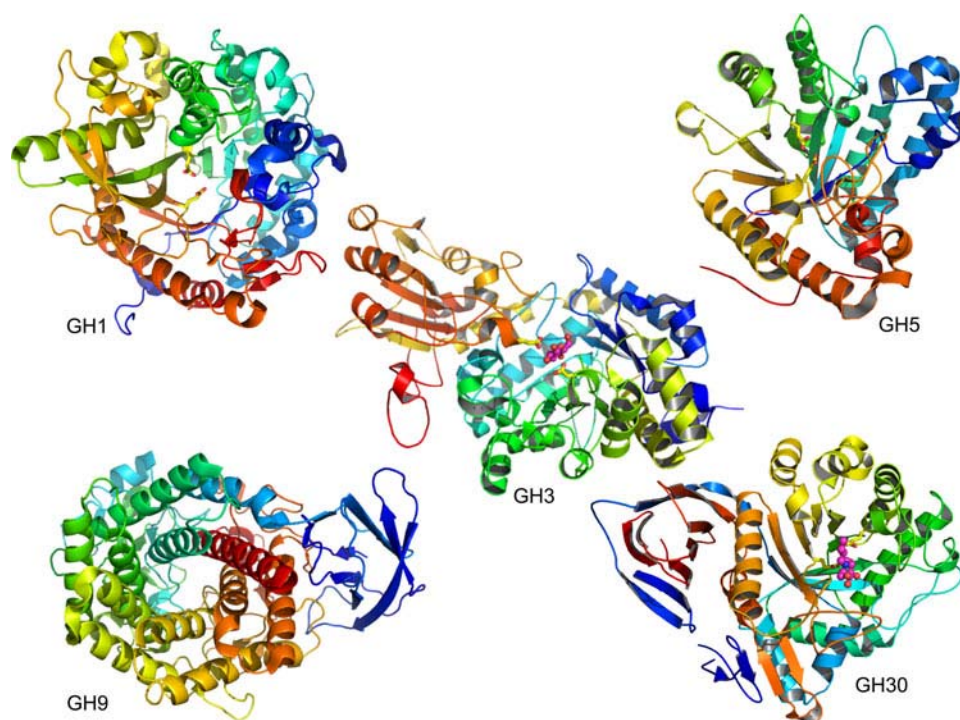


Fig. 2 Structures of β -glucosidases from different GH families. These include β -glucosidases or related enzymes from GH1 (*Zea mays* ZmGlu1, PDB code 1E1E), GH3 (*Hordeum vulgare* Exo I β -glucan glucohydrolase, PDB code 1EX1), GH5 (*Candida albicans* exo- β -(1,3)-glucanase Exg exoglucanase, PDB code 1CZ1), GH30 (*Homo sapiens*, acid β -glucosidase/glucocerebrosidase GBA1, PDB code 2V3D), and GH9 (*Vibrio parahaemolyticus*, putative exoglucanase, PDB code 3H7L). The structural cartoons are colored in a

spectrum from blue to red from their N- to C-termini, with the catalytic nucleophile and acid-base residues shown in stick for those enzymes in which they are known. The ligands shown are glucose in the GH3 barley ExoI and *N*-butyl-deoxynojirimycin in the GH30 human GBA1, both of which are shown with carbons in pink. The human GBA2 (bile acid β -glucosidase) shows low levels of sequence similarity to $(\alpha/\alpha)_6$ enzymes, suggesting its catalytic domain may be similar to the GH9 structure. Drawn with Pymol (DeLano)

barley Exo I is an aspartate at residue D285, which resides on the loop after β -strand 7 of the $(\beta/\alpha)_8$ barrel, while the catalytic acid/base is glutamate E491, which is on a long loop extending from the $(\alpha/\beta)_6$ sandwich domain [104].

Only a few GH9 proteins have been verified to be β -glucosidases [105, 106], as most proteins in this family are endoglycosidases. This family consists of $(\alpha/\alpha)_6$ barrels. Recently, the structure of a *Vibrio parahaemolyticus* protein with 69% amino acid sequence identity over 567 residues with the *Vibrio cholera* β -glucosidase, was determined (PDB accession 3H7L, Fig. 2). These family GH9 enzymes are the first β -glucosidases shown to act through an inverting mechanism (the prevailing mechanism in family GH9), which is unusual, since all other β -glucosidases described so far act through a retaining mechanism [106].

The human GBA2 and its relatives are not related to other β -glycosidases, but show weak similarities to certain $(\alpha/\alpha)_6$ enzymes in homology searches. The GBA2 sequence contains no secretory pathway signal sequence and a single putative transmembrane domain, but was predicted to have its N-terminus in the endoplasmic reticulum and C-terminus in the cytoplasm [11]. The position of this putative transmembrane α -helix falls in the middle of a sequence homologous to soluble $(\alpha/\alpha)_6$ amylohydrolase, chitobiose phosphorylase and α -L-rhamnosidase (3.2.1.40) enzymes, and the low confidence of the transmembrane topology prediction call this topology into question, but the enzyme is clearly associated with membranes by some means [10, 11].

Catalytic mechanisms

Glycoside hydrolases perform catalysis using two mechanisms, one with inversion and one with retention of chirality at the anomeric carbon [105]. Both of these mechanisms use a pair of acidic and nucleophilic residues, usually carboxylic acids, on either side of the sugar, approximately 5 Å apart in the retaining mechanism, and 10 Å apart in the inverting mechanism, in which a water molecule must fit between the catalytic base and the substrate. The GH9 β -glucosidases use an inverting mechanism, in which an activated water molecule makes a direct nucleophilic attack on the anomeric carbon to displace the aglycone in a single step, as shown in Fig. 3a [106]. The catalytic base extracts a proton from the incoming water molecule while the catalytic acid protonates the leaving group aglycone. In contrast, most β -glucosidases that have been characterized (i.e., GH1, GH3, and GH30 enzymes) are retaining enzymes, and they perform catalysis in two steps, glycosylation and deglycosylation (Fig. 3b). In glycosylation, the aglycone departs with the donation of a proton from the catalytic acid/base

and nucleophilic attack of the catalytic nucleophile on the anomeric carbon to yield an α -linked covalent enzyme-glycone intermediate. In the deglycosylation step, the process is reversed, as a water molecule attacks with basic assistance from the catalytic acid/base to displace the catalytic nucleophile from the glucose.

Both the glycosylation and deglycosylation steps are thought to pass through oxocarbenium-ion-like transition states. The glucose of the incoming substrate has sometimes been observed to be distorted into a 1S_3 skew boat as it moves toward the 4H_3 half-chair shape in the first transition state, although in other structures it is poorly defined by the electron density, apparently due to high mobility [108–112]. The structures of certain putative transition state mimics have also been solved in the active site and shown to have a structure close to the 4H_3 half-chair, although others appeared to inhibit by mechanisms other than transition state mimicry [108, 113–118].

The presence of the covalent intermediate was first demonstrated with the GH1 *Agrobacterium* sp. β -glucosidase by covalent labeling with 2,4-dinitrophenyl-2-deoxy-2-fluoroglucoside [119, 120]. In this inhibitor, the electronegative fluoride atom destabilizes the transition state for both half reactions, while the 2,4-dinitrophenylate provides an excellent leaving group to allow the glycosylation step to proceed rapidly. This traps the enzyme in the covalent intermediate and allows the catalytic nucleophile to be identified by tryptic digest and mass spectrometry. This covalent intermediate has also been observed in crystal structures for both the 2-F-glucoside and, in some cases, in the nonfluorinated glucosyl residue in certain acid/base catalyst mutants [104, 112, 121, 122]. The covalent inhibitor CBE has also been used to identify the catalytic nucleophile in some cases [104], but it is less specific and sometimes labels nearby amino acids. The acid/base catalyst of cassava β -glucosidase was also identified with a mechanism-based inhibitor, *N*-bromoacetyl- β -D-glucopyranosylamine [123], but most acid/base residues have been identified through homology, proximity to the glycosidic bond oxygen in crystal structures or site-directed mutagenesis [124].

The double-displacement mechanism for retaining β -glucosidases leads to predictions that mutants of these enzymes in which the acid/base or nucleophile is removed can be rescued by small nucleophiles and utilized for transglycosylation [107, 124–127]. When the acid/base of *Agrobacterium* sp. β -glucosidase was mutated to glycine (E170G), the hydrolysis of 2,4-dinitrophenyl β -D-glucoside (dNPGlc), which has a leaving group that does not require protonation ($pK_a = 3.96$), lost its pH dependence from 7 to 9 and could be rescued by various small nucleophiles, such as azide, which produced β -D-glucosyl azide [124]. This verified E170 as the catalytic acid/base and was consistent

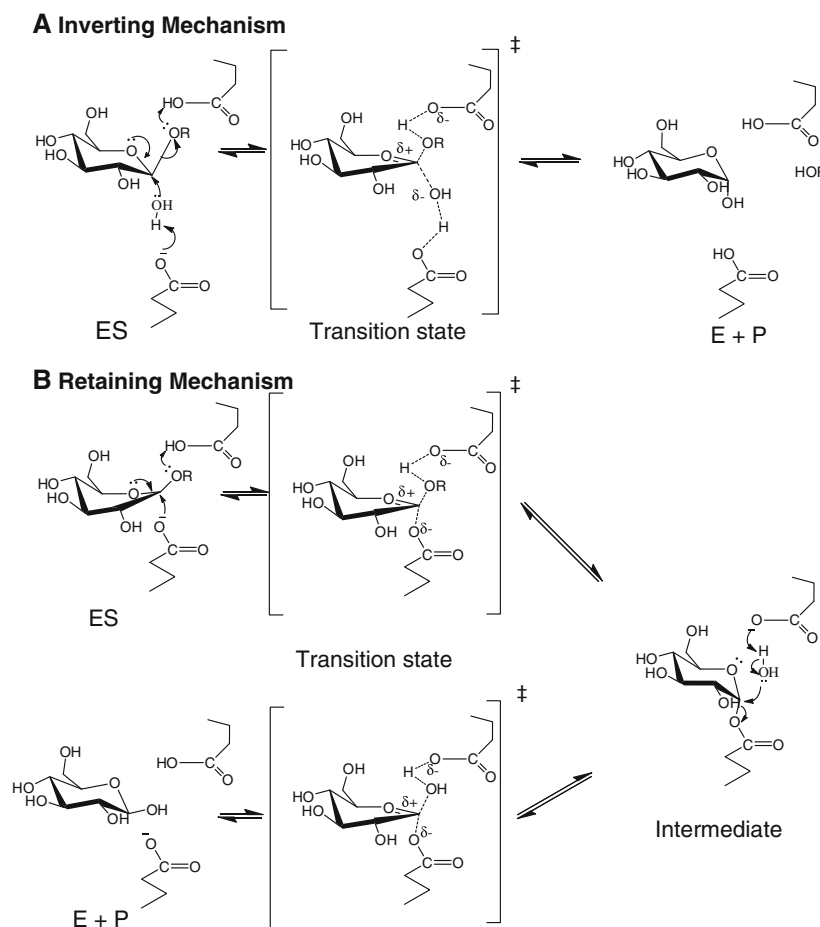


Fig. 3 Retaining catalytic mechanisms of inverting and retaining β -glucosidases. **a** The inverting mechanism that is seen in family GH9 glycoside hydrolases, including β -glucosidases. A single displacement of the aglycone by the water leads to an anomeric carbon with inverted chirality. **b** The commonly accepted mechanism for hydrolysis with retention of anomeric configuration as seen GH Clan A and family GH3 β -glucosidases. The glucosyl moiety is distorted into an 1S_3 skew boat upon binding to the enzyme in preparation to form the 4H_3 half chair conformation of the proposed transition state [107,

108]. The first step is glycosylation, in which the catalytic acid donates a proton to the leaving group, while the catalytic nucleophile attacks from the opposite side to form an α -linked intermediate. In the second, deglycosylation step, the catalytic base (the same carboxylate as the catalytic acid) extracts a proton from a water molecule, improving its nucleophilic power to attack at the anomeric carbon and displace the enzyme. Hydrolysis by either mechanism is equivalent in the organism, since mutarotation of the released glucose will lead to a racemic mixture of glucose in solution after a short time

with the double displacement mechanism in which the role of the catalytic acid has been circumvented. Similarly, conversion of the Abg catalytic nucleophile to a small nonnucleophilic amino acid, i.e., Ala, Ser, or Gly, resulted in an inactive enzyme that could be rescued by azide or fluoride to form α -D-glucosides, thereby converting a retaining enzyme to an inverting enzyme [125, 126]. Alternatively, the use of α -fluoroglucoside, in which the fluoride replaces the enzyme nucleophile in the covalent intermediate, allowed transfer of a β -linked glucosyl moiety onto a sugar or other alcohol. Since these nucleophile mutants have low hydrolytic activity, but relatively high transferase activity, they were designated glycosynthases [127]. Both the acid/base and the nucleophile mutants have potential uses in glycoconjugate synthesis.

Mechanism of substrate binding and specificity

Although the residues responsible for the hydrolytic mechanism are well characterized, how β -glucosidases recognize and interact with their substrates, which in large part determines their diverse functions, is less clear. GH1 enzymes are a prime model for these studies and the structures of 23 GH1 enzymes and their variants are currently available, including three from archaea, nine from bacteria, two from animals, one from a fungus, and eight from plants (CAZY website, [5]). The complexes of several of these enzymes with substrates, inhibitors and covalent intermediates are available, allowing in-depth analysis of residues likely to be critical to substrate and transition state binding. Although many of the prokaryotic enzymes show

rather similar and broad substrate specificities, the complexes of β -glycosidase S from the archaeon *Sulfolobus solfataricus* and β -glucosidase A from the eubacteria *Thermotoga maritima* with a range of inhibitors has provided a wealth of information on catalytic and inhibitory mechanisms [113–118]. In addition, site-directed mutagenesis of GH1 enzymes with and without experimentally determined structures has been done to test the roles of various residues.

The GH1 enzymes may have rather broad glycone specificity, for instance one enzyme may hydrolyze β -D-glucosides, β -D-fucosides, β -D-mannosides, β -D-galactosides and α -L-arabinosides, or may be specific for one or a few glycone sugars. Marana [128] analyzed GH1 specificity and concluded that a conserved glutamate, which bridges the glycone hydroxyl groups 4 and 6 in enzymes with β -glucosidase and β -galactosidase activities but is replaced in 6-phosphoglycosidases, is critical for the distinction between enzymes. However, it still remains to be determined how GH1 enzymes can be primarily β -glucosidases or β -mannosidases or show different ranges of allowed glycones, even though they bind the sugar with the same conserved residues [38, 65, 109, 112]. It is worth noting that binding of the aglycone has also been observed to affect the sugar binding position [109, 129], so residues more distant in the substrate binding pocket cannot be excluded from playing roles in glycone specificity.

The basis of the tremendous diversity in function of β -glucosidases, especially in plants, is the substrate aglycone specificity differences that determine their natural substrates. Structures of complexes of enzymes with inhibitors and mutant enzymes with substrates, along with mutagenesis and chimera studies comparing similar enzymes with divergent specificities, have suggested that the basis of aglycone specificity is complex. Although this includes mutagenesis and structural studies of human cytoplasmic β -glucosidase [13, 130], the plant GH1 enzymes have served as the primary model, due to their high diversity in aglycone specificity.

Maize ZmGlu1 and sorghum dhurrinase 1 (SbDhr1) are closely related, displaying 70% amino acid sequence identity, but have distinct specificities. ZmGlu1 has broad specificity, but cannot hydrolyze dhurrin, the natural substrate of SbDhr1, while SbDhr1 hydrolyzes only dhurrin. Studies of reciprocal ZmGlu1/SbDhr1 chimeric enzymes [131] and subsequent structural and site-directed mutagenesis studies [109–111, 129] indicated that the aglycone specificity determining sites are different in ZmGlu1 and SbDhr1. The structures of a catalytically inactive ZmGlu1 mutant (Glu1E191D) in complex with the natural substrate DIMBOAGlc (Fig. 4a), its free aglycone DIMBOA, and the unhydrolyzed competitive inhibitor dhurrin showed that the aglycone moiety of the substrate is sandwiched

between four aromatic residues, W378 on one side and F198, F205, and F466 on the other [109]. The 7-methoxy group of DIMBOA also has a hydrophobic contact with A467. All of these residues, except W378, are variable among β -glucosidases that differ in substrate specificity, which led to the conclusion that these sites and the active-site shape are the basis of aglycone binding specificity in β -glucosidases. In the case of Dhr1, the three phenylalanines are replaced with V196, L203, and S462, and the active site is smaller (Fig. 4b). A water-mediated H-bond between the dhurrin phenolic hydroxyl and Dhr1 S462 provides a more polar and stronger binding interaction than seen in ZmGlu1 [109–111]. This apparently led to a more stable 1S_3 skew boat conformation of the glucosyl residue, whereas in ZmGlu1 the conformation appeared to be variable, leading to poor electron density around the anomeric carbon. Mutagenesis of these aglycone-binding residues in the Zm60.8 isoform of ZmGlu1 confirmed their importance to hydrolysis of synthetic substrates [132].

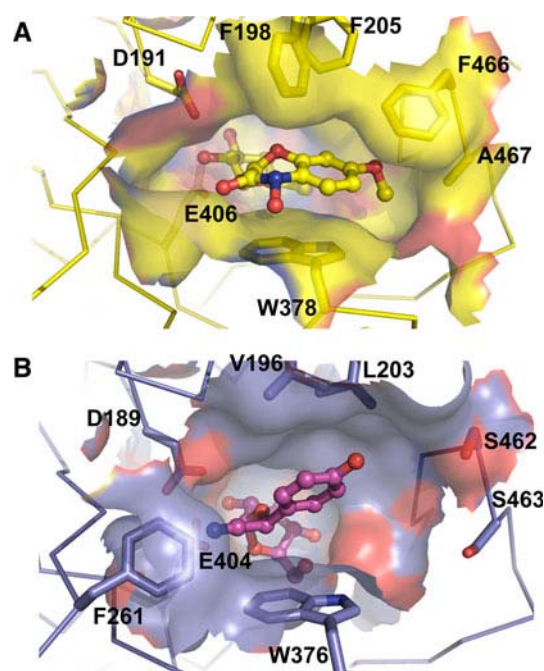


Fig. 4 Active site configurations of maize β -glucosidase 1 (Glu1, **a**) and sorghum dhurrinase 1 (Dhr1, **b**). The active sites of maize Glu1 and sorghum Dhr1 enzymes are shown for the structures of the Glu1 E189D mutant in complex with DIMBOA glucoside (PDB entry 1E56) and Dhr1 E191D mutant in complex with dhurrin (PDB entry 1VO3) [109, 111]. The sidechains of residues noted to interact with the aglycone are shown in stick representation behind the active site surface, which is colored as the underlying residues, which are colored with carbon in yellow for Glu1 and purple for Dhr1, nitrogen in blue, and oxygen in red. The ligands are shown in ball and stick representation with similar coloration. The Phe 261 residue, which narrows the active site in Dhr1, is also shown in front of the catalytic nucleophile Glu 404. Figure created with Pymol

However, the structural investigation of ZmGlu1 to SbDhr1 mutants and the subsequent structure of SbDhr1 and other GH1 hydrolases showed that the above-mentioned variable residues alone cannot designate substrate specificity [111, 131]. Although the Trp corresponding to ZmGlu1 W378 is nearly invariable in other plant GH1 enzymes, its positional variation was found to be critical for binding of substrates like dhurrin and strictosidine [81, 111, 131]. Even the closely related wheat β -glucosidase (TaGlu) was found to have different amino acids at the other aglycone-binding residues found in ZmGlu1, despite the fact that it also hydrolyzes DIMBOAGlc [52]. The oligosaccharide binding site in rice BGlu1 runs out of the active site in a different direction from that of DIMBOAGlc binding in ZmGlu1, so that L442, corresponding to ZmGlu F466, makes no interaction with the substrate, while N190, corresponding to ZmGlu F205, interacts only indirectly ([112], and PDB code 3F5K). Instead, N245 plays a key role in binding to the third glucosyl residue in cellobiosaccharides, while the corresponding residues in SbDhr1 (F261) and rice Os3BGlu6 (M251) appear to block off their active sites, which do not bind to such long substrates [71, 111]. Thus, a different, though overlapping, set of active site residues has been recruited to interact with the aglycone in each GH1 β -glucosidase that has been investigated, in contrast to the highly conserved glycone binding site.

Kinetic parameters for substrate hydrolysis

β -Glucosidases have variable kinetic parameters toward their substrates. The K_m values for natural substrates and other good substrates are typically 1 mM or less, but these values vary roughly 1,000-fold. Similarly, β -glucosidases have relatively low k_{cat} values ($\sim 300\text{ s}^{-1}$ or lower), which may be physiologically beneficial in some roles, but one must suspect the physiological relevance or the enzyme preparation if these values are too low. Comparison of the k_{cat}/K_m (efficiency coefficient) values is generally used to evaluate potential natural substrates, as two substrates with similar K_m values may have vastly different catalytic efficiencies. It is instructive to look at the published values for the hydrolysis of putative natural substrates by natural and recombinant preparations of human cytoplasmic β -glucosidases (Table 1). The plant-derived flavonoid luteolin 4'-glucoside and a synthetic fluorescent glycosceramide, C6-NBD-GalCer, give similar efficiency coefficients, although the former is hydrolyzed approximately five-fold faster than the latter. The natural glycosceramides that have longer acyl chains had much slower hydrolysis rates. Nonetheless, it was shown that RNAi knockdown of the cytoplasmic β -glucosidase resulted in an increase in glycosceramide concentrations, suggesting they may serve as natural substrates in the cell [18].

When β -glucosides with different efficiencies differ in the leaving-group ability of their aglycones, the rate-limiting step will be the glycosylation reaction, while either the glycosylation or deglycosylation step, or both, might be rate-limiting if substrates differ in the glycone. For example, many β -glucosidases hydrolyze *p*-nitrophenyl- β -D-fucoside (*p*NPFuc) with higher efficiency than *p*-nitrophenyl- β -D-glucoside (*p*NPGLc). In the case of maize β -glucosidases [50] and rice Os3BGlu7 [70], the K_m values are similar, and the V_{max} values for *p*NPFuc are clearly higher, however, in *Dalbergia* isoflavonoid β -glucosidases [46, 47, 133] and rice Os3BGlu6 [71], the V_{max} is similar or higher for *p*NPGLc and the K_m is ten-fold lower for *p*NPFuc.

Inhibitors and cofactors

β -Glucosidases are inhibited by transition-state sugar analogues, substrate analogue glycosides, and free aglycones of their substrates, as well as slowly hydrolyzed substrates, such as the fluoroglucosides mentioned earlier. Structural and thermodynamic analysis of 18 putative substrate analogues suggested many may not act as true transition-state analogues, but may nonetheless bind and inhibit *T. maritima* β -glucosidase [118]. Since the aglycone and glycone-binding pockets in the active site are distinct, sugar analogs shaped similar to the half-chair conformation of the transition state can bind to the glycone-binding site and inhibit the enzyme, whereas free aglycones may bind to the aglycone-binding site. Free glucose is a poor inhibitor (typically $K_i = 100\text{--}200\text{ mM}$) because glucose must be distorted toward the half-chair conformation for binding to the glycone-binding site, which is thought to require a portion of the energy of aglycone binding. In contrast, free aglycones can be potent competitive inhibitors because they bind to the aglycone-binding site without energetically unfavorable distortion.

Although most metal ions do not inhibit β -glucosidases, Ag^+ and Hg^{2+} , are potent β -glucosidase inhibitors, and inhibition by Cu^{2+} and Fe^{3+} has also been reported [134]. Although β -O-glucosidases are not known to require any cofactors, ascorbate is known to enhance the activity of β -S-glucosidases (myrosinases) by acting as a surrogate catalytic base [135]. The chelation of Zn^{2+} between the monomers in the biological dimer of myrosinase suggests that metal ions could act in stabilization of some GH1 enzymes.

pH and temperature optima and stability

β -Glucosidases show a range of pH optima and stabilities, depending on their source and amino acid sequence. The pH optima of most β -glucosidases range between pH 4 and 7.5,

depending on their source and cellular location, and they tend to be stable over a range of pH from 4 to 9. It is usually safe to store these enzymes at 0–4°C at pH 7–8, once major protease contaminants have been removed, but this should be tested with each enzyme. As with other proteins, pH extremes, copurifying proteases, and microbial contamination may result in degradation, although many β -glucosidases are resistant to proteases due to their tightly folded core structure. Nonetheless, proteolysis can result in recombinant proteins losing their purification tags and in purified active enzymes appearing to have two subunits on SDS-PAGE due to an internal cleavage that leaves the fold intact, in the authors' experience. Some β -glucosidases are resistant to denaturation by ionic detergents such as SDS, which allows extraction with buffers containing up to 3% SDS and zymogram development after SDS-PAGE when samples are applied without heating.

Thermostability and temperature optima vary greatly among enzymes. Mesophilic β -glucosidases may show highest activity at 30–65°C, but are generally inactivated at and above 55–70°C [46, 53, 58, 73]. High activity at temperatures above the extremes of the enzyme's natural environment is not physiologically relevant and may result in rapid heat denaturation, so assays are often run at 30–40°C. On the other hand, β -glucosidases from thermophilic bacteria, such as *T. maritima* BglA, may have temperature optima of over 100°C [136]. Engineering of a bacterial β -glucosidase to have the same N-terminal and C-terminal residues as *T. maritima* BglA, allowed hydrogen bonding between these termini and stabilized the enzyme, suggesting such interactions may be important for high stability [137]. This thermostability is also thought to be due to an increased number of proline residues, electrostatic bridges, and internal water molecules, and binding of more subunits in the quaternary structure compared to many mesophilic enzymes [138].

Summary and future prospects

The description of β -glucosidases in this review is limited in detail, due to the vast amounts of data that have been generated in the last several years. Nonetheless, we hope the reader will appreciate the wide variety of functions that β -glucosidases play in nature, from biomass degradation by microorganisms, to glycolipid and xenogenic β -glucoside breakdown in animals, to roles in defense, phytohormone regulation, cell wall metabolism, and secondary metabolism in plants, where the β -glucosidases have attained their greatest multiplicity and diversity of function. Although β -glucosidases like sorghum SbDhr1 can be very specific, others show overlapping ranges of activities for multiple substrates, such as the glucocerebrosidase and flavonoid

β -glucosidase activities of human β -glucosidases. It seems likely that β -glucosidases play many as yet undiscovered roles, as well as potential for many applications.

Although the catalytic mechanism is well understood for GH Clan A and GH3 β -glucosidases, the means by which their exact substrate specificity is established has proven to be divergent for even closely related β -glucosidases. This and the high multiplicity of putative β -glucosidase in plants limit the conclusions that can be drawn from genomic sequences as to the putative specificities and functions of new β -glucosidase homologs. Nonetheless, the insights gained from structural and mutagenic studies provide a starting point from which to investigate the functions of new β -glucosidases. As more substrate specificities and structures are determined, it should become more feasible to predict substrate specificity from the sequences of as yet uninvestigated β -glucosidases.

Acknowledgments Rodjana Opassiri and three anonymous reviewers are thanked for useful comments on the manuscript. JRKC was supported by Suranaree University of Technology and the Thailand Research Fund.

References

1. Henrissat B (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* 280:309–316
2. Henrissat B, Davies G (1997) Structural and sequence-based classification of glycosyl hydrolases. *Curr Opin Struct Biol* 7:637–644
3. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZY): an expert resource for glycogenomics. *Nucleic Acids Res* 37:D233–D238
4. Brunner R, Wirtz W, Rose JKC, Darbill AG, Govers F, Scheel D, Nurnberger T (2002) A β -glucosidase/xylosidase from the phytopathogenic oomycete, *Phytophthora infestans*. *Phytochemistry* 59:689–696
5. Opassiri R, Pomthong B, Akiyama T, Nakphaichit M, Onkoksoong T, Ketudat-Cairns M, Ketudat Cairns JR (2007) A stress-induced rice β -glucosidase represents a new subfamily of glycosyl hydrolase family 5 containing a fascin-like domain. *Biochem J* 408:241–249
6. Butters TD (2007) Gaucher disease. *Curr Opin Chem Biol* 11:412–418
7. Dvir H, Harel M, McCarthy AA, Toker L, Silman I, Futerman AH, Sussman JL (2003) X-ray structure of human acid- β -glucosidase, the defective enzyme in Gaucher disease. *EMBO Rep* 4:704–709
8. Liou B, Kazimierczuk A, Zhang M, Scott CR, Hedge RS, Grabowski GA (2006) Analyses of variant acid β -glucosidases: effects of Gaucher disease mutations. *J Biol Chem* 281:4242–4253
9. Lieberman RL, Wustman BA, Huertas P, Powe AC, Pine CW, Khanna R, Schlossmacher MG, Ringe D, Petsko GA (2007) Structure of acid β -glucosidase with pharmacological chaperone provides insight into Gaucher disease. *Nat Chem Biol* 3:101–107
10. Yildiz Y, Matern H, Thompson B, Allegood JC, Warren RL, Ramirez DMO, Hammer RE, Hamra FK, Matern S, Russell DW

- (2006) Mutation of β -glucosidase 2 causes glycolipid storage disease and impaired male fertility. *J Clin Invest* 116:2985–2994
11. Boot RG, Verhoeck M, Donker-Koopman W, Strijland A, van Marle J, Overkleef HS, Wennekes T, Aerts JM (2007) Identification of the non-lysosomal glucosylceramidase as beta-glucosidase 2. *J Biol Chem* 282:1305–1312
 12. Matern H, Boermans H, Lottspeich F, Matern S (2001) Molecular cloning and expression of human bile acid beta-glucosidase. *J Biol Chem* 276:37929–37933
 13. Tribolo S, Berrin J-G, Kroon PA, Czjzek M, Juge N (2007) The structure of human cytoplasmic β -glucosidase unravels substrate aglycone specificity of a family 1 glycoside hydrolase. *J Mol Biol* 370:964–975
 14. Mantei N, Villa M, Enzler T, Wacker H, Boll W, James P, Hunziker W, Semenza G (1988) Complete primary structure of human and rabbit lactase-phlorizin hydrolase: implications for biosynthesis, membrane anchoring and evolution of the enzyme. *EMBO J* 7:2705–2713
 15. Arribas JC, Herrero AG, Martín-Lomas M, Cañada FJ, He S, Withers SG (2000) Differential mechanism-based labeling and unequivocal activity assignment of the two active sites of intestinal lactase/phlorizin hydrolase. *Eur J Biochem* 267:6996–7005
 16. Day AJ, Canada FJ, Diaz JC, Kroon PA, Mclauchlan R, Faulds CB, Plumb GW, Morgan MR, Williamson G (2000) Dietary flavonoid and isoflavone glycosides are hydrolysed by the lactase site of lactase phlorizin hydrolase. *FEBS Lett* 468:166–170
 17. Daniels LB, Coyle PJ, Chiao YB, Glew RH, Labow RS (1981) Purification and characterization of a cytosolic broad specificity beta-glucosidase from human liver. *J Biol Chem* 256:13004–13013
 18. 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
 19. Day AJ, DuPont MS, Ridley S, Rhodes M, Rhodes MJ, Morgan MR, Williamson G (1998) Deglycosylation of flavonoid and isoflavonoid glycosides by human small intestine and liver beta-glucosidase activity. *FEBS Lett* 436:71–75
 20. Berrin J-G, McLauchlan WR, Needs P, Williamson G, Puigserver A, Kroon PA, Juge N (2002) Functional expression of human liver cytosolic β -glucosidase in *Pichia pastoris*. Insights into its role in the metabolism of dietary glucosides. *Eur J Biochem* 269:249–258
 21. Chang Q, Hoefs S, van der Kemp AW, Topala CN, Bindels RJ, Hoenderop JG (2005) The beta-glucuronidase klotho hydrolyzes and activates the TRPV5 channel. *Science* 310:490–493
 22. Kuro-o M, Matsumura Y, Aizawa H, Kawaguchi H, Suga T, Utsuki T, Ohyama Y, Kurabayashi M, Kaname T, Kume E, Iwasaki H, Iida A, Shiraki-Iida T, Nishikawa S, Nagai R, Nabeshima YI (1997) Mutation of the mouse klotho gene leads to a syndrome resembling ageing. *Nature* 390:45–51
 23. Nabeshima Y, Imura H (2008) α -Klotho: a regulator that integrates calcium homeostasis. *Am J Nephrol* 28:455–464
 24. Ito S, Fujimori T, Hayashizaki Y, Nabeshima Y (2002) Identification of a novel mouse membrane-bound family 1 glycosidase-like protein, which carries an atypical active site structure. *Biochim Biophys Acta* 1576:341–345
 25. Shiraki-Iida T, Aizawa H, Matsumura Y, Sekine S, Iida A, Anazawa H, Nagai R, Kuro-o M, Nabeshima Y (1998) Structure of the mouse klotho gene and its two transcripts encoding membrane and secreted protein. *FEBS Lett* 424:6–10
 26. Zagrobelny M, Bak S, Møller BL (2008) Cyanogenesis in plants and arthropods. *Phytochemistry* 69:1457–1468
 27. Marana SR, Jacobs-Lorena M, Terra WR, Ferreira C (2001) Amino acid residues involved in substrate binding and catalysis in an insect digestive β -glycosidase. *Biochim Biophys Acta* 1545:41–52
 28. Ferreira AHP, Marana SR, Terra WR, Ferreira C (2001) Purification, molecular cloning, and properties of a β -glycosidase isolated from midgut lumen of *Tenebrio molitor* (Coleoptera) larvae. *Insect Biochem Mol Biol* 31:1065–1076
 29. Jones AME, Bridges M, Bones AM, Cole R, Rossiter JT (2001) Purification and characterisation of a non-plant myrosinase from the cabbage aphid *Brevicoryne brassicae*. *Insect Biochem Mol Biol* 31:1–5
 30. Malboobi MA, Lefebvre DD (1997) A phosphate-starvation inducible β -glucosidase gene (*psr3.2*) isolated from *Arabidopsis thaliana* is a member of a distinct subfamily of the BGA family. *Plant Mol Biol* 34:57–68
 31. van de Ven WT, LeVesque CS, Perring TM, Walling LL (2000) Local and systemic changes in squash gene expression in response to silver winged whitefly feeding. *Plant Cell* 12:1409–1423
 32. Kawasaki S, Borchert C, Deyholos M, Wang H, Brazille S, Kawai K, Galbraith D, Bohnert HJ (2001) Gene expression profiles during the initial phase of salt stress in rice. *Plant Cell* 13:889–905
 33. Thorlby G, Fourier N, Warren G (2004) The SENSITIVE TO FREEZING2 gene, required for freezing tolerance in *Arabidopsis thaliana*, encodes a beta-glucosidase. *Plant Cell* 16:2192–2203
 34. Lipka V, Dittgen J, Bednarek P, Bhat R, Wiermer M, Stein M, Landtag J, Brandt W, Rosahl S, Scheel D, Llorente F, Molina A, Parker J, Somerville S, Schulze-Lefert P (2005) Pre- and postinvasion defenses both contribute to nonhost resistance in *Arabidopsis*. *Science* 310:1180–1183
 35. Xu Z, Escamilla-Treviño LL, Zeng L, Lalgondar M, Bevan DR, Winkler 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
 36. Opasiri R, Pomthong B, Okoksoong 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
 37. Marques AR, Coutinho PM, Videira P, Fialho AM, S-Correia I (2003) *Sphingomonas paucimobilis* beta-glucosidase Bgl1: a member of a new bacterial subfamily in glycoside hydrolase family 1. *Biochem J* 370:793–804
 38. Kuntothom T, Luang S, Harvey AJ, Fincher GB, Opasiri R, Hrmova M, Ketudat Cairns JR (2009) Rice family GH1 glycosyl hydrolases with β -D-glucosidase and β -D-mannosidase activities. *Arch Biochem Biophys* 491:84–95
 39. Arthan D, Kittakoop P, Esen A, Svasti J (2006) Furostanol glycoside 26-O- β -glucosidase from the leaves of *Solanum torvum*. *Phytochemistry* 67:27–33
 40. Niemeyer HM (1988) Hydroxamic acids (4-hydroxy-1,4-benzoxazin-3-ones), defense chemicals in the Gramineae. *Phytochemistry* 27:3349–3358
 41. Poulton JE (1990) Cyanogenesis in plants. *Plant Physiol* 94:401–405
 42. Morant AV, Jørgensen K, Jørgensen C, Paquette SM, Sanchéz-Perez R, Møller BL, Bak S (2008) β -Glucosidases as detonators of plant chemical defense. *Phytochemistry* 69:1795–1813
 43. Sheremeti I, Venus Y, Drzewiecki C, Tripathi W, Dan VM, Nitz I, Varma A, Grundler F, Oelmüller R (2008) PYK10, a β -glucosidase located in the endoplasmatic reticulum, is crucial for the beneficial interaction between *Arabidopsis thaliana* and the endophytic fungus *Piriformospora indica*. *Plant J* 54:428–439
 44. Morant AV, Bjarnholt N, Kragh ME, Kjaergaard CH, Jørgensen K, Paquette SM, Piotrowski M, Imbert A, Olsen CE, Møller BL, Bak S (2008) The beta-glucosidases responsible for

- bioactivation of hydroxynitrile glucosides in *Lotus japonicus*. *Plant Physiol* 147:1072–1091
45. Suzuki H, Takahashi S, Watanabe R, Fukushima Y, Fujita N, Noguchi A, Yokoyama R, Nishitani K, Nishino T, Nakayama T (2006) An isoflavone conjugate-hydrolyzing β -glucosidase from the roots of soybean (*Glycine max*) seedlings. *J Biol Chem* 281:30251–30259
46. Chuankhayan P, Hua Y, Svasti J, Sakdarat S, Sullivan PA, Ketudat Cairns JR (2005) Purification of an isoflavonoid 7-O- β -apiosyl-glucoside β -glycosidase and its substrates from *Dalbergia nigrescens* Kurz. *Phytochemistry* 66:1880–1889
47. Chuankhayan P, Rimlumduan T, Tantanuch W, Mothong N, Kongsaree PT, Metheenukul 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
48. Naoumkina M, Farag MA, Sumner LW, Tang Y, Liu CJ, Dixon RA (2007) Different mechanisms for phytoalexin induction by pathogen and wound signals in *Medicago truncatula*. *Proc Natl Acad Sci USA* 104:17909–17915
49. Esen A (1992) Purification and partial characterization of maize (*Zea mays* L.) β -glucosidase. *Plant Physiol* 98:174–182
50. Babcock GD, Esen A (1994) Substrate specificity of maize β -glucosidase. *Plant Sci* 101:31–39
51. Nikus J, Daniel G, Jonsson LM (2001) Subcellular localization of beta-glucosidase in rye, maize and wheat seedlings. *Plant Physiol* 111:466–472
52. Sue M, Yamazaki K, Yajima S, Nomura T, Matsukawa T, Iwamura H, Miyamoto T (2006) Molecular and structural characterization of hexameric beta-D-glucosidases in wheat and rye. *Plant Physiol* 141:1237–1247
53. Nisius A (1988) The stroma centre in *Avena* plastids: an aggregation of β -glucosidase responsible for the activation of oat-leaf saponins. *Planta* 173:474–481
54. Ahn YO, Shimizu B, Sakata K, Gantulga D, Zhou Z, Bevan DR, Esen A (2010) Scopulin-hydrolyzing β -glucosidases in the roots of Arabidopsis. *Plant Cell Physiol* 51:131–143
55. Hara-Nishimura I, Matsushima R (2003) A wound-inducible organelle derived from endoplasmic reticulum: a plant strategy against environmental stress? *Curr Opin Plant Biol* 6:538–588
56. Matsushima R, Kondo M, Nishimura M, Hara-Nishimura I (2003) A novel ER-derived compartment, the ER body, selectively accumulates a β -glucosidase with an ER retention signal in *Arabidopsis*. *Plant J* 33:493–502
57. Bednarek P, Pišlewski-Bednarek M, Svatoš A, Schneider B, Doubek J, Mansurova M, Humphry M, Consonni C, Panstruga R, Sanchez-Vallet A, Molina A, Schulze-Lefert P (2009) A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science* 323:101–106
58. Nagano AJ, Matsushima R, Hara-Nishimura I (2005) Activation of an ER-body-localized β -glucosidase via a cytosolic binding partner in damaged tissues of *Arabidopsis thaliana*. *Plant Cell Physiol* 46:1140–1148
59. Falk A, Taipalensuu J, Ek B, Lenman M, Rask L (1995) Characterization of rapeseed myrosinase-binding protein. *Planta* 195:387–395
60. Esen A, Blanchard DJ (2000) A specific β -glucosidase-aggregating factor (BGAF) is responsible for the β -glucosidase null phenotype in maize. *Plant Physiol* 122:563–572
61. Blanchard DJ, Cicek M, Chen J, Esen A (2001) Identification of beta-glucosidase aggregating factor (BGAF) and mapping of BGAF binding regions on maize beta-glucosidase. *J Biol Chem* 276:11895–11901
62. Kittur FS, Lalgondar M, Yu HY, Bevan DR, Esen A (2007) Maize β -glucosidase-aggregating factor is a polyspecific jacalin-related chimeric lectin, and its lectin domain is responsible for β -glucosidase aggregation. *J Biol Chem* 282:7299–7311
63. Nagano AJ, Fukao Y, Fujiwara M, Nishimura M, Hara-Nishimura I (2008) Antagonistic jacalin-related lectins regulate the size of ER body-type β -glucosidase complexes in *Arabidopsis thaliana*. *Plant Cell Physiol* 49:969–980
64. Leah R, Kigel J, Svedsen I, Mundy J (1995) Biochemical and molecular characterization of a barley seed β -glucosidase. *J Biol Chem* 270:15789–15797
65. Hrmova M, Harvey AJ, Wang J, Shirley NJ, Jones GP, Stone BA, Hoj PB, Fincher GB (1996) Barley β -D-glucan exohydrolases with β -D-glucosidase activity. *J Biol Chem* 271:5277–5286
66. Hrmova M, MacGregor EA, Biely P, Stewart RJ, Fincher GB (1998) Substrate binding and catalytic mechanism of a barley β -D-glucosidase/(1, 4)- β -D-glucan exohydrolase. *J Biol Chem* 273:11134–11143
67. Hrmova M, Burton RA, Biely P, Lahnstein J, Fincher GB (2006) Hydrolysis of (1,4)- β -D-mannans in barley (*Hordeum vulgare* L.) is mediated by the concerted action of (1,4)- β -D-mannan endohydrolase and β -D-mannosidase. *Biochem J* 399:77–90
68. Akiyama T, Kaku H, Shibuya N (1998) A cell wall-bound β -glucosidase from germinated rice: purification and properties. *Phytochemistry* 48:49–54
69. Opasiri R, Ketudat Cairns JR, Akiyama T, Wara-Aswapati O, Svasti J, Esen A (2003) Characterization of a rice β -glucosidase highly expressed in flower and germinating shoot. *Plant Sci* 165:627–638
70. 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 BGlu1. *Biochem J* 379:125–131
71. Seshadri S, Akiyama T, Opasiri R, Kuaprasert B, Ketudat Cairns J (2009) Structural and enzymatic characterization of Os3BGlu6, a rice β -glucosidase hydrolyzing hydrophobic glycosides and (1 $\rightarrow$ 3)- and (1 $\rightarrow$ 2)-linked disaccharides. *Plant Physiol* 151:47–58
72. Hösel W, Surholt E, Borgmann E (1978) Characterization of β -glucosidase isoenzymes possibly involved in lignification from chick pea (*Cicer arietinum* L.) cell suspension culture. *Eur J Biochem* 84:487–492
73. Dhamawardhana DP, Ellis BE, Carlson JE (1995) A β -glucosidase from lodgepole pine xylem specific for the lignin precursor coniferin. *Plant Physiol* 107:331–339
74. Escamilla-Treviño LL, Chen W, Card ML, Shih MC, Cheng CL, Poulton JE (2006) *Arabidopsis thaliana* β -Glucosidases BGLU45 and BGLU46 hydrolyse monolignol glucosides. *Phytochemistry* 67:1651–1660
75. Schliemann W (1984) Hydrolysis of conjugated gibberellins by β -glucosidases from dwarf rice (*Oryza sativa* L. cv. Tan-gin-bozu). *J Plant Physiol* 116:123–132
76. Brzobohatý B, Moore I, Kristoffersen P, Bako L, Campos N, Schell J, Palme K (1993) Release of active cytokinin by a β -glucosidase localized to the maize root meristem. *Science* 262:1051–1054
77. Dietz K-J, Sauter A, Wichert K, Messdaghi D, Hartung W (2000) Extracellular β -glucosidase activity in barley involved in the hydrolysis of ABA glucose conjugate in leaves. *J Exp Bot* 51:937–944
78. Jakubowska A, Kawalczyk S (2005) A specific enzyme hydrolyzing 6-O(4-O)-indole-3-ylacetyl- β -D-glucose in immature kernels of *Zea mays*. *J Plant Physiol* 162:207–213
79. 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

80. Stöckigt J, Zenk MH (1977) Strictosidine (isovincoside): the key intermediate in the biosynthesis of monoterpenoid indole alkaloids. *J Chem Soc Chem Commun* 1977:646–648
81. Barleben L, Panjikar S, Ruppert M, Koepke J, Stöckigt J (2007) Molecular architecture of strictosidine glucosidase: the gateway to the biosynthesis of the monoterpenoid indole alkaloid family. *Plant Cell* 19:2886–2897
82. Warzecha H, Gerasimenko I, Kutchan TM, Stöckigt J (2000) Molecular cloning and functional bacterial expression of a plant β -glucosidase specifically involved in alkaloid biosynthesis. *Phytochemistry* 54:657–666
83. Nomura T, Quesada AL, Kutchan TM (2008) The new beta-D-glucosidase in terpenoid-isoquinoline alkaloid biosynthesis in *Psychotria ipecacuanha*. *J Biol Chem* 283:34650–34659
84. Reuveni M, Sagi Z, Eynon D, Hetzroni A (1999) β -Glucosidase activity is involved in scent production in Narcissus flowers. *Plant Sci* 147:19–24
85. Mattiacci L, Dicke M, Posthumus MA (1995) Beta-glucosidase: an elicitor of herbivore-induced plant odor that attracts host-searching parasitic wasps. *Proc Natl Acad Sci USA* 92:2036–2040
86. Gilbert HJ, Stålbrand H, Brumer H (2008) How the walls come tumbling down: recent structural biochemistry of plant polysaccharide degradation. *Curr Opin Plant Biol* 11:338–348
87. Doi RH, Kosugi A (2004) Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nat Rev Microbiol* 2:541–551
88. Carvalho AL, Dias FM, Nagy T, Prates JA, Proctor MR, Smith N, Bayer EA, Davies GJ, Ferreira LM, Romão MJ, Fontes CM, Gilbert HJ (2007) Evidence for a dual binding mode of dockerin modules to cohesins. *Proc Natl Acad Sci USA* 2007(104):3089–3094
89. Lyman ES, Li B, Renganathan V (1995) Purification and characterization of a cellulose-binding β -glucosidase from cellulose-degrading cultures of *Phanerochaete chrysosporium*. *Appl Environ Microbiol* 61:2976–2980
90. Igarashi K, Tani T, Kawal R, Samejima M (2003) Family 3 β -glucosidase from cellulose-degrading culture of the white-rot fungus *Phanerochaete chrysosporium*. *J Biosci Bioeng* 95:572–576
91. Tsukada T, Igarashi K, Yoshida M, Samejima M (2006) Molecular cloning and characterization of two intracellular β -glucosidases belonging to glycoside hydrolase family 1 from the basidiomycete *Phanerochaete chrysosporium*. *Appl Microbiol Biotechnol* 73:807–814
92. Günata Z (2003) Flavor enhancement in fruit juices and derived beverages by exogenous glycosidases and consequences of the use of enzyme preparations. In: Whitaker JR, Voragen AGJ, Wong DWS (eds) *Handbook of food enzymology*. Marcel Dekker Inc., New York, pp 303–330
93. Esen A (2003) β -Glucosidases. In: Whitaker JR, Voragen AGJ, Wong DWS (eds) *Handbook of food enzymology*. Marcel Dekker Inc., New York, pp 791–804
94. Yasumoto K, Tsuji H, Iwami K, Mitsuda H (1977) Isolation from rice bran of a bound form of vitamin B6 and its identification as 5'-O- β -D-glucopyranosyl-pyridoxine. *Agric Biol Chem* 41:1061–1067
95. Chuankhayon P, Rimlumduan T, Svasti J, Ketudat Cairns JR (2007) Hydrolysis of soybean isoflavonoid glycosides by *Dalbergia* β -glucosidases. *J Agric Food Chem* 55:2407–2412
96. Ismail B, Hayes K (2005) β -Glycosidase activity toward different glycosidic forms of isoflavones. *J Agric Food Chem* 53:4918–4924
97. Higdon JV, Delage B, Williams DE, Dashwood RH (2007) Cruciferous vegetables and human cancer risk: epidemiologic evidence and mechanistic basis. *Pharmacol Res* 55:224–236
98. Sánchez-Pérez R, Jørgensen K, Olsen CE, Dicenta F, Möller BL (2008) Bitterness in almonds. *Plant Physiol* 146:1040–1052
99. Crout DH, Vic G (1998) Glycosidases and glycosynthetases in glycoside and oligosaccharide synthesis. *Curr Opin Chem Biol* 2:98–111
100. Henrissat B, Callebaut I, Fabrega S, Lehn P, Mornon 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
101. 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
102. Sanz-Aparicio J, Hermoso JA, Martinez-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
103. Varghese JN, Hrmova M, Fincher GB (1999) Three-dimensional structure of a barley β -D-glucan exohydrolase; a family 3 glycosyl hydrolase. *Structure* 7:179–190
104. Hrmova M, Varghese JN, De Gori R, Smith BJ, Driguez H, Fincher GB (2001) Catalytic mechanisms and reaction intermediates along the hydrolytic pathway of a plant beta-D-glucan glucosylhydrolase. *Structure* 9:1005–1016
105. Park JK, Wang L-X, Patel HV, Roseman S (2002) Molecular cloning and characterization of a unique β -glucosidase from *Vibrio cholerae*. *J Biol Chem* 277:29555–29560
106. Qi M, Jun H-S, Forsbert CW (2008) Cel9D, an atypical 1, 4- β -D-glucan glucosylhydrolase from *Fibrobacter succinogenes*: characteristics, catalytic residues, and synergistic interactions with other cellulases. *J Bacteriol* 190:1976–1984
107. Rye CS, Withers SG (2000) Glycosidase mechanisms. *Curr Opin Chem Biol* 4:573–580
108. Davies GJ, Ducros VM-A, Varrot A, Zechel DL (2003) Mapping the conformational itinerary of β -glucosidases by X-ray crystallography. *Biochem Soc Trans* 31:523–527
109. Czjzek M, Cicek M, Zamboni V, Bevan DR, Henrissat B, Esen A (2000) The mechanism of substrate (aglycone) specificity in β -glucosidases is revealed by crystal structures of mutant maize β -glucosidase-DIMBOA, -DIMBOAGlc, and dhurrin complexes. *Proc Natl Acad Sci USA* 97:13555–13560
110. Czjzek M, Cicek M, Zamboni V, Burmeister WP, Bevan DR, Henrissat B, Esen A (2001) Crystal structure of a monocotyledon (maize ZMGLu1) β -glucosidase and a model of its complex with p-nitrophenyl β -D-thiogluconate. *Biochem J* 354:37–46
111. 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
112. Chuenchor W, Pengthaisong S, Robinson RC, Yuvaniyama J, Oonanant W, Bevan DR, Esen A, Chen C-J, Opasiri R, Svasti J, Ketudat Cairns JR (2008) Structural insights into rice BGLu1 β -glucosidase oligosaccharide hydrolysis and transglycosylation. *J Mol Biol* 377:1200–1215
113. Zechel DL, Boraston AB, Gloster TM, Boraston CM, Macdonald JM, Tilbrook DMG, Stick RV, Davies GJ (2003) Iminosugar glycosidase inhibitors: structural and thermodynamic dissection of the binding of isofagomine and 1-deoxynojirimycin to β -glucosidases. *J Am Chem Soc* 125:14313–14323
114. Vincent F, Gloster TM, Macdonald J, Morland C, Stick RV, Dias FM, Prates JA, Fontes CM, Gilbert HJ, Davies GJ (2004) Common inhibition of both beta-glucosidases and beta-mannosidases by isofagomine lactam reflects different

- conformational itineraries for pyranoside hydrolysis. *ChemBiochem* 5:1596–1599
115. Gloster TM, Macdonald JM, Tarling CA, Stick RV, Withers SG, Davies GJ (2004) Structural, thermodynamic, and kinetic analyses of tetrahydrooxazine-derived inhibitors bound to β -glucosidases. *J Biol Chem* 279:29236–49242
116. Gloster TM, Madsen R, Davies GJ (2006) Dissection of conformationally restricted inhibitors binding to a beta-glucosidase. *ChemBiochem* 7:738–742
117. Gloster TM, Roberts S, Perugino G, Rossi M, Moracci M, Panday N, Terinek M, Vasella A, Davies GJ (2006) Structural, kinetic, and thermodynamic analysis of glucoimidazole-derived glycosidase inhibitors. *Biochemistry* 45:11879–11884
118. Gloster TM, Meloncelli P, Stick RV, Zechel D, Vasella A, Davies GJ (2007) Glycosidase inhibition: an assessment of the binding of 18 putative transition-state mimics. *J Am Chem Soc* 129:2345–2354
119. Withers SG, Street IP, Bird P, Dolphin DH (1987) 2-Deoxy-2-fluoroglucosides: a novel class of mechanism based inhibitors. *J Am Chem Soc* 109:7530–7531
120. Withers SG, Warren RAJ, Street IP, Rupitz K, Kempton JB, 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
121. Burmeister WP, Cottaz S, Driguez H, Iori R, Palmieri S, 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
122. Noguchi J, Hayashi Y, Baba Y, Okino N, Kimura M, Ito M, Kakuta Y (2008) Crystal structure of the covalent intermediate of human cytosolic beta-glucosidase. *Biochem Biophys Res Commun* 374:549–552
123. Keresztessy Z, Kiss L, Hughes MA (1994) Investigation of the active site of the cyanogenic beta-D-glucosidase (linamarase) from *Manihot esculenta* Crantz (cassava). II. Identification of Glu-198 as an active site carboxylate group with acid catalytic function. *Arch Biochem Biophys* 315:323–330
124. Wang Q, Trimbur D, Graham R, Warren RA, Withers SG (1995) Identification of the acid/base catalyst in *Agrobacterium faecalis* beta-glucosidase by kinetic analysis of mutants. *Biochemistry* 34:14554–14562
125. Wang Q, Graham RW, Trimbur D, Warren RAJ, Withers SG (1994) Changing enzymic reaction mechanisms by mutagenesis: conversion of a retaining glucosidase to an inverting enzyme. *J Am Chem Soc* 116:11594–11595
126. Mackenzie LF, Wang Q, Warren RAJ, Withers SG (1998) Glycosynthases: mutant glycosidases for oligosaccharide synthesis. *J Am Chem Soc* 120:5583–5584
127. Ly HD, Withers SG (1999) Mutagenesis of glycosidases. *Annu Rev Biochem* 68:487–522
128. Marana SR (2006) Molecular basis of substrate specificity in family 1 glycoside hydrolases. *IUBMB Life* 58:63–73
129. 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
130. Berrin J-G, Czjzek M, Kroon PA, McLauchlan WR, Puigserver A, Williamson G, Juge N (2003) Substrate (aglycone) specificity of human cytosolic β -glucosidase. *Biochem J* 373:41–48
131. Cicek M, Blanchard D, Bevan DR, Esen A (2000) The aglycone specificity-determining sites are different in 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA)-glucosidase (maize beta-glucosidase) and dhurrinase (sorghum beta-glucosidase). *J Biol Chem* 275:20002–20011
132. Zouhar J, Vévodová J, Marek J, Damborský J, Su X-D, Bryzobohatý B (2001) Insights into the functional architecture of the catalytic center of a maize β -glucosidase Zm-p60.1. *Plant Physiol* 127:973–985
133. Srisomsap C, Svasti J, Surarit R, Champattanachai V, Sawangareetrakul P, Boonpuan K, Subhasitanont P, Chokchai-chamnankit D (1996) Isolation and characterization of an enzyme with beta-glucosidase and beta-fucosidase activities from *Dalbergia cochinchinensis* Pierre. *J Biochem* 119:585–590
134. Zollner H (1989) Handbook of enzyme inhibitors. VCH, Weinheim, pp 94–95
135. 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
136. Kengen SW, Luesink EJ, Stams AJ, Zehnder AJ (1993) Purification and characterization of an extremely thermostable beta-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Eur J Biochem* 213:305–312
137. Nam KH, Kim S-J, Kim M-Y, Kim JH, Yeo Y-S, Lee C-M, Jun H-K, Hwang KY (2008) Crystal structure of engineered beta-glucosidase from soil metagenome. *Proteins* 73:788–793
138. Chi YI, Martinez-Cruz LA, Jancarik J, Swanson RV, Robertson DE, Kim SH (1999) Crystal structure of the beta-glycosidase from the hyperthermophile *Thermosphaera aggregans*: insights into its activity and thermostability. *FEBS Lett* 445:375–383

Article

Crystallisation of Wild-Type and Variant Forms of a Recombinant Plant Enzyme β -D-Glucan Glucohydrolase from Barley (*Hordeum vulgare* L.) and Preliminary X-ray Analysis

Sukanya Luang¹, James R. Ketudat Cairns¹, Victor A. Streltsov² and Maria Hrmova^{3,*}

¹ School of Biochemistry, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand; E-Mail: jeab_bc@hotmail.com (S.L.); cairns@sut.ac.th (J.R.K.C.)

² Molecular and Health Technologies, CSIRO-Commonwealth Scientific Research Organization, Victoria 3052, Australia; E-Mail: victor.streltsov@csiro.au

³ Australian Centre for Plant Functional Genomics, University of Adelaide, Waite Campus, Glen Osmond, South Australia 5064, Australia

* Author to whom correspondence should be addressed; E-Mail: maria.hrmova@adelaide.edu.au; Tel.: +61-8-830-372-80; Fax: +61-8-830-371-02.

Received: 28 May 2010; in revised form: 16 July 2010 / Accepted: 16 July 2010 /

Published: 19 July 2010

Abstract: Wild-type and variant crystals of a recombinant enzyme β -D-glucan glucohydrolase from barley (*Hordeum vulgare* L.) were obtained by macroseeding and cross-seeding with microcrystals obtained from native plant protein. Crystals grew to dimensions of up to $500 \times 250 \times 375 \mu\text{m}$ at 277 K in the hanging-drops by vapour-diffusion. Further, the conditions are described that yielded the wild-type crystals with dimensions of $80 \times 40 \times 60 \mu\text{m}$ by self-nucleation vapour-diffusion in sitting-drops at 281 K. The wild-type and recombinant crystals prepared by seeding techniques achieved full size within 5–14 days, while the wild-type crystals grown by self-nucleation appeared after 30 days and reached their maximum size after another two months. Both the wild-type and recombinant variant crystals, the latter altered in the key catalytic and substrate-binding residues Glu220, Trp434 and Arg158/Glu161 belonged to the $P4_32_12$ tetragonal space group, *i.e.*, the space group of the native microcrystals was retained in the newly grown recombinant crystals. The crystals diffracted beyond 1.57–1.95 Å and the cell dimensions were between $a = b = 99.2$ – 100.8 Å and $c = 183.2$ – 183.6 Å . With one molecule in the asymmetric unit, the calculated Matthews coefficients were between 3.4 – $3.5 \text{ Å}^3 \cdot \text{Da}^{-1}$

and the solvent contents varied between 63.4% and 64.5%. The macroseeding and cross-seeding techniques are advantageous, where a limited amount of variant proteins precludes screening of crystallisation conditions, or where variant proteins could not be crystallized.

Key words: macro- and cross-seeding; wild-type and mutant protein; X-ray diffraction

1. Introduction

Plant β -D-glucan glucohydrolase enzymes are classified in the GH3 family of glycoside hydrolases (<http://www.cazy.org/>) that currently includes nearly 3,000 entries [1]. The majority of these entries represent nucleotide sequences obtained from sequencing of bacterial genomes. Almost all GH3 entries have two, three or more individually folded domains, while the spatial arrangement of the domains varies [2,3]. A large proportion of the GH3 entries are enzymes that are variously annotated as β -D-glucosidases, glucan-1,4- β -D-glucosidases, (1,3)- β -D-glucan exohydrolases, (1,3;1,4)- β -D-glucan exohydrolases, exo-1,3-1,4- β -D-glucanases, N-acetyl- β -D-glucosaminidases, xylan-1,4- β -D-xylosidases and α -L-arabinofuranosidases [1]. Although the GH3 family contains predominantly enzymes, their substrate specificities have not been satisfactorily defined and the most of annotations are based on sequence similarities with enzymes for which biochemical data are available [4].

We have focused our attention in the past upon the biochemical and structural characterization of β -D-glucan glucohydrolases from higher plants, in particular on barley β -D-glucan glucohydrolase, isoenzyme ExoI (designated as HvExoI). This enzyme was crystallized from a native plant source obtained from barley seedlings [5], although the procedure for isolation of a crystallisable quality of HvExoI is both time consuming and technically challenging [6]. Over the past ten years several structures (Protein Data Bank references 1EX1, 1IEQ, 1IEV, 1IEW, 1IEX, 1J8V, 1LQ2, 1X38, 1X39) have been solved from the native HvExoI crystals with substrate analogues and mechanism-based inhibitors to explain the enzyme's catalytic mechanism and substrate specificity [2,7-10]. However, a few facets of catalytic and substrate binding mechanisms remain to be explained. To this end we are interested in the specific roles of amino acid residues in the vicinity of the catalytic pair and how specific structural determinants at the entrance of the catalytic pocket control spatial dispositions of isomeric oligosaccharides entering the catalytic site. Further, we have been intrigued by so-called substrate-product trafficking events at the mouth of the catalytic site [2,7]. Here, the glucose product of the hydrolytic reaction, which is released from the non-reducing termini of substrates, remains bound to the enzyme's active site. However, when a new substrate molecule approaches the enzyme, the glucose product diffuses away from the pocket and the incoming substrate enters the active site [4]. Both events are linked and a precise mechanism, how they proceed, is not known. We presume that the incoming substrate binds in the vicinity of the active site and could, by some delicate mechanism, instigate a product for new substrate interchange. We believe that this substrate/product interchange event represents an ideal model system to study, how products and incoming substrates interact in and/or near enzymes' catalytic sites in general.

Recently, we have reported the high level expression of recombinant HvExoI (designated as rHvExoI) from a codon-optimized HvExoI cDNA in protease-deficient *Pichia pastoris* under low

temperature conditions [11]. We expect that rHvExoI could be used as a suitable enzyme model to study the roles of amino acid residues in catalysis and substrate specificity of this enzyme. To this end, we prepared variants of wild-type rHvExoI, altered (into Ala) in the key catalytic and substrate binding residues Glu220, Trp434 and Arg158/Glu161 by site-directed mutagenesis [12]. From the available structural data [2,7-10] we would expect that Glu220, Trp434 and Arg158 are surface exposed, while Glu161 is buried.

In the current work, we describe the techniques and conditions for preparation of the wild-type and variant rHvExoI crystals that showed excellent diffraction parameters. These crystals were prepared by the macroseeding and cross-seeding techniques with microcrystals obtained from a native plant protein. Seeding techniques have previously been used successfully in bio-macromolecular crystallography [13]. The advantage of these techniques is that they use crystals that provide a preformed, regular surface, onto which new molecules may be added in a regular fashion, generally at a lower degree of supersaturation than is required for nucleation [13].

2. Results and Discussion

2.1. Protein Expression and Purification

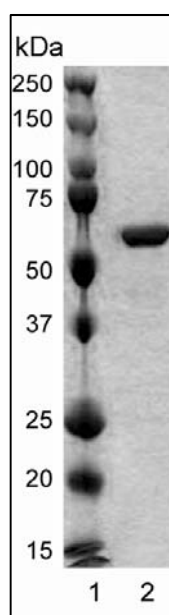
Protein expression in *P. pastoris* at a temperature of 293 K, deglycosylation and purification yielded near-homogenous recombinant rHvExoI (Figure 1) with a molecular mass of 67,169 Da [11]. It is of note that rHvExoI contained at its NH₂-terminus an 8x-His-tag and additional three Ala residues flanking the 8x-His-tag (AHHHHHHHHHAA). The Ala residues resulted from the ligation-based cloning process, while the 8x-His-tag was added for protein purification purposes [11]. This 11-residue tag was not removed from the rHvExoI protein that was subjected to crystallisation trials. Expressed rHvExoI was catalytically competent with the catalytic efficiency value $k_{\text{cat}} \cdot K_M^{-1} = 14 \text{ mM}^{-1} \cdot \text{s}^{-1}$ towards 4-nitrophenyl β -D-glucoside. This second-order rate constant was similar to that reported for native fully N-glycosylated HvExoI [14] and a recombinant N-deglycosylated form [11]. Further, the biophysical properties of rHvExoI such as pH optimum and thermostability were also similar to those reported for native or recombinant HvExoI [11,14].

2.2. Protein Crystallisation

The concentrated rHvExoI was subjected to four types of crystallisation trials (Figures 2 and 3). We firstly set-up the microbatch under paraffin oil trials at 277 K and 287 K with the goal to screen approximately 130 independent conditions, including those that were previously found to be successful with a native plant enzyme [5]. These conditions (1.7 M ammonium sulfate, 75 mM HEPES-NaOH buffer, pH 7, containing 7.5 mM sodium acetate and 1.2% (w/v) PEG 400) previously produced crystals of native plant HvExoI that belonged to the primitive tetragonal space group $P4_32_12$ and yielded high-resolution diffraction patterns [2,5,7]. Nevertheless, despite intensive effort, no diffraction-quality crystals grew, and after approximately 14 days only the formation of a highly intricate interlaced web of thin needles of approximately 40-60 μm in their longest dimensions was observed. Figure 2A and its inset show the appearance of these needles that grew from 1.6 M magnesium sulfate at pH 6.5 and 277 K.

Our second approach to growing the diffraction quality crystals led us to set-up close to 400 trials at the Bio21 Collaborative Crystallisation Centre at 281 K and 293 K by a sitting-drop vapour-diffusion method. After approximately 21 days at 281 K, we could observe formation of short thin needle-shaped crystals that either remained dispersed throughout the droplets (Figure 2B) or formed well-organised round balls (Figure 2C). However, after about a month, in some of the droplets with needles (Figures 2B and 2C) truncated bi-pyramidal crystals formed, which reached dimensions of $80 \times 40 \times 60 \mu\text{m}$ after about 97 days (Figure 2B). It was of note that these crystals were only observed at 281 K in droplets with 1.6 to 2.2 M ammonium sulfate, containing 10 mM malate-MES-Tris buffer, pH 5. These conditions were similar to the conditions that we found previously for native HvExoI [5], except that here the pH value of 5 was more acidic than that used previously (pH 7), and that the protein concentration was almost twice as high ($12.5 \text{ mg}\cdot\text{mL}^{-1}$ versus $6.8 \text{ mg}\cdot\text{mL}^{-1}$). We expected that these truncated bi-pyramidal crystals could belong to the tetragonal $P4_32_12$ space group found for the native protein crystals [5], although at this stage we were not able to collect their diffraction patterns. Notably, in an identical screen at a higher temperature of 293 K the bi-pyramidal crystals were not formed, so it seemed that the temperature was a critical factor for crystal formation of rHvExoI. Thirdly, we also attempted to grow crystals by self-nucleation at 277 K by hanging-drop vapour diffusion and using the conditions developed for native HvExoI [5]. However, we could not observe crystal formation within 180 days and rHvExoI mostly precipitated or the drops remained clear.

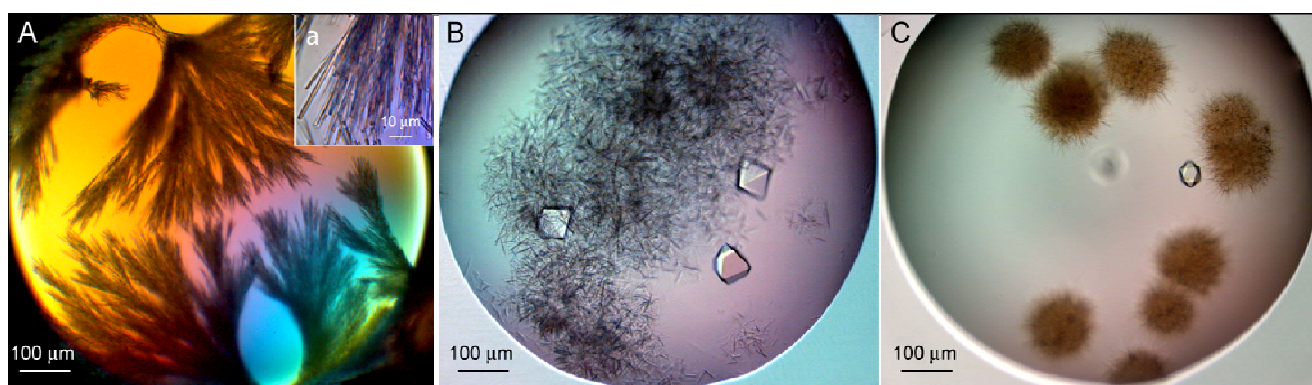
Figure 1. SDS-PAGE profile of recombinant wild-type rHvExoI (lane 2) used for crystallisation by self-nucleation and macro- and cross-seeding using hanging-drop and sitting-drop vapour-diffusion methods. The rHvExoI protein (25 μg) was visualised with a Coomassie Brilliant Blue stain. Protein standards (5 μL of the Precision Plus Protein Standards' stock) are indicated in lane 1.



As no diffraction-quality crystals were obtained with the three crystallisation approaches described above, we turned our attention to seeding at 277 K using a hanging-drop vapour-diffusion method that was used successfully for preparation of large well-diffracting crystals of native HvExoI [7-10]. We

first examined, if we could use microcrystals obtained from native HvExoI to seed the wild-type recombinant rHvExoI protein, despite the differences between the two proteins. These differences included the 11 additional residues of the affinity tag at the NH₂-terminus, as described above. Further, the rHvExoI protein was N-deglycosylated by endoglycosidase H, such that only one N-linked N-acetyl- β -D-glucosaminyl residue [12] remained attached to each of the three N-glycosylation sites Asn221, Asn498 and Asn600 [2,11]. Approximately 48 h after rHvExoI was macroseeded with the native microcrystals of the sizes between $10 \times 5 \times 7.5 \mu\text{m}$ and $20 \times 10 \times 15 \mu\text{m}$, the original native microcrystals started growing in size. The fully grown crystals of wild-type rHvExoI reached dimensions that varied between $100 \times 50 \times 75 \mu\text{m}$ (Figure 3B) and $500 \times 250 \times 375 \mu\text{m}$ after 5 to 7 days and these crystals had a similar bi-pyramidal morphology as the native microcrystals (Figures 3A). Having succeeded in growing wild-type crystals, using native microcrystal seeds, we prepared large recombinant crystals of three rHvExoI variants, specifically Glu220Ala (Figure 3C), Trp434Ala and of the double mutant Arg158Ala/Glu161Ala, also using native seed crystals. The newly grown variant crystals reached sizes of $100 \times 50 \times 75 \mu\text{m}$ (Figure 3C) to $500 \times 250 \times 375 \mu\text{m}$ and showed a similar bi-pyramidal morphology as the native microcrystals. The variant crystals grew slightly slower and reached their maximum dimensions after 10-14 days.

Figure 2. Crystals of recombinant wild-type rHvExoI grown by self-nucleation. Various forms of crystals grown for 14 (A) and 97 (B-C) days using a microbatch-under-paraffin-oil technique from 1.6 M magnesium sulfate, pH 6.5 at 277 K (A) or obtained using sitting-drop vapour-diffusion from 1.8 (B) or 2.2 M (C) ammonium sulfate, both at pH 5 and 281 K. The crystals in (A) formed thin needles of approximately 40-50 μm in their longest dimensions (inset), while in B and C short thin needles are shown that appeared within 14 days and after another 30 days the truncated bi-pyramidal crystals grew that reached dimensions of $80 \times 40 \times 60 \mu\text{m}$ after 97 days.

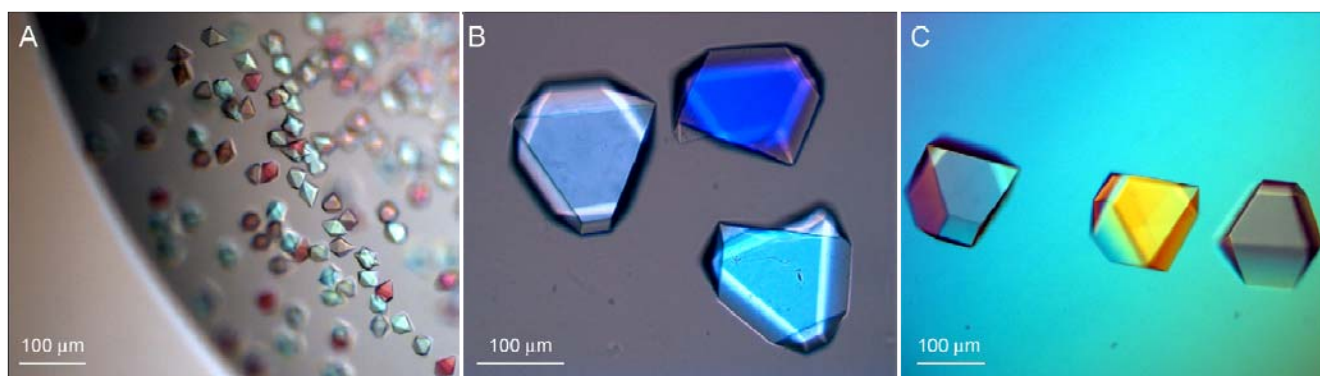


2.3. X-ray Diffraction

Single wild-type and variant rHvExoI crystals were cryo-protected and subjected to diffraction at the MX1 beamline of the Australian Synchrotron. All of the X-ray diffraction data sets were virtually complete beyond 1.57-1.95 Å (Figure 4; Table 1). The HKL2000 indexing and systematic absences calculated that the space groups of the wild-type and variant rHvExoI crystals were consistent with a

primitive tetragonal space group $P4_32_12$ (Figure 4 and Table 1), and as expected, these space group characteristics were similar to their native counterparts [2,7]. In line with these observations was our previous finding that the native HvExoI microcrystals of the sizes $20 \times 10 \times 15 \mu\text{m}$ diffracted beyond $2.80 \text{ \AA}$ on an in-house rotating anode X-ray source and belonged to a primitive tetragonal space group $P4_32_12$ [15]. Thus, our data are in agreement with other reports, where the space group characteristics of macroseeds and fully grown crystals were identical [16–18], or that as a bonus, the resolution of the newly grown crystals has improved [18]. On the contrary, other authors have reported that during seeding trials mutant crystals often crystallized in different space groups than their macroseeds [19].

Figure 3. Microcrystals ($10 \times 5 \times 7.5 \mu\text{m}$ to $20 \times 10 \times 15 \mu\text{m}$) of native HvExoI (A) used to grow the recombinant wild-type (B) and variant Glu220Ala (C) rHvExoI crystals that grew to their full-sizes within 5–14 days. The crystals obtained by seeding in hanging-drops by vapour-diffusion grew to dimensions of up to $500 \times 250 \times 375 \mu\text{m}$ at 277 K .



The best diffraction data, to $1.57 \text{ \AA}$ were collected from the wild-type crystals, followed by the Arg158Ala/Glu161Ala, Glu220Ala and Trp434Ala variants that diffracted to $1.65 \text{ \AA}$, $1.90 \text{ \AA}$ and $1.95 \text{ \AA}$, respectively (Figure 4, Table 1). The lattice dimensions of wild-type and variant crystals varied between $a = b = 99.2\text{--}100.8 \text{ \AA}$ and $c = 183.2\text{--}183.6 \text{ \AA}$, and there appeared to be one molecule in the asymmetric units, according to the Matthews coefficient calculation [20]. The Matthews coefficients of recombinant crystals were between $3.4\text{--}3.5 \text{ \AA}^3 \cdot \text{Da}^{-1}$ with solvent contents of 63.4% to 64.5% . The R_{merge} values of 5.6% to 10.1% were obtained with $\langle I/\sigma(I) \rangle$ of 37.0 to 71.2 , whereas the multiplicity of the individual datasets was well above 20 and varied between 26 to 29 (Table 1). It was of note that the completeness for the highest resolution shells of the wild-type and Glu220Ala datasets was lower, despite the high multiplicity and crystal symmetry, because the data were integrated into the corners of the detector. Also, the mean $\langle I/\sigma(I) \rangle$ values for these datasets indicated that they actually diffracted to higher resolution than that stated in Table 2. It was not surprising that the most favourable diffraction statistics was obtained with the wild-type crystals that were seeded with the wild-type native macroseeds (Figure 4 and Table 1).

Table 1. Data collection statistics from the wild-type and variant rHvExoI crystals, calculated by the HKL2000 suite of programs.

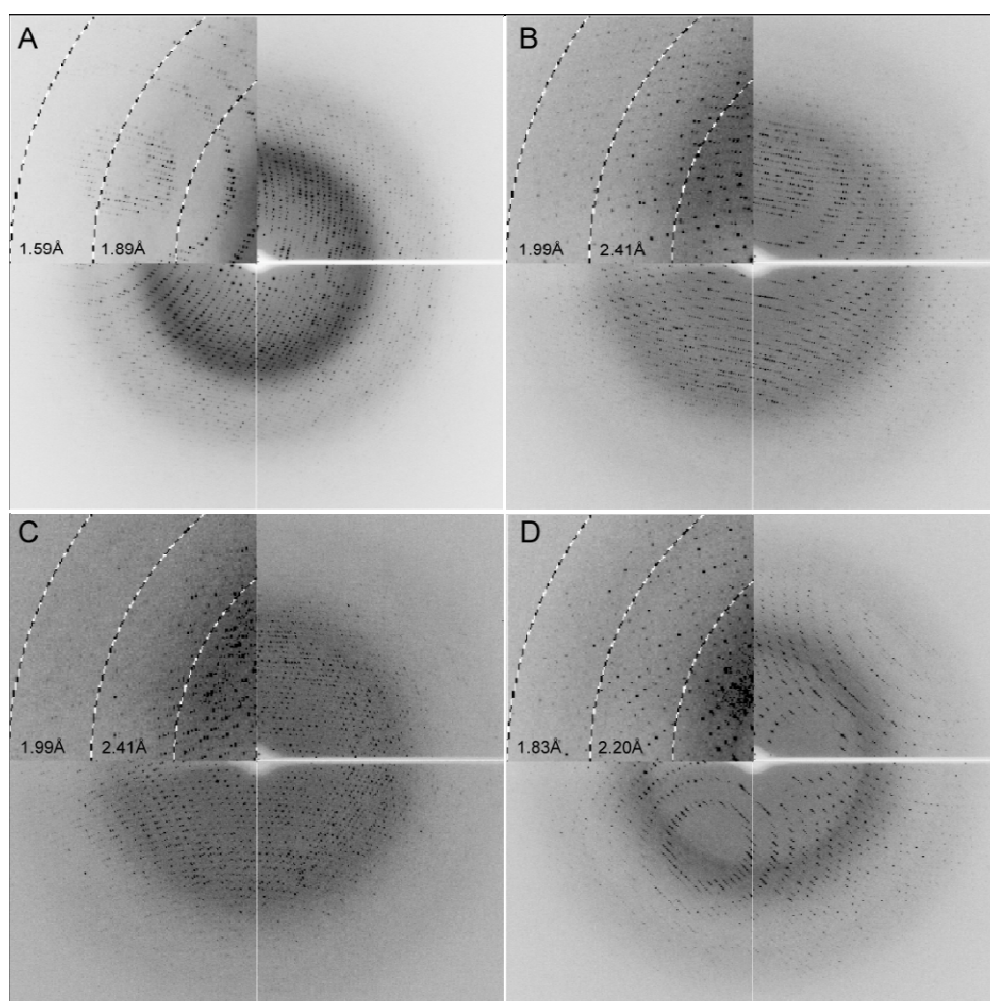
	Wild-type	Glu220Ala	Trp434Ala	Arg158Ala/Glu161Ala
Unique reflections	119968	74171	64460	107601
Resolution ^a (Å)	1.57 (1.6-1.57)	1.90 (1.94-1.90)	1.95 (1.98-1.95)	1.65 (1.68-1.65)
Mean multiplicity ^{a,b}	29 (26)	27 (16)	27 (16)	26 (12)
Completeness ^{a,b} (%)	99 (86)	99 (88)	100 (100)	99.8 (98)
Mean $\langle I/\sigma(I) \rangle$ ^a	71.2 (5.6)	54.8 (4.3)	37.0 (2.2)	58.7 (2.3)
R_{merge} ^{a-c} (%)	6.7 (47)	8.9 (64)	10.1 (87)	5.6 (82)
$a = b$ (Å) ^d	99.2	100.2	100.1	100.8
c (Å) ^d	183.5	183.2	183.6	183.2

^a Numbers in parenthesis represent the values in the highest resolution shell.

^b The numbers were rounded to no decimal place.

^c $R_{\text{merge}} = 100[\sum(I_i - \langle I \rangle)^2 / \sum I_i^2]$ is summed over all independent reflections.

^d The numbers were rounded to the 1st decimal place.

Figure 4. X-ray diffraction patterns of the recombinant wild-type (A), and variant Glu220Ala (B), Trp434Ala (C) and Arg158Ala/Glu161Ala (D) rHvExoI crystals. The left top insets show diffraction intensities to 1.59-1.89 Å (A), 1.99-2.41 Å (B and C) and 1.83-2.20 Å (D).

3. Experimental Section

3.1. Expression and Purification of Wild-Type and Variant Forms of rHvExoI

Wild-type (GenBank accession No. GU441535) and variant codon-optimized cDNAs, encoding a mature barley β -D-glucan exohydrolase I (HvExoI) inserted in pPICZ α BNH₈ expression vectors, were expressed in *P. pastoris*, strain SMD1168H and purified by ion exchange, immobilized metal affinity chromatography (IMAC), N-deglycosylation by endoglycosidase H and a second round of IMAC, as described previously [11]. The variant forms included Glu220Ala, Trp434Ala and a double mutant Arg158Ala/Glu161Ala, for which the constructions of the cDNA fusions will be described elsewhere [12]. At the final purification step before crystallisation, the N-deglycosylated rHvExoI wild-type and variant enzymes were eluted from a BioGel-P100 size-exclusion column with 50 mM sodium acetate buffer, pH 5.25 containing 200 mM NaCl and 1 mM dithiothreitol at a linear flow rate of 0.5 cm·h⁻¹. The protein purities of the rHvExoI fractions were analyzed by SDS-PAGE, using 12.5% w/v polyacrylamide and bis-polyacrylamide gels and standard techniques [6]. The protein concentration was estimated with a Bio-Rad protein assay kit (Bio-Rad Laboratories, Gladesville, New South Wales, Australia) using bovine serum albumin (Sigma Chemical Company, St. Louis, MO, USA) as a standard. The protein standards 'Precision Plus Protein Standards' used for SDS-PAGE were from Bio-Rad Laboratories.

3.2. Enzyme Assays

The activities of pooled and concentrated (using 10 kDa cut-off centrifugal filter units (Millipore, Bedford, MA, USA) wild-type and variant enzymes were assayed against 4-nitrophenyl β -D-glucopyranoside (Sigma) in 50 mM sodium acetate buffer, pH 5.25.

3.3. Crystallisation of Wild-Type and Variant Forms of rHvExoI

Near-homogenous N-deglycosylated wild-type and variant rHvExoI proteins were concentrated to 12.5 mg·mL⁻¹ in 20 mM sodium acetate pH 5.25 and filtered through a 0.22 μ m filter (Millipore). Screening of crystallisation conditions was performed by four experimental approaches. Firstly, initial crystallisation conditions were screened using a microbatch under paraffin oil technique. Here, 1 μ L of precipitant solutions and 1 μ L of the solution containing 12.5 mg·mL⁻¹ of rHvExoI were added into 10 μ L of 100% Paraffin oil (Hampton Research, Aliso Viejo, CA, USA) that was previously dispensed in microbatch 72 well Greiner (Terasaki style) plates (Hampton Research). The formulations of Crystal Screen 2, Crystal Screen Lite and Grid Screen Ammonium Sulfate (Hampton Research) were used as precipitants and the crystals grew at 277 K or 287 K in a vibration-free crystallographic cabinet (Molecular Dimensions, Suffolk, UK). Secondly, crystallisation trials were set-up robotically (Phoenix Nano-Dispenser, Art Robbins Instruments, Sunnyvale, CA, USA) in sitting drops, in which 300 nL droplets of rHvExoI were mixed with the same volumes of precipitants and crystal growth proceeded at 281 K and 293 K at the Bio21 Collaborative Crystallisation Centre (CSIRO, Parkville, Australia) [21]. The precipitants from the PSS_1_Com5 and PS gradient-mid range formulation screens were used, whereas both screens were prepared in-house at the Bio21 Centre

(http://www.csiro.au/c3/Facility/c3_centre_robotic_crystal.htm), following the recommendation from Emerald BioSystems (Bainbridge Island, WA, USA) for the first screen and those reported by Newman [22] for the second screen. The PSS_1_Com5 screen uses the inorganic precipitants such as sulfates, chlorides, citrates and phosphates, but also 2-methyl-2,4-pentanediol, glycerol and polyethyleneglycols (PEGs) in the pH ranges of 5-8.7. On the other hand, the PS gradient-mid range formulation relies on ammonium sulfate and sodium malonate as precipitants in a 10 mM malate-MES-Tris buffer system in the pH ranges of 4.5-9 [22]. Thirdly, attempts were made to grow crystals under the conditions developed for native HvExoI at 277 K using a hanging-drop vapour diffusion method [5]. Lastly, and most importantly, the rHvExoI crystals were grown in hanging drops at 277 K that were seeded with the native HvExoI microcrystals prepared as described previously [5]. The sizes of the microcrystals for the latter conditions varied between $10 \times 5 \times 7.5 \mu\text{m}$ and $20 \times 10 \times 15 \mu\text{m}$. The hanging drops were prepared at 277 K as follows. The volume of 4 to 6 μL of rHvExoI at $12.5 \text{ mg}\cdot\text{mL}^{-1}$ was added to 4 μL of the precipitant solution A (100 mM HEPES-NaOH buffer pH 7, 2.4% (w/v) PEG 400, 1.6 M ammonium sulfate) on 22-mm siliconized circular glass cover slips (Hampton Research). A few microcrystals of native HvExoI were transferred into the hanging drop with a cat whisker. Here, the whisker gently touched the surface of a macroseed stock of native HvExoI and subsequently the whisker was swiftly immersed into a new rHvExoI drop. The cover slips with the seeded hanging drops were placed over 1 mL of reservoir solutions (1.7 M ammonium sulfate in 50 mM HEPES-NaOH buffer, pH 7) contained in the 24 well Linbro plates (Hampton Research), and the wells were sealed with vacuum grease (Dow Corning Corporation, Midland, MI, USA). Crystals from the seeded drops appeared within 5-14 days and were suitable for X-ray data collection. The crystals were photographed through a Leica Laser Microdissection microscope (Leica, North Ryde, Australia) equipped with fluorescence and differential interference contrast.

3.4. X-ray Data Collection and Processing

Single enzyme crystals with the longest dimensions of 100 to 500 μm were cryo-protected in 20% (v/v) glycerol concentration in solution A (as specified above in Section 3.3.) [10] and flash cooled in the cold N_2 stream at the beamline MX1 of the Australian Synchrotron. X-ray diffraction data sets were collected at 0.5° oscillations (1 sec exposures) through 360° on the ADSC Quantum 210r Detector [23]. The data were processed with the HKL2000 suite of programs [24].

4. Conclusions

In summary, excellent X-ray diffraction data were obtained from the recombinant wild-type and variant rHvExoI crystals grown by seeding from a native plant source protein in hanging-drops by vapour-diffusion. The recombinant crystals grew relatively fast and within 5-14 days reached dimensions of up to $500 \times 250 \times 375 \mu\text{m}$. The fully grown recombinant crystals retained the space group characteristics of their native macroseeds and diffracted beyond $1.57 \text{ \AA}$ to $1.95 \text{ \AA}$. As reported for other proteins, this technique could be valuable, where a limited amount of variant proteins is available precluding crystallisation trials, or where variant protein forms could not be crystallized. We project that cross-seeding using native protein as a source of microcrystals could be successfully used for generation of large recombinant wild-type and variant crystals that could potentially yield high

resolution diffraction patterns. Lastly, the diffraction data collected from the wild-type and variant rHvExoI crystals reported here could be used successfully for structure solution. The structural data are currently being prepared for publication [12].

Acknowledgements

The authors acknowledge the MX1 beamline at the Australian Synchrotron (Victoria, Australia) and Janet Newman from the Bio21 Collaborative Crystallisation Centre (Melbourne, Australia). We are also grateful to Gwenda Mayo (Australian Centre for Plant Functional Genomics) for taking photographs of crystals. This work was supported by grants from the Australian Research Council to M.H., the Thailand Research Fund (TRF) to J.R.K.C. and a TRF Royal Golden Jubilee PhD Fellowship to S.L. The constructive help of both anonymous referees is also acknowledged.

References

1. Cantarel, B.L.; Coutinho, P.M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res.* **2008**, *37*, D233-D238.
2. Varghese, J.N.; Hrmova, M.; Fincher, G.B. Three-dimensional structure of a barley β -D-glucan exohydrolase; a family 3 glycosyl hydrolase. *Struct. Fold. Des.* **1999**, *7*, 179-190.
3. Pozzo, T.; Pasten, J.L.; Karlsson, E.N.; Logan, D.T. Structural and functional analyses of beta-glucosidase 3B from *Thermotoga neapolitana*: A thermostable three-domain representative of glycoside hydrolase 3. *J. Mol. Biol.* **2010**, *397*, 724-739.
4. Hrmova, M.; Fincher, G.B. Dissecting the catalytic mechanism of a plant β -D-glucan glucosylhydrolase through structural biology using inhibitors and substrate analogues. *Carbohydr. Res.* **2007**, *305*, 209-221.
5. Hrmova, M.; Varghese, J.N.; Høj, P.B.; Fincher, G.B. Crystallization and preliminary X-ray analysis of β -glucan exohydrolase isoenzyme ExoI from barley (*Hordeum vulgare*). *Acta Cryst.* **1998**, *D54*, 6887-689.
6. Hrmova, M.; Harvey, A.J.; Wang, J.; Shirley, N.J.; Jones, G.P.; Høj, P.B.; Fincher, G.B. Barley β -D-glucan exohydrolases with β -D-glucosidase activity. Purification and determination of primary structure from a cDNA clone. *J. Biol. Chem.* **1996**, *271*, 5277-5286.
7. Hrmova, M.; Varghese, J.N.; DeGori, R.; Smith, B.J.; Driguez, H.; Fincher, G.B. Catalytic mechanisms and reaction intermediates along the hydrolytic pathway of plant β -D-glucan glucosylhydrolase. *Struct. Fold. Des.* **2001**, *9*, 1005-1016.
8. Hrmova, M.; DeGori, R.; Smith, B.J.; Driguez, H.; Varghese, J.N.; Fincher, G.B. Structural basis for a broad specificity in higher plant β -D-glucan glucosylhydrolases. *Plant Cell* **2002**, *14*, 1033-1052.
9. Hrmova, M.; De Gori, R.; Smith, B.J.; Vasella, A.; Varghese, J.N.; Fincher, G.B. Three-dimensional structure of the barley β -D-glucan glucosylhydrolase in complex with a transition state mimic. *J. Biol. Chem.* **2004**, *279*, 4970-4980.

10. Hrmova, M.; Streltsov, V.A.; Smith, B.J.; Vasella, A.; Varghese, J.N.; Fincher, G.B. Structural rationale for low nanomolar binding of transition state mimics to a family GH3 β -D-glucan glucohydrolase from barley. *Biochemistry* **2005**, *44*, 16529-16539.
11. Luang, S.; Hrmova, M.; Ketudat Cairns, J.R. High-level expression of barley β -D-glucan exohydrolase HvExo I from a codon-optimized cDNA in *Pichia pastoris*. *Prot. Expr. Purif.* **2010**, *73*, 90-98.
12. Luang, S.; Streltsov, V.A.; Ketudat Cairns, J.R.; Hrmova, M. Characterization of variant forms of β -D-glucan glucohydrolase from barley (*Hordeum vulgare* L). *Biochem. J.* **2010**, to be submitted.
13. Stura, E.; Wilson, I.A. Analytical and production seeding techniques. *Methods* **2000**, *1*, 38-49.
14. Hrmova, M.; Fincher, G.B. Barley β -D-glucan exohydrolases. Substrate specificity and kinetic properties. *Carbohydr. Res.* **1998**, *305*, 209-221.
15. Varghese, J.N.; van Donkelaar, A.; Balaic, D.X.; Barnea, Z. CSIRO-Commonwealth Scientific Research Organization. Personal communication, Melbourne, Victoria, 2001.
16. Radaev, S.; Agniswamy, J.; Sun, P.D. A case of structure determination using pseudosymmetry. *Acta Cryst.* **2009**, *D65*, 1334-1340.
17. Stec, B.; Holtz, K.M.; Wojciechowski, C.L.; Kantrowitz, E.R. Structure of the wild-type TEM-1 β -lactamase at 1.55 Å and the mutant enzyme Ser70Ala at 2.1 Å suggest the mode of noncovalent catalysis for the mutant enzyme. *Acta Cryst.* **2005**, *D61*, 1072-1079.
18. Heaslet, H.; Rosenfeld, R.; Giffin, M.; Lin, Y.-C.; Tam, K.; Torbett, B.E.; Elder, J.H.; McRee, D.E.; Stout, C.D. Conformational flexibility in the flap domains of ligand free HIV protease. *Acta Cryst.* **2007**, *D63*, 866-875.
19. Mizutani, H.; Saraboji, K.; Malathy Sony, S.M.; Ponnuswamy, M.N.; Kumarevel, T.; Krishna Swamy, B.S.; Simanshu, D.K.; Murthy, M.R.; Kunishima, N. Systematic study on crystal-contact engineering of diphthine synthase: Influence of mutations at crystal-packing regions on X-ray diffraction quality. *Acta Cryst.* **2008**, *D64*, 1020-1033.
20. Matthews, B.W. Solvent content of protein crystals. *J. Mol. Biol.* **1968**, *33*, 491-497.
21. Newman, J.; Pham, T.M.; Peat, T.S. Phoenix experiments: combining the strengths of commercial crystallization automation. *Acta Cryst.* **2008**, *F64*, 991-996.
22. Newman, J. Optimization of buffer solutions for protein crystallization. *Acta Cryst.* **2004**, *D60*, 610-612.
23. McPhillips, T.M.; McPhillips, S.E.; Chiu, H.J.; Cohen, A.E.; Deacon, A.M.; Ellis, P.J.; Garman, E.; Gonzalez, N.K.; Sauter, N.K.; Phizackerley, R.P.; Soltis, S.M.; Kuhn, P. Blu-ice and the distributed control system: Software for data acquisition and instrument control at macromolecular crystallography beamlines. *J. Synchrotron Radiat.* **2002**, *9*, 401-406.
24. Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997**, *276*, 307-326.

Binding of β -D-Glucosides and β -D-Mannosides by Rice and Barley β -D-Glycosidases with Distinct Substrate Specificities[†]

Teerachai Kuntothom,^{‡,§,▽} Michal Raab,^{‡,§} Igor Tvaroška,[§] Sebastien Fort,^{||} Salila Pengthaisong,[‡] Javier Cañada,[⊥] Luis Calle,[⊥] Jesús Jiménez-Barbero,[⊥] James R. Ketudat Cairns,[‡] and Maria Hrmova^{*,@,●}

[‡]School of Biochemistry, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima, Thailand,

[§]Department of Structure and Function of Saccharides, Institute of Chemistry, Center for Glycomics, Slovak Academy of Sciences, Bratislava, Slovak Republic, ^{||}Centre de Recherches sur les Macromolécules Végétales, Grenoble, France, [⊥]Centro de Investigaciones

Biológicas, CSIC, Madrid, Spain, and [@]Australian Centre for Plant Functional Genomics, University of Adelaide, Glen Osmond, Australia. [‡]These authors made equal contributions to this work. [▽]Current address: Department of Chemistry,

Maharakham University, Maharakham, Thailand. [●]Also affiliated with Institute of Chemistry, Slovak Academy of Sciences, Bratislava, Slovak Republic.

Received July 12, 2010; Revised Manuscript Received September 7, 2010

ABSTRACT: Predominantly, rice Os3BGlu7 operates as a β -D-glucosidase (EC 3.2.1.21), while barley HvBII acts as a β -D-mannosidase (EC 3.2.1.25). Saturation transfer difference nuclear magnetic resonance (STD NMR) and transferred nuclear Overhauser effect (trNOE) spectroscopy in conjunction with quantum mechanics/molecular mechanics (QM/MM) modeling and docking at the 6-31+G* level were used to investigate binding of S- and O-linked *gluco*- and *manno*-configured aryl- β -D-glycosides to Os3BGlu7 and HvBII. Kinetic analyses with 4-nitrophenyl β -D-thioglucoside (4NP-S-Glc) and 4-nitrophenyl β -D-thiomannoside (4NP-S-Man) indicated that the inhibitions were competitive with apparent K_i constants of 664 and 710 μ M for Os3BGlu7 and 95 and 266 μ M for HvBII, respectively. The STD NMR and trNOESY experiments revealed that 4NP-S-Glc and 4NP-S-Man bound weakly in ⁴C₁ conformations to Os3BGlu7; 4NP-S-Glc adopted ³S₅ (B_{3,0}) or ¹S₃ (^{1,4}B) conformations, and 4NP-S-Man preferred ⁴C₁ geometry, when bound to HvBII. The QM modeling and docking, based on GLIDE scores, predicted that 4NP-O-Glc, 4NP-O-Man, and 4NP-S-Man bound preferentially in ¹S₃ geometries to both enzymes, contrary to 4NP-S-Glc that could also adopt a ⁴C₁ conformation, although in a “flipped-down” ring position. The experimental and computational data suggested that in glycoside recognition and substrate specificity of Os3BGlu7 and HvBII, a combination of the following determinants is likely to play key roles: (i) the inherent conformational and spatial flexibilities of *gluco*- and *manno*-configured substrates in the enzymes’ active sites, (ii) the subtle differences in the spatial disposition of active site residues and their capacities to form interactions with specific groups of substrates, and (iii) the small variations in the charge distributions and shapes of the catalytic sites.

The glycoside hydrolase GH1 family includes enzymes with approximately 20 known substrate specificities (1), including β -D-glucosidases (EC 3.2.1.21), 6-phospho- β -D-glucosidases (EC 3.2.1.86), β -D-mannosidases (EC 3.2.1.25), β -D-galactosidases (EC 3.2.1.23), β -D-glucuronidases (EC 3.2.1.31), and others (2). A detailed examination of the substrate specificity of the barley β -D-glucosidase isoenzyme β II (HvBII) (3, 4), also designated Hv β MannosI (5), revealed that it exhibits a marked preference for manno-oligosaccharides and that the rate of hydrolysis increases with the degree of polymerization of both cello- and manno-oligosaccharides (3–6). Hence, the substrate specificity and action patterns of HvBII are characteristic of an oligosaccharide

exohydrolase, rather than of an enzyme with a preference for low-molecular mass cello-oligosaccharides. Similar conclusions were drawn for an Os3BGlu7 β -D-glucosidase from rice (also called BGlu1), although binding energies at individual subsites differ somewhat (6, 7). Both plant enzymes are capable of catalyzing transglycosylation reactions with 4NP-O-Glc¹ (3, 4, 6, 7), but not with 4NP-O-Man. The presence of an extended series of subsites in these two plant β -D-glycosidases indicates that their biological

[†]This work was supported by grants from the Australian Research Council to M.H. and from the Thailand Research Fund (BRG5080007) to J.R.K.C. T.K. was sponsored by The Institute for the Promotion of Teaching Science and Technology of Thailand. M.R. and I.T. are thankful for support from the Science and Technology Assistance Agency under Contract APVV-0607-07. J.C., L.C., and J.J.-B. thank the Ministerio de Ciencia e Innovación for support (CTQ2009-08536).

*To whom correspondence should be addressed. Telephone: +61 8 8303 7160. Fax: +61 8 8303 7102. E-mail: maria.hrmova@adelaide.edu.au.

¹Abbreviations: CAZy, Carbohydrate-Active enZymes; DP, degree of polymerization; CORCEMA, Complete Relaxation and Conformational Exchange Matrix; DPGSE, double pulse field-gradient spin-echo; EA, catalytic nucleophile; EB, catalytic acid/base; ESP, electrostatic potential; GH, glycoside hydrolase; HDO, hydrogen/deuterium water; ISPA, isolated spin-pair approximation; NOEs, nuclear Overhauser effects; PDB, Protein Data Bank; rmsd, root-mean-square deviation; QM/MM, quantum mechanics/molecular mechanics; STD NMR, saturation transfer difference nuclear magnetic resonance; trNOESY, transferred nuclear Overhauser effect spectroscopy; 1D, one-dimensional; 2D, two-dimensional; 3D, three-dimensional; 4NP, 4-nitrophenyl; 4NP-O-Glc, 4-nitrophenyl β -D-glucopyranoside; 4NP-O-Man, 4-nitrophenyl β -D-mannopyranoside; 4NP-S-Glc, 4-nitrophenyl β -D-thioglucopyranoside; 4NP-S-Man, 4-nitrophenyl β -D-thiomannopyranoside.

functions could lie in the hydrolysis of longer oligosaccharides, possibly derived from cell wall (1,3;1,4)-glucans and (1,4)- β -D-(gluco)mannans (3–6, 8).

Substrate hydrolysis in the GH1 family proceeds with retention of anomeric configuration in a two-step double-displacement catalytic mechanism (9). The catalytic event advances with participation of a pair of glutamate residues, a catalytic acid/base and a catalytic nucleophile. The exceptions to this rule in the GH1 family are plant myrosinases, for example, that from *Sinapsis alba* (10), which do not have proton donors. Most information about the catalytic mechanism of the GH1 β -D-glucosidases has been derived from chemical modification (11, 12), mutagenesis (13), molecular modeling (3, 14), or X-ray diffraction studies (e.g., refs 15 and 16). The catalysts in a range of β -D-glucosidases were determined using mechanism-based inhibitors; e.g., the catalytic nucleophile in the *Agrobacterium* sp. β -D-glucosidase (17) and rice Os3BGlu7 (16) were identified with 2',4'-dinitrophenyl 2-deoxy-2-fluoro- β -D-glucoside, while conduritol B epoxide was used with the *Aspergillus wentii* β -D-glucosidase (18) and barley HvBII (3). In the HvBII enzyme, the two catalytic amino acid residues, Glu179 and Glu386, are positioned near the bottom of the substrate-binding pocket and are separated by approximately 5–6 Å, although it is not clear how the nonreducing end of the substrate is preferably recognized versus the reducing end by the active site residues (3). It is also not apparent how the remainder of the substrate dissociates from the active site after glucose is cleaved, to provide sufficient space for the product to diffuse out of the pocket.

Approximately 30 unique three-dimensional (3D) structures or molecular models (1) have been reported for the GH1 enzymes that fold into (β/α)₈ barrel projections. In contrast to the open cleft structures of endohydrolases, plant and microbial β -D-glucosidases align their glycosidic substrate in dead-end funnels (3, 19). The substrates are brought into juxtaposition with the catalytic pairs close to a glycosidic linkage at the nonreducing terminal residue (8, 15, 20). Much attention has been devoted to dissecting structural features that are responsible for glycon and aglycon substrate specificities in the GH1 group of enzymes (e.g., refs 8, 12, 15, and 20). However, despite this effort, so far we have not been able to dissect what precise conformations the *gluco*- and *manno*-configured substrates adopt in bound states in the active sites of plant GH1 β -D-glucosidases and β -D-mannosidases, what transition states develop during their catalytic cycles, and what structural determinants underlie the substrate specificities of these two closely related β -D-glycosidases.

One approach that has previously been fruitful for descriptions of hydrolase–glycoside interactions is to use thio analogues that mimic natural substrates (e.g., refs (21–23)). In the work presented here, we have used the S-linked *gluco*- and *manno*-configured aryl- β -D-glycosides and the GH1 enzymes Os3BGlu7 from rice and HvBII from barley. These enzymes represent two types of β -D-glycosidases with distinct substrate specificities; i.e., these catalysts prefer to hydrolyze β -D-glucosides and β -D-mannosides, respectively (3, 4, 6, 7, 24). But contrary to employing X-ray crystallography as an experimental tool that has been used in the majority of previous studies, here we interchangeably used saturation transfer difference nuclear magnetic resonance (STD NMR) with transferred nuclear Overhauser effect (trNOE) spectroscopy, in conjunction with predictive computational quantum mechanics/molecular mechanics (QM/MM) calculations. The approaches specified above determined that O- and S-linked *gluco*- and *manno*-configured aryl- β -D-glycosides

adopted a range of conformations. The resultant data are discussed in relation to the substrate preferences of the Os3BGlu7 and HvBII enzymes.

EXPERIMENTAL PROCEDURES

Materials. 4-Nitrophenyl β -D-glucopyranoside (4NP-O-Glc) and 4-nitrophenyl β -D-mannopyranoside (4NP-O-Man) were obtained from Sigma Chemical Co. (St. Louis, MO), and 4NP-S-Glc and 4NP-S-Man were prepared through the published organo-synthetic procedures for 1-thioglycosides (25). The source of all other chemicals was specified elsewhere (4, 6).

Cloning, Heterologous Expression, and Purification. The cDNAs encoding Os3BGlu7 and HvBII were subcloned into the pET32a expression vector encoding a thioredoxin folding partner protein (Novagen, Madison, WI), at the *Nco*I and *Xho*I restriction sites as described previously (6). The HvBII and Os3BGlu7 cDNAs encoding mature proteins corresponded to the EU807965 and U28047 sequences, respectively. Protein expression was conducted after chemical transformation in *Escherichia coli* Origami(DE3) cells (Novagen) and selection on Luria-Bertani agar containing 50 μ g/mL ampicillin, 15 μ g/mL kanamycin, and 12.5 μ g/mL tetracycline. The selected clones were induced for 8 h at 20 °C with 0.3 mM isopropyl β -D-thiogalactopyranoside as previously described (24). The Os3BGlu7 and HvBII enzymes were synthesized as His₆-tagged fusion proteins that were purified by immobilized metal affinity chromatography using a Talon resin (Invitrogen Life Technologies, Carlsbad, CA). The NH₂-terminal thioredoxin-His tags were removed by digestion with enterokinase (New England Biolabs, Beverly, MA) (6, 26). The recombinant enzymes were reconstituted in 50 mM sodium acetate buffer at their pH optima, i.e., at pH 5.0 for Os3BGlu7 and pH 4.0 for HvBII (6, 7).

Kinetic Constants and Inactivation by Thio Inhibitors. Kinetic constants of Os3BGlu7 and HvBII were determined as previously detailed (6). The reaction mixtures contained 0.038–1.5 mM (0.2–3 times the K_m value) 4NP-O-Glc or 0.042–3.81 mM (0.2–3 times the K_m value) 4NP-O-Man, 0.016% (w/v) BSA, and 2–4 pmol (24–48 nM) of Os3BGlu7 or HvBII in 50 mM sodium acetate buffer at pH 5.0 or 4.0, respectively. The enzyme inactivation constants were determined using 0–600 μ M 4NP-S-Glc or 4NP-S-Man. Each inhibitor was tested at four concentrations that were 0.4–3 times the individual K_i (dissociation constant of the enzyme–inhibitor complex) values. The reactions were developed with 2 M Na₂CO₃, and the amount of 4-nitrophenol (4NP) released was measured in microtiter plates using a PolarStar Optima Plate Reader (BMG LabTech, Offenburg, Germany). The enzyme activity was monitored at 405 nm, and the K_i values were determined from Dixon plots (inhibitor concentrations vs 1/v). Substrate hydrolyses during incubations never exceeded 10% of their initial concentrations, and kinetic constants were determined in triplicate at 30 °C. Hydrolytic and inhibition parameters were calculated by non-linear regression of the Michaelis–Menten curves with Graft version 5.0 (Erithacus Software, Horley, Surrey, U.K.).

STD NMR. STD NMR experiments were performed in a phosphate buffer of two different pH values as specified below (D₂O, uncorrected for isotope effects), between 290 and 300 K without saturation of a residual HDO (hydrogen/deuterium water) signal at 4NP-S-Glc or 4NP-S-Man to enzyme molar ratios between 20:1 and 50:1. A train of Gaussian-shaped pulses of 50 ms for each measurement was employed, with a total saturation time of the protein envelope of 2 s and a maximum B1

field strength of 50 Hz. An off-resonance frequency (δ) of 40 ppm and an on-resonance frequency (δ) of -1.0 ppm (protein aliphatic signal region) were applied. In all cases, the line broadening of inhibitor protons was monitored. In each case, the intensities were normalized with respect to the strongest response, which always corresponded to one of the H2 protons. The STD NMR experiments were repeated twice and the results averaged.

For the nuclear Overhauser effects (NOEs) in free states, selective experiments that employed a double-pulse field-gradient spin-echo (DPFGSE) module were conducted. NOE intensities were normalized with respect to the diagonal peak at zero mixing time. Selective T_1 measurements were performed on anomeric and several other protons to obtain the values mentioned above. Experimental NOEs were fitted to the double-exponential function $f(t) = p_0(e^{-p_1 t})(1 - e^{-p_2 t})$, as described previously (27, 28), where p_0 , p_1 , and p_2 are adjustable parameters. The initial slope was determined from the first derivative at time zero [$f'(0) = p_0 p_2$]. Interproton distances were obtained from the initial slopes by employing the isolated spin-pair approximation (ISPA).

The trNOESY experiments with HvBII were performed with freshly prepared mixtures containing HvBII and 4NP-S-Glc or 4NP-S-Man inhibitors, at approximately 30:1 molar ratios, in 20 mM phosphate buffer (pH 5.0) (near the pH optimum), with mixing times of 50, 100, 150, and 200 ms (28), resulting in final thio inhibitor concentrations of approximately 2–3 mM. No purging spin-lock period was employed to remove the NMR signals arising from the HvBII background, because typically the background signals are not observed with large proteins such as HvBII. First, line broadening of the inhibitor protons was monitored after the addition of the enzyme. The theoretical analysis of the trNOEs of 4NP-S-Glc or 4NP-S-Man protons was performed using Complete Relaxation and Conformational Exchange Matrix (CORCEMA), and a relaxation matrix with exchange as described previously (28). Varying exchange rate constants were employed to obtain optimal matches between the experimental and theoretical results of the methylene protons of 4NP-S-Glc or 4NP-S-Man at the C6 atoms, which could be separated by a fixed distance of 1.8 Å. The overall correlation time (τ_c) for the free states of inhibitors was always set to 50 ps for the monosaccharides, while τ_c was set to 10 ns for bound states. To fit the experimental trNOE intensities, off-rate constants between 100 and 1000 s $^{-1}$ were tested. An optimal agreement was achieved for a k_{off} of 100–300 s $^{-1}$.

Interactions of Os3BGlu7 with 4NP-S-Glc and 4NP-S-Man were investigated as described above, except that 20 mM Tris-HCl buffer (pH 8.0) or 20 mM phosphate buffer (pH 6.0) (near the pH optimum) was used, to avoid protons in NMR spectra in the latter case. The other experimental details used were as specified for HvBII.

All spectra were recorded between 290 and 300 K on a Bruker AVANCE 500 MHz spectrometer, equipped with a triple-channel ^1H , ^{13}N , ^{15}N gradient probe and processed with the Topspin Bruker software.

Homology Modeling of HvBII. A rice β -D-glucosidase Os3BGlu7 [PDB entry 2RGL (16)] was used as a structural template. The template sequence was aligned with that of HvBII using ClustalW (29), and the alignment was checked manually to maintain the integrity of secondary structure elements (30). The levels of sequence similarity and identity between HvBII and 2RGL were 92 and 66%, respectively. The structurally aligned sequences were used as input parameters to generate 200 models of HvBII with Modeler 9v2 (31). The top five models with the

lowest value of the Modeler 9v2 Objective Function were assessed by the DOPE module of Modeler 9v2, and the model with the most favorable value was chosen for evaluation. The overall G factors (estimates of stereochemical parameters) in both coordinate files, assessed by PROCHECK (32), were 0.31 and 0.04 for 2RGL and HvBII, respectively, with 99.8 and 0.2% of the residues in the favored, additional, and generously allowed regions and disallowed regions, respectively. WHATIF (33) evaluated the packing environment in the structures and returned quality control values of -0.97 and -0.77 for 2RGL and HvBII, respectively, while Verify 3D (34), which scores the fitness of protein sequences in their 3D environment, showed that 98.3% (2RGL) and 97.7% (HvBII) of the residues had an acceptable 3D–1D score. The z score values (35) reflecting combined statistical potential energy for 2RGL and HvBII were -11.59 and -12.38 , respectively. The rmsd value in the C α positions between the model and its template was 0.147 Å over 470 residues, from a total of 474 and 479 residues in 2RGL and HvBII, respectively, as determined with Stamp (36). Surfaces of Os3BGlu7 and HvBII were calculated with PyMol (37) with a probe radius of 1.4 Å, electrostatic potentials with the Adaptive Poisson–Boltzmann Solver (the dielectric constants of the solvent and solute were 80 and 2, respectively) (<http://apbs.sourceforge.net/>) implemented in PyMol as a plugin, and mapped on the protein molecular surfaces. The overall buriedness parameters of the HvBII and 2RGL active site funnels were calculated by the PocketPicker plugin (38) in PyMol. Molecular graphics were generated with PyMol and the Maestro graphical interface within the Schrödinger software suite (39).

Computational Docking of 4NP-O- and 4NP-S-Glycosides. Twelve starting structures for 4NP-Glc, 4NP-Man, 4NP-S-Glc, and 4NP-S-Man were constructed (39). For 1-O- and 1-S-glucos- and -mannopyranosides, three ring conformations were considered, namely, $^1\text{S}_3$ (^1B) (1B, 2B, 3B, and 4B), $^3\text{S}_5$ (B_3O) (1S, 2S, 3S, and 4S), and $^4\text{C}_1$ (1C, 2C, 3C, and 4C). All 12 structures were fully optimized using the DFT/M05-2X method (40–42) at the 6-31+G* basis. The electrostatic potential (ESP) charges fitted to the atom centers were calculated at the same level of theory that was used for the optimized structures. Calculations were conducted using Jaguar (39).

The 12 optimized structures of β -D-glycosides were docked in the active sites of the crystal structure of Os3BGlu7 (PDB entry 2RGM) and the HvBII model. The preparation of proteins for docking consisted of the following steps. First, both protein structures were reduced to single monomeric units, and water and bound inhibitor molecules were removed. The protein preparation utility within the Schrödinger suite was used to add hydrogen atoms, explicitly define disulfide linkages, and assign protonation states of residues. The structures of both proteins were minimized through 1000 cycles with MacroModel (39) using the OPLS-AA force field (43). The last structures from minimizations were selected for the calculation of the ESP charges for the residues in the active sites utilizing the hybrid QM/MM approach, as implemented in the QSite subroutine (39). For this purpose, the DFT/M05-2X method with the lacvp basis set was used for calculations, over the QM region that consisted of residues within an ~ 4 Å radius around the -1 subsite. The MM part contained the remainder of the proteins that were treated with an OPLS-AA all-atom force field approximation. The QM/MM boundaries were treated with a hydrogen capping approach. Both proteins were assigned charges from the QM/MM calculations that were used to construct the grid models for docking. The docking sites

Table 1: Kinetic Constants for Hydrolysis and Inhibition by Os3BGlu7 and HvBII

	K_m^a (mM)	k_{cat}^a (s^{-1})	$k_{cat}K_m^{-1b}$ ($s^{-1} mM^{-1}$)	K_i (μM)	ΔG^c ($kJ mol^{-1}$)
Os3BGlu7					
4NP-O-Glc	0.23 ± 0.02	7.90 ± 0.4	35.1 ± 1.0		
4NP-O-Man	1.30 ± 0.1	1.32 ± 0.05	1.01 ± 0.02		
4NP-S-Glc				664.3 ± 64.4	-18.4
4NP-S-Man				710.3 ± 115.1	-18.3
HvBII					
4NP-O-Glc	0.50 ± 0.03	0.50 ± 0.07	1.0 ± 0.05		
4NP-O-Man	0.25 ± 0.01	3.10 ± 0.02	12.7 ± 0.2		
4NP-S-Glc				94.7 ± 16.8	-23.3
4NP-S-Man				265.9 ± 57.2	-20.7

^aRounded to a single decimal digit, when higher than 1.0. ^bData from refs 6 and 7. ^cCalculated according to the equation $\Delta G = -RT \ln(1/K_i)$ (59).

were restricted to a cube with a 12 Å radius around the active site; no additional constraints were defined, and all 12 β -D-glycosides structures were docked into the active sites of both enzymes. The GLIDE program from the Schrödinger suite (39, 44) was used to calculate and evaluate descriptors for docking fitness. The resultant complexes were inspected for proper orientations and distances of bound substrates with respect to the corresponding active site residues.

RESULTS

Kinetic Parameters for Inhibition by β -D-Thioglycosides.

The parameters K_i and ΔG resulting from measurements of inhibition kinetics of rice Os3BGlu7 and barley HvBII with 4NP-S-Glc and 4NP-S-Man are listed in Table 1. The values of the K_i constants varied between 95 and 710×10^{-6} M with both enzymes, and the thioglycosides exhibited competitive-type inhibition (data not shown). Hence, the $v = (V_{max}S)/(K_m(1 + I/K_i) + S)$ equation was used for calculations of K_i constants. The ΔG values of inactivation for HvBII with 4NP-S-Glc and 4NP-S-Man were approximately 2–5 kJ/mol more favorable than the respective values for Os3BGlu7. This difference represented a small but significant variation between the two enzymes (Table 1). Table 1 also contains the k_{cat} and K_m values for hydrolysis of 4NP-O-Glc and 4NP-O-Man that were previously determined (4, 6, 7), but for comparison to our work, these constants are included. The catalytic efficiency factor for Os3BGlu7 with 4NP-O-Glc was ~ 35 -fold higher than that for 4NP-O-Man (6, 7), while for HvBII, this second-order rate constant ($k_{cat}K_m^{-1}$) was ~ 13 -fold higher with 4NP-O-Man than with 4NP-O-Glc (Table 1).

STD and *tr*NOESY NMR Investigations. As a first step, the conformations of free thio inhibitors 4NP-S-Glc and 4NP-S-Man were investigated (Figure 1). The analysis of vicinal proton–proton coupling constants permitted deduction of the exclusive existence of $^4C_1(D)$ chair conformations for both inhibitors in solution, as expected. In these cases, NOE experiments were performed to measure the cross relaxation rates that could be correlated with interproton distances. Positive NOEs were obtained at a variety of mixing times from 200 to 1000 ms (Figure 1), as expected for small molecules that display fast tumbling in solution. Excellent agreement was found between the experimentally estimated distances from NOE signals and those measured for the MM3*-minimized $^4C_1(D)$ 4NP-S-Glc and 4NP-S-Man chair geometries (data not shown).

The STD NMR experiments with HvBII permitted the unambiguous deduction of that the enzyme recognized both inhibitor molecules, because clear STD signals were observed for all the protons in 4NP-S-Glc and 4NP-S-Man (Figure 2; the data

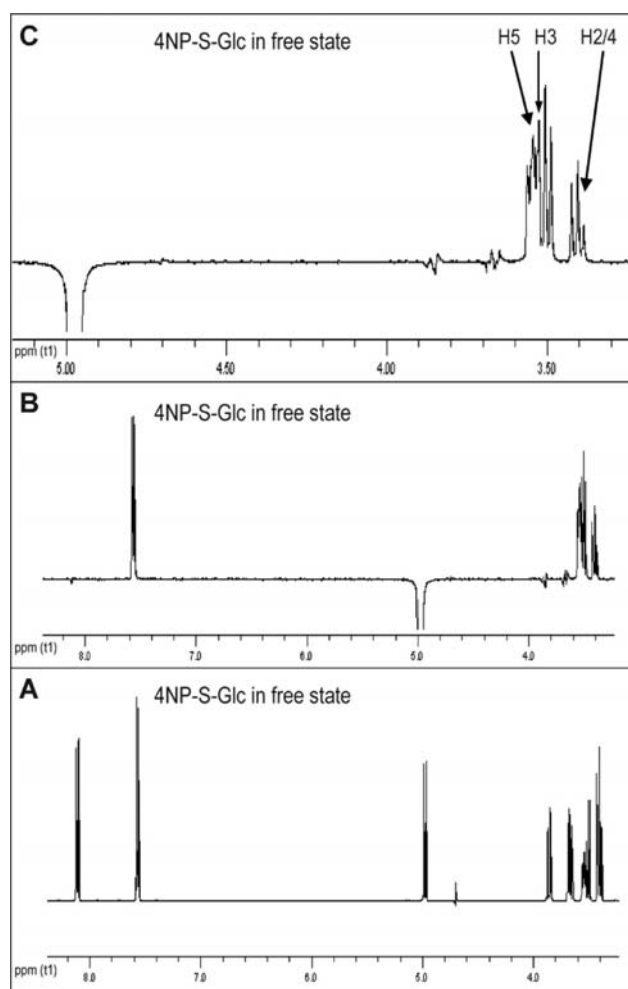


FIGURE 1: Nuclear Overhauser effects observed for free 4NP-S-Glc. (A) The 500 MHz NMR spectrum of 4NP-S-Glc in a free state. Overlapping of H2 and H4 occurs at the high field signal (δ 3.40). (B) NOE enhancements (mixing time of 600 ms) measured upon inversion of an anomeric signal (δ 4.98). Positive NOEs at the ortho-4NP, H5, H3, and H2/4 protons are clearly observed. (C) Expansion of the ring sugar protons, showing that the signal for the H1–H5 NOE was slightly larger than that for the H1–H3 proton pair. The latter pair was significantly larger than that for the H1–H2/4 proton pair. The observations were valid for all mixing times between 200 and 1000 ms.

shown for 4NP-S-Glc). Because of the small size of the thio inhibitors, a clear quantitative definition of the inhibitor epitopes was not attempted, as it was expected that the entire ring of thioglycosides would be recognized by HvBII. Here, a small transfer of saturation to the hydroxymethyl protons was observed with

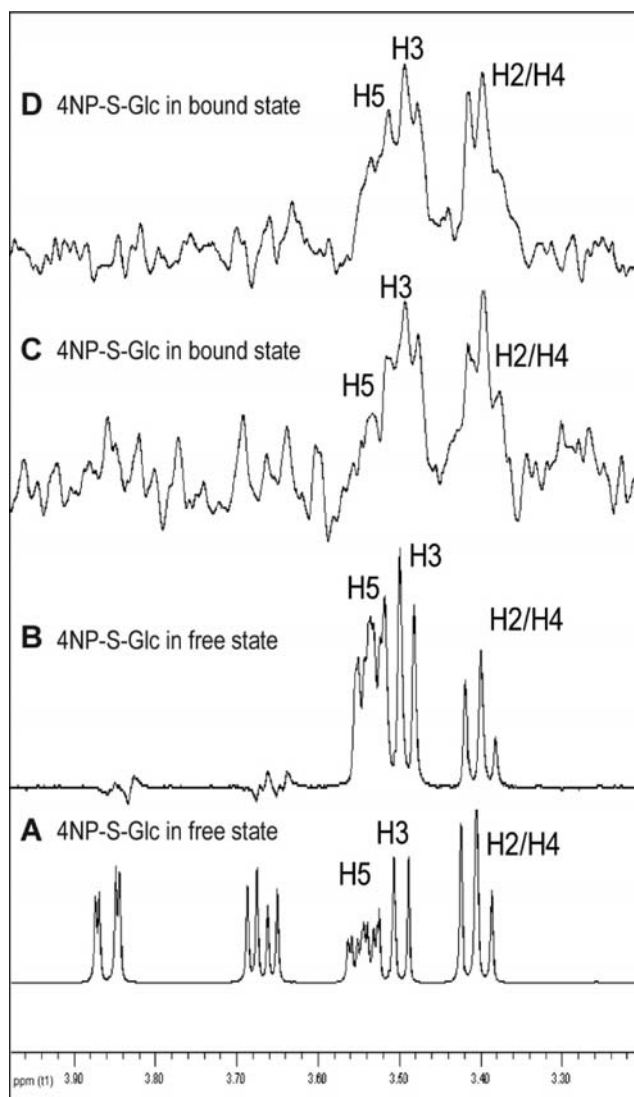


FIGURE 2: Binding of 4NP-S-Glc to HvBII. (A) The 500 MHz NMR spectrum of 4NP-S-Glc in the free state. The arrows indicate the positions of the H1–H5, H1–H3, and H1–H2/4 proton pairs. The overlapping of H2 and H4 occurred at a high field signal (δ 3.40). (B) Section of the NOESY spectrum showing the NOE enhancements (mixing time of 600 ms) measured upon inversion of the anomeric signal of free 4NP-S-Glc (δ 4.98). Positive NOEs at the H1–H5, H1–H3, and H1–H2/4 proton pairs were observed. The NOE signal for the H1–H5 proton pair was slightly larger than that for the H1–H3 proton pair, which was significantly larger than that for the H1–H2 proton pair. (C) Section of the trNOESY spectrum (4NP-S-Glc:HvBII molar ratio of 30:1, mixing time of 100 ms) showing the trNOE enhancements measured upon inversion of the anomeric signal of free 4NP-S-Glc (δ 4.98). (D) Negative NOEs at the H1–H3 and H1–H2/4 proton pairs. The signal for the H1–H5 proton pair was weaker and the signal for the H1–H3 proton pair similar to that of the H1–H2/4 proton pair.

both thio inhibitors, while strong STD signals were observed for the 4NP protons, indicating that this moiety also binds to the HvBII active site (Figure 2A,B; the data shown for 4NP-S-Glc).

Further, strong and negative trNOESY cross-peaks were observed for both inhibitors at a 30:1 ligand:enzyme molar ratio (Figure 2C) with the 100 ms NOE mixing time, as well as at other regimes of 50–200 ms (data not shown). The data for 4NP-S-Glc contrasted with those of 4NP-S-Glc in a free state, where the NOE cross-peaks were positive (cf. Figures 1 and 2). This change from positive to negative cross-peaks upon addition of the

enzyme again indicated binding and that, importantly, these trNOESY experiments could be used to deduce the conformations of bound 4NP-S-Glc and 4NP-S-Man. A CORCEMA-based full-relaxation matrix analysis of the cross-peaks was performed to deduce the experimental proton–proton distances in the bound state, and these cross-peaks were then compared with those that were estimated for all possible chair and skew boat conformers of both inhibitors (Table S1 of the Supporting Information). It was evident that for 4NP-S-Glc, the relative cross-relaxation rates (σ) were rather different from those observed in a free state. Here, the σ values were measured for the H1–H2 pair (σ_{12}), and the data indicated that these values were similar to σ_{13} but were larger than σ_{15} (Table S1 of the Supporting Information). These observations contrasted with those obtained for free 4NP-S-Glc. In this case, σ_{15} was larger than σ_{13} , which in turn was 2-fold larger than σ_{12} . On the basis of these data, we concluded that, in the state bound to HvBII, 4NP-S-Glc could adopt a geometry for which the interproton H1–H2 distance was similar to the H1–H3 distance but was shorter than that between the H1 and H5 protons. These analyses suggested that a conformational change of the six-member ring of 4NP-S-Glc had occurred upon binding to HvBII, and thus, the 4NP-S-Glc inhibitor in the bound state no longer adopted a 4C_1 chair conformation (Table S1 of the Supporting Information).

In contrast, 4NP-S-Man bound to HvBII had very similar relative cross-relaxation rates for the key proton pairs σ_{12} , σ_{13} , and σ_{15} (Table S2 of the Supporting Information). Essentially, identical cross-relaxation rates were observed for both free and bound states, although they had different signs, a negative one for a bound state and a positive one for a free state. This quantitative data suggested that, in principle, no changes in the interproton distances occurred in 4NP-S-Man, when bound to HvBII, and thus no significant conformational geometry changes from the ground 4C_1 conformation occurred during the molecular recognition by HvBII (Table S2 of the Supporting Information).

The nature of the distortion of the pyranose ring of 4NP-S-Glc upon binding to HvBII was assessed by comparison of the experimental interproton distances of bound 4NP-S-Glc with those measured in the MM3*-minimized geometries of the pseudorotational itinerary of 4NP-S-Glc. Here, we considered two extreme chair geometries, ${}^4C_1(D)$ and ${}^1C_4(D)$, and five skew boat conformers, 0S_2 , 1S_3 , 3S_5 , 4S_0 , and 1S_5 . Obviously, a chair conformation could not explain the observed NOE enhancements and the corresponding cross-relaxation rates (Table S1 of the Supporting Information). The best match between the observed and computed distances was found for the 3S_5 skew boat conformer (Table S1 of the Supporting Information and Figure 3A), although given that there were intrinsic uncertainties for the trNOESY experiments and for the conversion of the cross-relaxation rates into distances, the existence of the 1S_3 geometry for bound 4NP-S-Glc could not completely be excluded (Figure 3B). It was characteristic to observe that, for the 3S_5 and 1S_3 geometries, the anomeric C–S bond adopted a pseudoaxial orientation with a proper orientation for the aglycone moiety to depart. The same pseudoequatorial arrangement was also adopted by the hydroxyl groups at the C2, C3, and C4 atoms of the glucose ring. The major difference between the 1S_3 and 3S_5 conformers of 4NP-S-Glc is in the orientation of the C6 hydroxymethyl group, which is pseudoaxial for the 3S_5 conformer and pseudoequatorial for the 1S_3 geometry (Figure 3A,B).

The analogous STD NMR and trNOESY analyses for 4NP-S-Man showed that no major distortion occurred upon binding

to HvBII, because the estimated interproton distances were in full agreement with those computed for the ${}^4\text{C}_1(\text{D})$ chair conformer (Figure 3C). Our findings indicated that the data obtained for 4NP-S-Man could function as an “additional internal reference point” for validating the conclusions deduced for the conformational changes of bound 4NP-S-Glc, given that both experiments were performed under identical experimental conditions.

The STD experiments with Os3BGlu7 interacting with 4NP-S-Glc provided weak transferred signals (data not shown). These experiments proceeded in 20 mM Tris-HCl buffer at a pH value of 8.0 that was similar to the pH value at which the crystal structure of Os3BGlu7 was determined [PDB entry 2RGL (16)]. A high magnetization transfer was observed only for the aromatic ring protons, while the peaks in the carbohydrate part of 4NP-S-Glc were barely visible. Also, the two-dimensional (2D) trNOESY method did not provide cross-peaks for the 4NP-S-Glc bound to Os3BGlu7 at a molar ratio of 20:1 or 50:1. We have further investigated 1D selective NOESY on H1 of the glucose residue using different mixing times; however, no NOESY peaks were observed in the carbohydrate or aromatic regions of the spectra (data not shown). Hence, we concluded that a rather weak binding of 4NP-S-Glc to Os3BGlu7 could be responsible for a lack of measurable signals. These conclusions were supported by the weak K_i constant (664 μM) for binding of 4NP-S-Glc to Os3BGlu7 (Table 1).

One of the reasons that explains why we did not detect the measurable NOE signals with Os3BGlu7 and 4NP-S-Glc at pH 8.0 could be that at this particular pH value, the enzyme's affinity for the thio inhibitor could be low. Hence, the measurements were repeated at pH 6.0 in 20 mM phosphate buffer, whereas under these conditions, the pH value was closer to the pH optimum of 5.0 of Os3BGlu7 (7). Nevertheless, under these conditions, Os3BGlu7 unexpectedly hydrolyzed 4NP-S-Glc to a

small extent, as evidenced by the appearance of new spectral peaks that corresponded to hydrolytic products of 4NP-S-Glc (data not shown). Hence, it could be concluded that with Os3BGlu7, 4NP-S-Glc should be cautiously used as a substrate mimic for investigations of binding interactions.

To investigate binding of 4NP-S-Man to Os3BGlu7, the experimental approach used for binding of 4NP-S-Glc to Os3BGlu7 was adopted, where we first used 20 mM Tris-HCl buffer (pH 8.0) and then 20 mM phosphate buffer (pH 6.0). The STD and transferred NOESY experiments conducted at pH 8.0 with 4NP-S-Man and Os3BGlu7 produced better signals at a molar ratio of 50:1 than at 20:1 (data not shown). A major saturation transfer was again observed on the aromatic protons, followed by the H1 and H2 protons of the mannosyl residue (data not shown). As indicated in Figure 4, in 20 mM phosphate buffer (pH 6) using 1D selective trNOESY conditions, the cross-peaks from H1 to H2, H3, and H5 exhibited the same intensity, as seen for the free inhibitor. These data suggested that no major distortion of the mannosyl ${}^4\text{C}_1$ chair ring had occurred.

QM Modeling and Docking Simulations. The geometry optimizations of the starting ${}^1\text{S}_3$, ${}^3\text{S}_5$, and ${}^4\text{C}_1$ conformers of the O- and S-linked 4NP-Glc and 4NP-Man were conducted at the M05-2X/6-31+G* level of theory, and the calculations yielded the β -D-glycoside conformers in their local minima (Figure 5). The optimized values of selected bond lengths, bond angles, and dihedral angles and ESP partial charges of conformers are summarized in Table S3 of the Supporting Information. A comparison of the structural features of the O- and S-linked 4NP-Glc and 4NP-Man conformers revealed that the glycosides displayed the following structural characteristics. The ESP partial charge on the anomeric carbon of the O-linked glycosides was less negative than the charges on the S-linked glycosides, ranging from 0.36 to 0 for the O-linked versus 0.05 to -0.51 for the S-linked glycosides. Further, the ESP charge on the glycosidic linkage atoms was less negative for the S-linkage (-0.01 to -0.24) than for the O-linkage (-0.29 to -0.54). The partial charge of the endocyclic oxygen was similar for both types of glycosides (-0.36 to -0.53 and -0.22 to -0.40 for the O- and the S-linked glycosides, respectively). The lengths of the C1–O1 linkages in the O-linked glycosides were approximately 0.4–0.5 Å shorter than the lengths of the C1–S1 linkages. Hence, the distances in the O-linked glycosides were between 1.41 and 1.44 Å, while the same distances in the S-linked glycosides were between 1.85 and 1.92 Å (Table S3 of the Supporting Information). The same trend was observed for the lengths of the O1–C7 linkages, which were shorter by approximately 0.44 Å than the lengths of the S1–C7 linkages (Table S3 of the Supporting Information). The calculated C1–X1–C7 angles

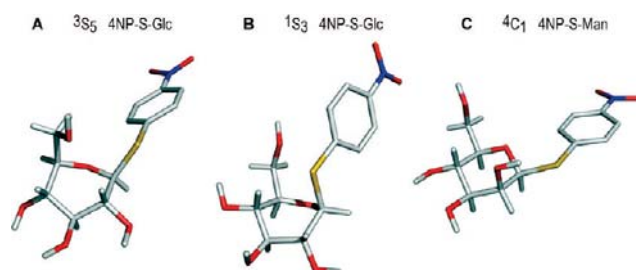


FIGURE 3: Geometries of 4NP-S-Glc and 4NP-S-Man bound to HvBII, as identified by STD NMR. 4NP-S-Glc is shown in the ${}^3\text{S}_5$ (A) and ${}^1\text{S}_3$ (B) conformations, while 4NP-S-Man is in the ${}^4\text{C}_1$ conformation (C).

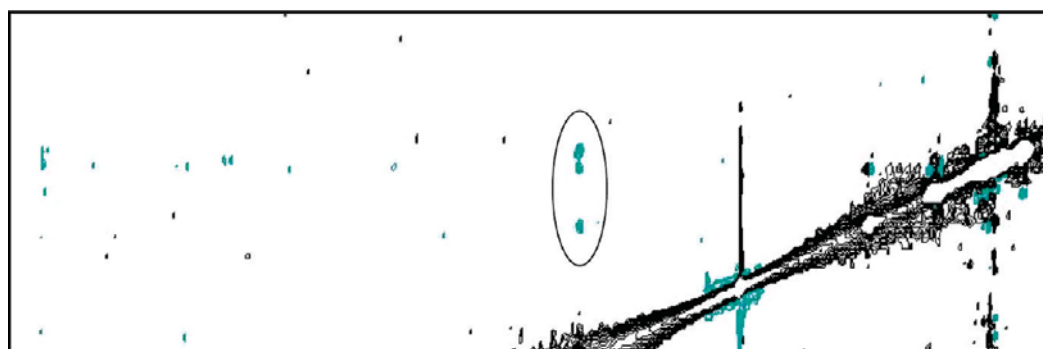


FIGURE 4: Binding of 4NP-S-Man to Os3BGlu7 at a molar ratio of 10:1. In the 2D spectrum, the circled cross-peaks between the anomeric proton and the H2, H3, and H5 protons of 4NP-S-Man showed similar intensities, suggesting that the glycoside was bound in the ${}^4\text{C}_1$ conformation.

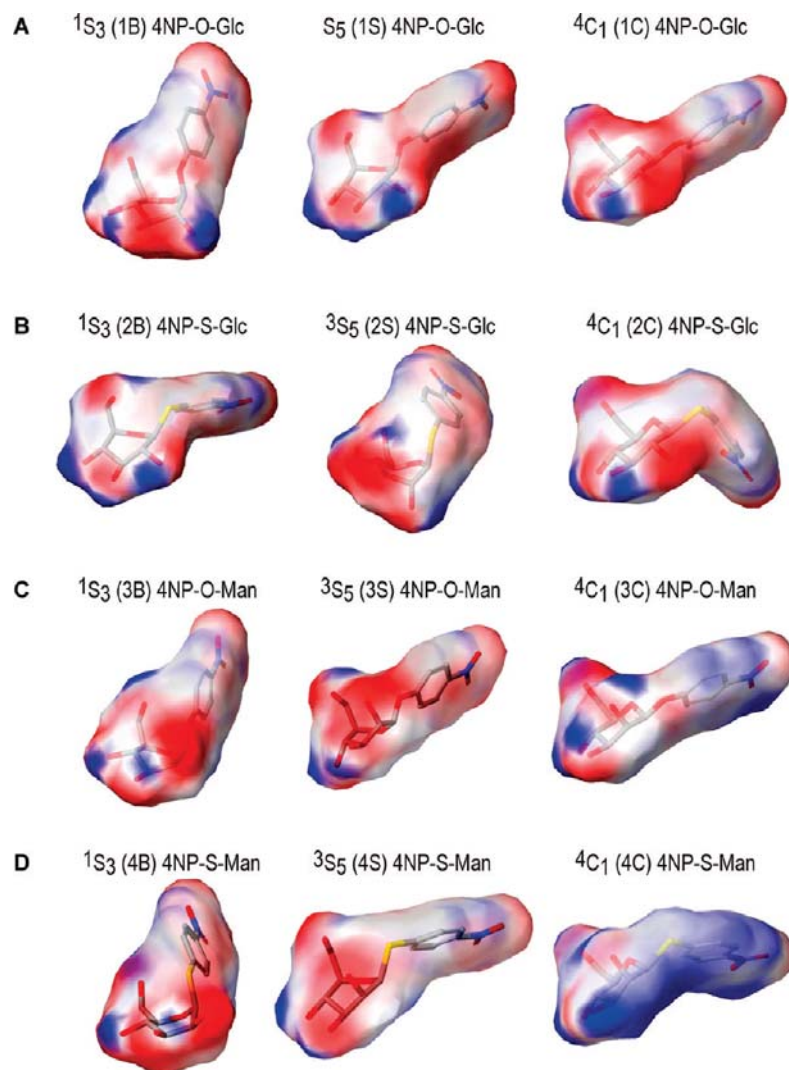


FIGURE 5: Optimized pyranose ring conformations of the *gluco*- and *manno*-configured aryl-glycosides in their local minima that were used for docking: (A) 4NP-O-Glc in 1S_3 (1B), 3S_5 (1S), and 4C_1 (1C); (B) 4NP-S-Glc in 1S_3 (2B), 3S_5 (2S), and 4C_1 (2C); (C) 4NP-O-Man in 1S_3 (3B), 3S_5 (3S), and 4C_1 (3C); and (D) 4NP-S-Man in 1S_3 (4B), 3S_5 (4S), and 4C_1 (4C). The contours denote surfaces and their electrostatic potentials derived from the ESP charges.

of the glycosidic linkages also varied significantly, whereby the C1–S1–C7 bond angles were sharper by as much as 21–30° than the C1–O1–C7 angles. The values were between 87° and 95° for the C1–S1–C7 angle and between 116° and 118° for the C1–O1–C7 angle (Table S3 of the Supporting Information). The minima around the C1–X1 linkages of all aryl-glycoside conformers, predefining their dihedral angles Θ , Φ , and Ψ , corresponded to the synclinal (sc) orientations of the phenyl groups with respect to the pyranose rings, and to the antiperiplanar (ap) orientations at the C2 atoms, in agreement with the exoanomeric effect (45).

Twelve optimized aryl-glycosides in 1S_3 , 3S_5 , and 4C_1 geometries, that is, 4NP-O-Glc in 1B, 1S, and 1C, 4NP-S-Glc in 2B, 2S, and 2C, 4NP-O-Man in 3B, 3S, and 3C, and 4NP-S-Man in 4B, 4S, and 4C, were docked in the active sites of Os3BGlu7 and HvBII, using GLIDE (Table 2 and Figures 6–8). As expected for a *gluco*-O-configured aryl-glucoside, the 1S_3 conformer, also known as the 1,4B (1B) skew boat conformer, fitted most favorably in Os3BGlu7 and HvBII, as supported by the GLIDE score values of –7.2 kcal/mol in both instances; these scores were the highest from all conformers for both enzymes (Table 2). The predicted affinities for the 3S_5 conformer, also known as $B_{3,O}$ (1S)

and 4C_1 (1C) conformers of 4NP-O-Glc, were between –5.2 and –6.0 kcal/mol for Os3BGlu7 and HvBII (Table 2). However, a similar GLIDE score value was predicted for the 1S_3 (2B) (–6.7 kcal/mol) and 4C_1 (2C) (–6.8 kcal/mol) conformers of 4NP-S-Glc docked in Os3BGlu7. Here the difference between the 1S_3 and 4C_1 conformers of 4NP-S-Glc was smaller than the precision of the scoring function. Therefore, on the basis of the GLIDE scores, we concluded that the 1S_3 (2B) or 4C_1 (2C) conformers of 4NP-S-Glc could equally be fitted in Os3BGlu7 (Table 2). On the other hand, for 4NP-S-Glc bound to HvBII, the higher scores were predicted for the 1S_3 and 4C_1 conformers (–5.8 and –5.7 kcal/mol, respectively) than for the 3S_5 conformer (–4.9 kcal/mol). Similarly, for the *manno*-configured aryl-glycosides, the best predicted score values were obtained for the 1S_3 conformations. In the case of 4NP-O-Man in 1S_3 (3B), the binding to Os3BGlu7 was slightly more favorable than that to HvBII, with scores of –7.3 kcal/mol versus –6.9 kcal/mol, while 4NP-O-Man in the 3S_5 and 4C_1 conformers bound less favorably, with affinities of –6.3 kcal/mol for Os3BGlu7 versus –5.6 kcal/mol for HvBII, and –5.5 kcal/mol for Os3BGlu7 versus –5.4 kcal/mol for HvBII (Table 2). The predicted affinities for 4NP-S-Man

Table 2: Summary of Docking of 4NP-O-Glc (¹S₃, ³S₅, and ⁴C₁ or 1B, 1S, and 1C), 4NP-S-Glc (¹S₃, ³S₅, and ⁴C₁ or 2B, 2S, and 2C), 4NP-O-Man (¹S₃, ³S₅, and ⁴C₁ or 3B, 3S, and 3C), and 4NP-S-Man (¹S₃, ³S₅, and ⁴C₁ or 4B, 4S, and 4C) in Os3BGlu7 and HvBII, Illustrating the Relative Conformational “Strain” Energy (Δ*E*) of Docked Conformers at the M05-2X/6-31+G* Level of Theory

	GLIDE score		<i>d</i> ^a (C1⋯O _{EA} ^b)		<i>d</i> ^a (O1/S1⋯O _{EB} ^b)		<i>d</i> ^a (HO ₂ ⋯O _{EB} ^b)		strain energy (Δ <i>E</i> ^a)	
	Os3BGlu7	HvBII	Os3BGlu7	HvBII	Os3BGlu7	HvBII	Os3BGlu7	HvBII	Os3BGlu7	HvBII
1B	−7.24	−7.15	3.19	2.89	3.64	3.42	1.61	3.20	17.21	18.31
1S	−6.02	−5.34	3.51	3.33	3.47	3.16	1.60	1.46	30.90	22.72
1C	−5.82	−5.20	5.13	5.02	4.38	4.02	7.27	3.39	24.05	16.45
2B	−6.66	−5.77	4.67	4.66	4.32	4.03	4.37	3.14	26.30	26.83
2S	−6.43	−4.86	4.64	3.48	3.40	3.37	7.08	1.50	26.58	33.80
2C	−6.76	−5.74	3.84	4.82	4.37	4.90	7.25	3.85	23.30	24.29
3B	−7.28	−6.89	3.33	3.00	4.22	3.69	4.10	3.01	30.44	26.10
3S	−6.33	−5.57	3.87	3.40	3.89	3.23	4.49	3.12	21.21	14.84
3C	−5.51	−5.36	4.39	4.92	3.51	3.87	4.28	2.46	25.61	21.80
4B	−7.56	−7.26	3.10	2.95	3.59	3.27	1.47	1.56	15.31	22.46
4S	−5.97	−5.70	4.96	4.40	3.13	4.46	4.21	5.20	26.39	15.66
4C	−6.61	−5.68	4.42	4.72	5.93	3.14	5.01	1.70	27.35	26.38

^aLengths (*d*) in angstroms and GLIDE scores and strain energies (Δ*E*) in kilocalories per mole. ^bEA denotes the catalytic nucleophile residues Glu386 in Os3BGlu7 and Glu389 in HvBII, and EB denotes the catalytic acid/base residues Glu176 in Os3BGlu7 and Glu179 in HvBII.

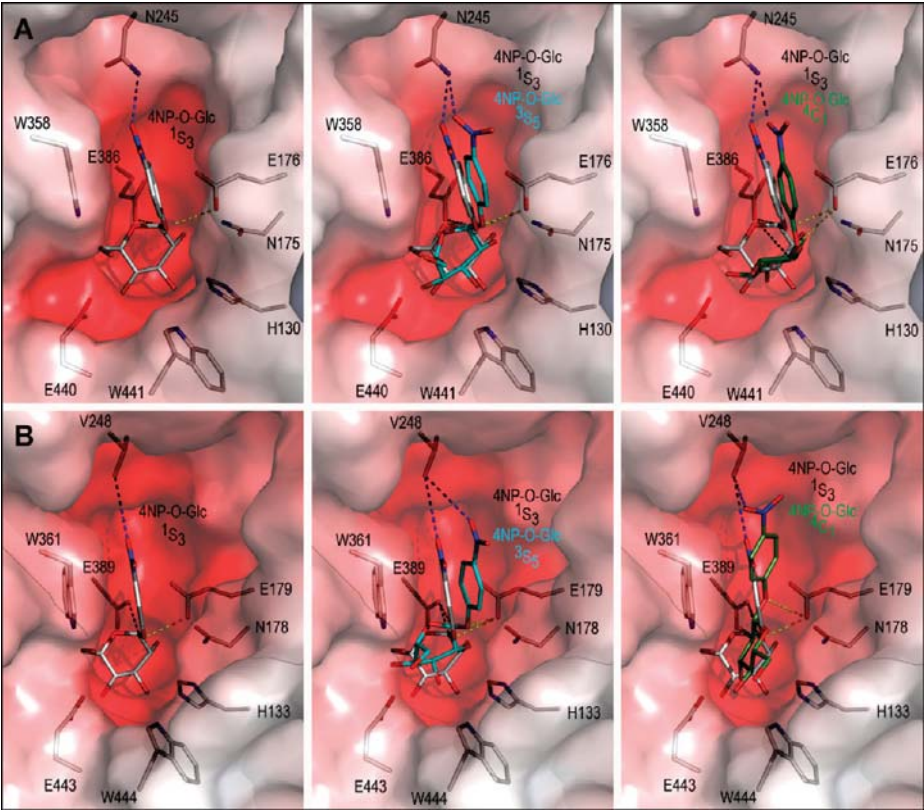


FIGURE 6: Predicted binding positions of 4NP-O-Glc in the active sites of Os3BGlu7 (A) and HvBII (B). The positions of 4NP-O-Glc (A) [left panel, in ¹S₃ (cpk); middle panel, in ¹S₃ (cpk) and ³S₅ (cpk cyan); right panel, in ¹S₃ (cpk) and ⁴C₁ (cpk green)] were predicted by GLIDE. The separations between the catalytic nucleophiles E386 (Os3BGlu7) and E389 (HvBII) and acid/base residues E176 (Os3BGlu7) and E179 (HvBII), and the O1 and C1 atoms in 4NP-O-Glc, respectively, are shown as yellow and black dashes. Blue dashes indicate separations between the O atom of the 4NP group and the nearest amino acid residue. Blue and red patches indicate electropositive and electronegative areas contoured at +5 and −10*kT*/e, respectively. Electrostatic potentials were calculated with Adaptive Poisson–Boltzmann Solver implemented in PyMol.

in the ⁴C₁ (4B) conformer were very similar to those of 3B with both enzymes, whereas the highest values of −7.6 kcal/mol (Os3BGlu7) and −7.3 kcal/mol (HvBII) were calculated for 4NP-S-Man in the ¹S₃ conformer. The GLIDE scores of 4NP-S-Man in the ³S₅ conformer with Os3BGlu7 (−6.0 kcal/mol) and HvBII (−5.7 kcal/mol) ranked relatively close to each other, although 4NP-S-Man bound in the ⁴C₁ conformer

was more favored for Os3BGlu7 (−6.6 kcal/mol) than HvBII (−5.7 kcal/mol).

To evaluate the predicted orientations of docked *gluco*- and *manno*-configured O- and S-linked 4NP-glycosides, closer inspections that are summarized in Figures 6–8 were conducted. The best docking positions of 4NP-O-Glc in the ¹S₃ (1B) geometry indicated that the pyranose rings were located in the intermediate

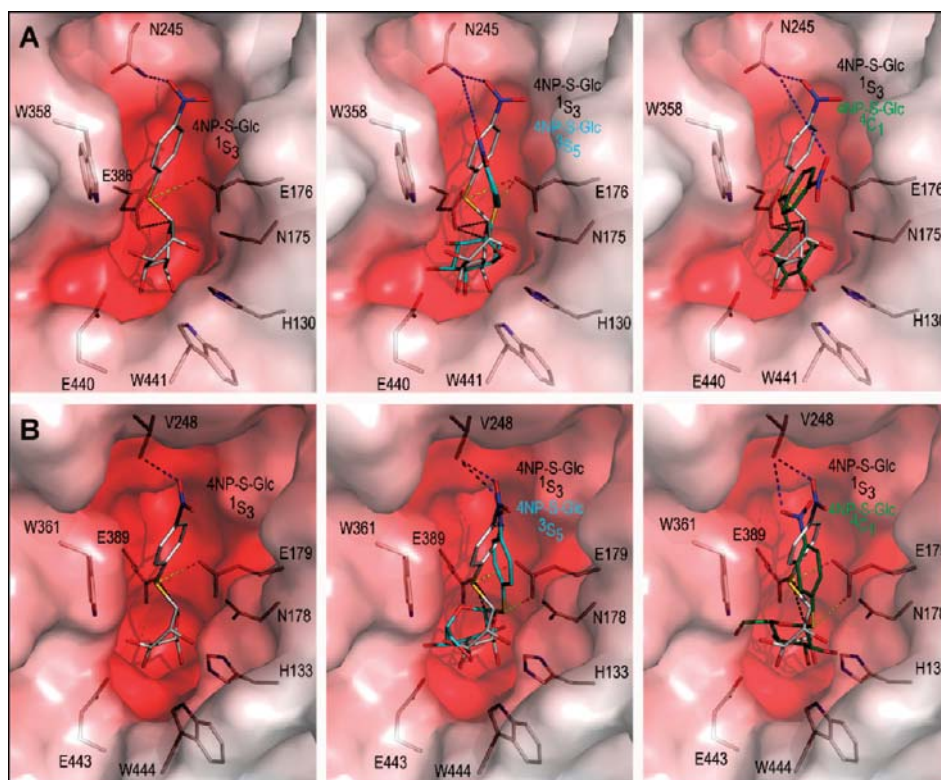


FIGURE 7: Predicted binding positions of 4NP-S-Glc in the active sites of Os3BGlu7 (A) and HvBII (B). In panel A, the positions of 4NP-S-Glc in (left) 1S_3 (cpk), (middle) 1S_3 (cpk) and 3S_5 (cpk cyan), and (right) 1S_3 (cpk) and 4C_1 (cpk green) were predicted by GLIDE. The separations between the catalytic nucleophiles E386 (Os3BGlu7) and E389 (HvBII) and acid/base residues E176 (Os3BGlu7) and E179 (HvBII), and the S1 and C1 atoms in 4NP-S-Glc, are shown as yellow and black dashes, respectively. Blue dashes indicate the separations between the O atom of the 4NP group and the nearest residue.

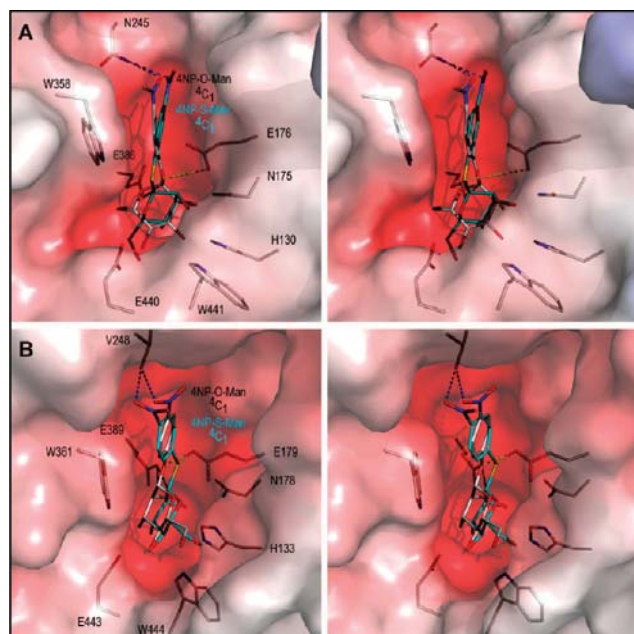


FIGURE 8: Stereoviews of predicted binding positions of 4NP-O-Man and 4NP-S-Man in the active sites of Os3BGlu7 (A) and HvBII (B). The positions of 4NP-O-Man (cpk) and 4NP-S-Man in the 4C_1 conformer (cpk cyan) were predicted by GLIDE. The separations between the catalytic nucleophiles E386 (Os3BGlu7), and E389 (HvBII) and acid/base residues E176 (Os3BGlu7) and E179 (HvBII), and the O1 (S1 for 4NP-S-Man) and C1 atoms in 4NP-mannosides, respectively, are shown as yellow and black dashes. Blue dashes indicate the separations between the O atom of the 4NP group and the nearest residue.

positions in the active sites of both enzymes and formed networks of hydrogen bonds with the surrounding residues (Figure 6). The 4NP aglycon of 4NP-O-Glc in 1S_3 was in a favorable orientation in the aromatic/hydrophobic funnel, although it did not stretch through its entire length (Figure 6A, right panels). The distances between the anomeric carbons of 4NP-O-Glc and one of the oxygen atoms of the catalytic nucleophiles (EA) Glu386 and Glu389 were 3.2 and 2.9 Å in Os3BGlu7 and HvBII, respectively. The predicted C1–O_{EA} distances in the complexes of Os3BGlu7 and HvBII with 4NP-O-Glc in the 1S_3 conformer (black dashes in Figure 6A,B, left panels) suggested that the anomeric carbons were positioned optimally for nucleophilic attacks by the oxygen atoms of catalytic nucleophiles. Further, the glycosidic oxygens in 1S_3 were in the reasonable proximities of the catalytic acid/base residues (EB) Glu176 and Glu179, with O1–O_{EB} distances of 3.6 and 3.4 Å in Os3BGlu7 and HvBII, respectively. In addition, the hydrogen atoms of the C2–OH groups pointed to the oxygen atoms of the catalytic acid/base residues. The separations between the protons and the oxygen atoms of EB (HO2–O_{EB}) were 1.6 and 3.2 Å for Os3BGlu7 and HvBII, respectively. However, in the case of HvBII, the hydrogen atom was rotated toward Glu389, forming a strong hydrogen bond that was shorter than the HO2–O_{EB} distance of 3.2 Å. These rather short distances suggested that exceptionally strong hydrogen bonds were formed between the participating atoms that may potentially stabilize the enzyme–substrate complex and contribute to the higher affinity of the 1S_3 conformer of 4NP-O-Glc to both enzymes, as predicted by GLIDE (Table 2). It was also clear from Table 2 that the affinities of 4NP-O-Glc in the 3S_5 and 4C_1 geometries were less favorable. Here, the distances between the catalytic residues and C1, O1, and

C2-OH were significantly longer (Table 2). It was remarkable to observe that the 4NP aglycons of 4NP-O-Glc in the 3S_5 and 4C_1 geometries stretched further through the pocket than that of the 1S_3 conformer; in particular, the 4NP aglycon of the 4C_1 conformer stretched all the way through the length of the HvBII active site (cf. Figure 6A,B, middle and right panels). While the conformational penalty of the 1S_3 conformer was by only 1 kcal/mol lower for Os3BGlu7 than for HvBII, the penalties of the 3S_5 and 4C_1 conformers of 4NP-O-Glc were clearly higher for HvBII, with energy differences of 8.2 and 7.6 kcal/mol, respectively (cf. ΔE values in Table 2, rows 3 and 4).

In Os3BGlu7, GLIDE predicted a different binding mode for 4NP-S-Glc in the 1S_3 geometry (Figure 7). In this instance, the predicted position of the pyranose ring of 4NP-S-Glc overlapped well with that of 2-deoxy-2-fluoroglucoside, observed in the crystal structure of Os3BGlu7 (16). However, upon comparison of the positions of the pyranose rings of 4NP-S-Glc with that of 4NP-O-Glc in the 1S_3 geometries in both enzymes, the pyranose rings of the S-linked 4NP-glycoside were slightly "out of alignment" in the active sites (Figures 6 and 7, right panels). Moreover, it could be observed that unlike the 1S_3 geometry of 4NP-O-Glc, the 1S_3 conformer of 4NP-S-Glc in Os3BGlu7 adopted a more "equatorial-like" orientation for the glycosidic linkage with respect to the pyranose ring. However, in HvBII, the 3S_5 and 4C_1 geometries of 4NP-S-Glc had the glycosidic linkages in equatorial orientations with respect to the pyranose rings, and these linkage orientations seemed to be optimally predisposed for hydrolysis (Figure 7). Here, the distances between C1 of 4NP-S-Glc in 1S_3 and O atoms of the catalytic nucleophiles stretched to 4.7 Å in both enzymes, and the distances between S1 of 4NP-S-Glc and O atoms of catalytic acid/base residues were 4.3 and 4.0 Å for Os3BGlu7 and HvBII, respectively. Further, the distances between the C2-OH groups of 4NP-S-Glc and the O atoms of catalytic acid/base residues were 4.4 and 3.1 Å in Os3BGlu7 and HvBII, respectively (Table 2). In summary, the distances between the C1 and S1 atoms of 4NP-S-Glc in 1S_3 (2B) conformer and the catalytic residues in both enzymes were significantly longer than those between the C1 and O1 atoms of 4NP-O-Glc in the 1S_3 (1B) conformer (Table 2).

Clear differences in the binding of 4NP-S-Glc in the 3S_5 geometry were observed between Os3BGlu7 and HvBII (Table 2 and Figure 7, middle panels). In Os3BGlu7, the participating respective distances of 4.6 and 3.4 Å between the EA and EB catalysts and the C1 atoms of 4NP-S-Glc in 3S_5 were slightly shorter than those for the 1S_3 conformer. Further, the pyranose ring of 4NP-S-Glc in the 3S_5 conformer adopted a "flipped-down" orientation compared to that of 4NP-O-Glc in 1S_3 , as demonstrated by the HO2–O_{EB} distance of 7.1 Å (Table 2). Notably, this conformation posed steric limitations for the interactions between the C1 atom of 4NP-S-Glc in 3S_5 and the O1 atom of Glu386 in Os3BGlu7. On the other hand, the C1 atom in the HvBII complex with 4NP-S-Glc in the 3S_5 geometry showed a better associative power with EA and EB than 4NP-S-Glc in the 1S_3 geometry. Here, the corresponding C1–O_{EA}, S1–O_{EB}, and HO2–O_{EB} distances were 3.5, 3.4, and 1.5 Å, respectively (Table 2). However, the large conformational penalty of 34 kcal/mol for the HvBII complex with 4NP-S-Glc in the 3S_5 geometry most likely contributed to a low GLIDE score of –4.9 kcal/mol for this docking position (Table 2). Finally, binding of 4NP-S-Glc in the 4C_1 geometry to both enzymes resulted in scores similar to those that were calculated for 4NP-S-Glc in the 1S_3 conformer, although these conformations again adopted the flipped-down pyranose ring

orientations. The latter orientations produced longer separations between the C1 atom of 4NP-S-Glc and the O atoms of the EA and EB catalysts and also generated steric hindrances between C1 of 4NP-S-Glc and EA in both enzymes (Figure 7, right panels).

The binding modes of the 4C_1 conformers of 4NP-O-Man and 4NP-S-Man in Os3BGlu7 and HvBII that were predicted by GLIDE are illustrated in Figure 8. The stereoviews of binding modes revealed that the two *manno*-configured O- and S-linked 4NP-glycosides adopted very different positions in both enzymes, although the 4NP aglycon moieties of 4NP-O-Man and 4NP-S-Man overlapped reasonably well. As illustrated in Figure 8B, the binding modes of *manno*-configured O- and S-linked 4NP-glycosides in HvBII were almost identical, while in Os3BGlu7, the 4NP aglycons adopted different orientations (Figure 8A). Compared to the 1S_3 mannose geometry, the pyranose rings of 4NP-O-Man in the 4C_1 geometry in both enzymes adopted flipped-down orientations, where the C1–O_{EA}, O1–O_{EB}, and HO2–O_{EB} distances were 4.4 and 4.9 Å, 3.5 and 3.9 Å, and 4.3 and 2.5 Å in Os3BGlu7 and HvBII, respectively. The conformational penalty of 4NP-O-Man in the 4C_1 conformer favored the HvBII complex over that of Os3BGlu7, as illustrated by the energy difference of nearly 4 kcal/mol (Table 2). On the other hand, the 4C_1 conformer of 4NP-S-Man scored slightly better with both enzymes, although the binding modes were dissimilar. In Os3BGlu7, 4NP-S-Man in the 4C_1 conformer adopted an orientation that was similar to that of 4NP-O-Glc in the 1S_3 conformer, and where the C1 and S1 atoms interacted weakly with EA and EB, as denoted by the C1–O_{EA}, S1–O_{EB}, and HO2–O_{EB} distances of 4.4, 5.9, and 5.0 Å, respectively (Table 2). However, in HvBII, although the C1 atom of 4NP-S-Man in 4C_1 interacted weakly with EA at a C1–O_{EA} distance of 4.7 Å, the S1 atom and the C2-OH group of 4NP-S-Man formed strong interactions with EB with S1–O_{EB} and HO2–O_{EB} distances of 3.1 and 1.7 Å, respectively. Hence, in the HvBII complexes with 4NP-O-Man and 4NP-S-Man in the 4C_1 geometry, the axial hydroxyl groups in the C2 positions and the oxygens of EB formed strong interactions, as indicated by their respective separations of 1.7 and 2.5 Å (Table 2 and Figure 8). Noteworthy is the fact that the conformational penalty with 4NP-S-Man in 4C_1 was similar for both Os3BGlu7 and HvBII, regardless of the separations between the axial hydroxyl groups in the C2 positions and the EB oxygens (Table 2).

DISCUSSION

A detailed knowledge of binding of β -D-glycosides has paramount importance for understanding of hydrolytic processes. Naturally, the deduction of molecular features of glycosidase–substrate complexes is a complex task as substrates are transformed into products at fast rates. To dissect how binding of β -D-glycosides proceeds at atomistic levels and in particular what specific sugar geometries participate during hydrolysis, mechanism-based inhibitors and substrate analogues have been used to mimic natural glycoside substrates (46). In other approaches, variant enzymes containing mutated active site residues have been employed (e.g., ref 47). The information gained from these studies is critical for glycosidase inhibitor design that has applications in a wide range of biotechnologies, including food and dietary fiber processing and production of pharmaceuticals, nutraceuticals, paper, pulp, wood, and biofuels.

There is mounting evidence that enzymes recognize substrates with a preferential conformational selection mechanism and that

in the absence of substrates, a dynamic equilibrium exists between substrate-free and substrate-bound configurations (48). It has been suggested that the most realistic information for substrate binding events is obtained with probes resembling putative transition states (46, 49). Here, binding of transition-state probes is linked to energy gains, resulting from binding of high-energy conformers on a transition-state itinerary of hydrolytic enzymes, where oxocarbenium ion-like-shaped intermediates have been suggested to play key roles (11). These events have been investigated predominantly by X-ray diffraction (e.g., refs 15 and (5052)), although STD NMR techniques have contributed significantly (27, 28, 53).

Against this background, the goal of our study was to shed light on the structural basis of binding of S- and O-linked *gluco*- and *manno*-configured aryl- β -D-glycosides to the Os3BGlu7 and HvBII enzymes that preferentially hydrolyze β -D-glucoside and β -D-mannoside substrates, respectively (3–7). Our aim was also to understand how chemical and structural features in substrates and enzymes' active sites underlie the substrate specificity of these enzymes that are classified in the GH-A clan of CAZY classification (1).

We first determined and compared kinetic parameters of inhibition (K_i) and hydrolysis (K_m) with S- and O-linked *gluco*- and *manno*-configured aryl- β -D-glycosides (Table 1). The comparisons of the kinetic data of HvBII and Os3BGlu7, using the corresponding O-glycoside substrates, revealed that the observed trends in K_i values of the thio analogues did not necessarily follow those in the K_m values. For example, with Os3BGlu7, where both thio inhibitors were almost equally effective, the K_m values for 4NP-O-Glc and 4NP-O-Man differed by a factor of ~ 6 . This conclusion was also valid for HvBII, where, while 4NP-S-Glc was nearly 3-fold more potent than 4NP-S-Man, the K_m value for the O-linked glucoside was 2 times higher than that for 4NP-O-Man (Table 1). Overall, the inhibition experiments clearly demonstrated that HvBII was far more sensitive to the inhibition by 4NP-S-Glc and 4NP-S-Man than Os3BGlu7 and that the K_m and K_i values with O- and S-linked *gluco*- and *manno*-configured aryl- β -D-glycosides could not be correlated.

To reconcile the differences in binding of the O- and S-linked aryl glycosides that serve as respective substrates and inhibitors in Os3BGlu7 and HvBII, STD NMR and trNOESY investigations coupled with QM/MM modeling and docking experiments were undertaken. Here, we aimed to investigate precise conformations of S-linked *gluco*- and *manno*-configured aryl- β -D-glycosides in bound states, and if the distortions from the 4C_1 ground states could be observed by STD NMR and trNOESY spectroscopy. These studies were complemented with predictive molecular modeling approaches. We surmised that these two approaches could provide useful insights, in particular, when crystal structures of inhibitor–ligand complexes are not available. The former technique has become a powerful tool in recent years for the characterization of binding of mechanistic probes to proteins (e.g., refs 27 and 28) and relies on saturation transfer from protein to ligand molecules. Thus, when a protein is selectively irradiated, the ligand between bound and free forms also becomes saturated in a bound state. Subtraction of the spectra acquired with off-resonance irradiation readily reveals conformations of a bound ligand, and the differential STD effects within a ligand molecule provide the information about the proximity of individual protons present in the vicinity of a protein. The advantage of this technique is that it is rapid and does not require isotope labeling of proteins or excessively large quantities of protein.

The STD NMR and trNOESY experiments were set up in several stages. As previously shown (28), when ligands bind weakly and when exchange rates between free and bound states occur at reasonably fast rates, trNOESY signals could provide adequate means for determining conformations of glycosides. After the S-linked *gluco*- and *manno*-configured aryl- β -D-glycosides were incubated with HvBII, strong negative NOE cross-peaks were observed, indicating binding of inhibitors to the enzymes, in contrast to the free states of inhibitors. Via the trNOESY experiments with HvBII, we could unambiguously deduce conformational changes of the six-member rings of 4NP-S-Glc and 4NP-S-Man. Upon comparison of the experimental interproton distances of bound 4NP-S-Glc with those measured in the MM3*-minimized geometries, we concluded that 4NP-S-Glc adopted the 3S_5 or 1S_3 skew boat geometry, while 4NP-S-Man did not deviate from a starting low-energy 4C_1 geometry. The analogous STD NMR and trNOESY analyses for 4NP-S-Glc and 4NP-S-Man bound to Os3BGlu7 were also performed and showed that both thioglycosides remained in the ground 4C_1 conformations and that major saturation transfers were observed only on aromatic protons. The latter conclusions were supported by inhibition kinetics indicating that both β -D-thioglycosides were rather weak inhibitors (Table 1). The existence of no major contacts with the enzyme, in principle, suggests that no distortion of the ring is taking place upon binding. Further, the lack of observable NOESY cross-peaks for the sugar part during binding of 4NP-S-Glc and 4NP-S-Man to Os3BGlu7 may also suggest that future experiments with other inhibitors such as S-alkyl or S-benzyl derivatives are necessary and will ultimately clarify if the distortions of 4NP-S-glycosides upon binding to Os3BGlu7 occur.

The STD NMR and trNOESY experiments were complemented with QM/MM docking computations. During the docking procedure, the geometry of active site side chains was optimized, and this approach coupled with the QM-derived partial charges provided a basis for binding predictions by molecular docking. For these computations, we used a crystal structure of Os3BGlu7 (PDB entry 2RGM), while a homology model was constructed for HvBII, in the absence of its 3D structure. During the docking procedure, the orientations of the key amino acid residues in both proteins were sufficiently relaxed for the docked molecules to adjust. Although caution should be exercised in interpreting the computational data using homology models, as subtle errors in modeled HvBII could lead to erroneous miscalculations of docked positions and docking energies, in our experience the docking procedure using reliable homology models leads to reasonable predictions. Nevertheless, the use of the same models of, e.g., variant enzymes, where large conformational changes or variations in partial charges of amino acid residues occur, must be interpreted carefully because they could lead to erroneous prediction of binding affinities.

For docking of β -D-glycosides, we used as a guide an experimentally determined position of 2-deoxy-2-fluoroglucoside in Os3BGlu7 (16). When the characteristics of the active sites of Os3BGlu7 and HvBII were compared, it became obvious that they contained a plethora of charged and hydrophobic residues, where the catalytic nucleophiles Glu386 (Os3BGlu7) and Glu389 (HvBII) (4, 16) at subsite -1 were expected to form covalent linkages with the anomeric carbons of the substrates during formation of glycosyl–enzyme intermediates. The other important catalysts in the active sites of Os3BGlu7 and HvBII are the catalytic acid/base residues Glu176 (Os3BGlu7) and Glu179

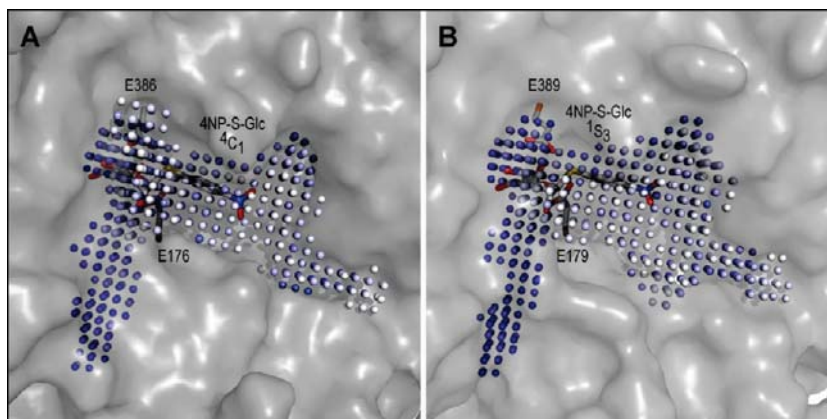


FIGURE 9: Shapes and buriedness of active site funnels of Os3BGlu7 and HvBII. The positions of 4NP-S-Glc (cpk) are shown in preferred geometries, 4C_1 for Os3BGlu7 (A) or 1S_3 for HvBII (B), as identified by STD NMR. The catalytic nucleophiles E386 (Os3BGlu7) and E389 (HvBII) and acid/base residues E176 (Os3BGlu7) and E179 (HvBII) are shown in cpk. Binding sites are given in the PocketPicker descriptors (38), with darker spheres indicating greater buriedness.

(HvBII) that, according to the canonical mechanism, should protonate glycosidic oxygens during hydrolysis. Moreover, in Os3BGlu7, the catalytic site is surrounded by a series of bulky hydrophobic/aromatic residues, namely, Trp433 and Trp441, and other residues such as Trp358 and Tyr315 that form a hydrophobic funnel (16). Here, a particular emphasis is placed on Trp433 positioned at the bottom of the active site pocket that mediates stacking interactions with the pyranose ring of 2-deoxy-2-fluoroglucoside (16).

One would expect that these interactions might be one of the major forces that keep substrates in favorable orientations for a nucleophilic attack. Further, at subsite 1, residues Ile179, Leu183, Tyr315, and Trp358 formed a hydrophobic funnel that presumably could accommodate the aglycon moieties of substrates. Thus, it would be expected that a specific and characteristic shape, buriedness, and geometry of active sites are required for particular conformations of substrates to bind in the active site regions. Here, a comparison of shape properties, using Pocket-Picker descriptors (38), indicated that both enzymes had very similar shapes, depths, and geometries of active site funnels (Figure 9). One would also expect that significant differences in conformations and charges in O- and S-linked 4NP-Glc and 4NP-Man might have serious consequences for the shapes of electrostatic surfaces of glycosides that could influence binding modes of substrates in enzymes' active sites, in particular those in subsites -1 and 1 . These features are summarized in Figures 5 and 9 and in Table S3 of the Supporting Information, collectively indicating that O- and S-linked aryl- β -D-glycosides exhibit certain characteristic electrostatic potentials, and that these potentials are dependent on geometries of the individual pyranose rings. In particular, the data in Table S3 of the Supporting Information indicated that significant differences in the ESP charge distributions in the O- and S-linked glycosides occurred in the vicinity of their anomeric C1 atoms. It was therefore not unexpected that, after the glycosidic oxygens were replaced with sulfurs, the ESP charges on the C1 atoms changed from the positive values to the negative ones in all conformers (Table S3 of the Supporting Information). On the other hand, the overall ESP charges on the X1 and O4 atoms became less negative (Table S3 of the Supporting Information).

A few observations from the modeling experiments are worth analyzing. It seemed that in addition to characteristic positions of the S-linked *gluco*- and *manno*-configured aryl- β -D-glycoside

conformers in the active site funnels, defined by their separations from the respective active site residues, the overall sidedness of the pyranose ring orientations varied. The so-called flipped-down 4C_1 conformations of 4NP-S-Glc and 4NP-O-Man were adopted in Os3BGlu7 and HvBII (Figures 6 and 8), and a similar orientation was also seen in the 3S_5 conformer of 4NP-S-Glc that bound to Os3BGlu7 (Figure 7A, middle panel), compared to a 2-deoxy-2-fluoroglucoside moiety in the covalent complex of Os3BGlu7 (16). These flipped-down conformations resulted in unusually long C2-OH–O_{EB} separations of more than 7 Å in Os3BGlu7 (Table 2), except for that of 4NP-O-Man, and were also associated with longer separations between β -D-glycosides and respective active site residues EA and EB. This observation indicated that both enzymes could select incorrect orientations of incoming substrates that bind less favorably in their catalytic funnels. The importance of hydrogen bonds between the C2-OH group and catalytic residues was emphasized in a theoretical study of the catalytic mechanisms of β -D-glycoside hydrolases, where Bras and co-workers (52) concluded that the C2-OH–O_{EA} hydrogen bond is responsible for a low-energy activation barrier in the *E. coli* β -galactosidase catalytic mechanism.

Further, scoring algorithms that evaluate the accuracy of docking (Figures 6–8 and Table 2) and that could distinguish between various binding modes of ligands deserve mentioning, although the accuracy of predicting precise affinities in quantitative terms could be relatively low (44, 54). Nevertheless, the data from our modeling experiments clearly demonstrated that binding behavior of 4NP-S-Glc and 4NP-S-Man differed between the two plant enzymes. It was noteworthy that the bound O- and S-linked *gluco*- and *manno*-configured aryl- β -D-glycosides did not adopt necessarily their lowest-energy conformations in bound states (Table 2). To estimate the strain energy imposed by the active site environments, we calculated the energy differences of bound aryl- β -D-glycosides at the M05-2X/6-31+G* level of theory. These energy differences arose from the differences between their local minima in free and bound states (Table 2). From these values, one could infer how binding of O- and S-linked *gluco*- and *manno*-configured aryl- β -D-glycosides in the higher-energy states influences the hydrolytic activation barrier.

Finally, it has been suggested for the members of the GH26 group of β -D-mannanases and (1,3;1,4)- β -D-glucan endohydrolases, which like the β -D-glucosidases and β -D-mannosidases

studied here are the members of clan GH-A, that while the interactions around the C1 and C2-OH groups were similar, environments around C3-OH groups could contribute to the specificity of the reaction coordinate (55). The authors further suggested that although the chemistry of the bound sugar is important, it does not dominate substrate specificity; whereas for *manno*-configured substrates, both $B_{2,5}$ and 3H_4 transition states participate, for *gluco*-configured substrates, the 4H_3 and $B^{2,5}$ transition states participate (55). Subsequently, a Michaelis complex of a GH2 β -D-mannosidase BtMan2A was published, which suggested that a boatlike transition state develops on a 1S_5 - $B_{2,5}$ - 0S_2 itinerary in this class of hydrolases (56). However, in the work described here, we concluded that other conformations, such as those participating in the 1S_5 - $B_{2,5}$ - 0S_2 itinerary for BtMan2A, are highly unlikely.

Our computational data indeed showed that the so-called flipped-down 4C_1 conformation of 4NP-O-Man was adopted in Os3BGlu7 and HvBII. On the other hand, the QM/MM modeling experiments indicated that the most favorable geometry for 4NP-O-Man with both Os3BGlu7 and HvBII was 1S_3 (Table 2). Here we did not test the significance of 1S_5 geometry, although we presume that geometrical interconversions between 1S_5 and 1S_3 should be less significant than those between 3S_5 and 1S_3 , and thus, the 1S_3 conformer could be considered as an intermediate between the 3S_5 ($B_{3,O}$) and 1S_5 conformers. The 1S_3 geometry was also identified in 4NP-O-Glc with both β -D-glycosidases through the QM modeling and docking experiments, although STD NMR and trNOESY experiments confirmed its existence only with 4NP-S-Glc. To complicate the matters even further, it has recently been reported with a β -D-glucosidase from a soil metagenome library (57) that in addition to the glycon, the nature of the aglycon could contribute to substrate specificity. Similar observations were previously reported with a maize β -glucosidase (20) and diverse β -D-mannanases (58).

In summary, STD NMR and trNOESY experiments in conjunction with molecular docking simulations provided information about binding of S- and O-linked *gluco*- and *manno*-configured aryl- β -D-glycosides to rice Os3BGlu7 and barley HvBII that operate predominantly as a β -D-glucosidase and β -D-mannosidase, respectively. Kinetic analyses with 4NP-S-Glc and 4NP-S-Man indicated that the inhibitions were competitive. The STD NMR and trNOESY experiments revealed that 4NP-S-Glc and 4NP-S-Man bound weakly in 4C_1 conformations to Os3BGlu7, 4NP-S-Glc adopted the 3S_5 or 1S_3 conformation, and 4NP-S-Man preferred 4C_1 geometry with HvBII. The docking and modeling studies predicted that 4NP-O-Glc, 4NP-O-Man, and 4NP-S-Man bound preferentially in the 1S_3 geometries to both enzymes, contrary to 4NP-S-Glc that could adopt a range of geometries (1S_3 , 3S_5 , or 4C_1). Although NMR, kinetic, and QM/MM examinations presented in this study provide valuable information about the conformations of sugars bound by the two plant β -D-glycosidases with distinct substrate specificities, we further surmise that the depth of information could be further enhanced via X-ray crystallography, and thus the amalgamation of the four techniques would ideally represent the most powerful approach for evaluation of binding affinities (60).

We conclude that in the plant β -D-glucoside hydrolase Os3BGlu7 and β -D-mannoside hydrolase HvBII a combination of determinants is likely to play roles in substrate recognition, such as (i) the inherent conformational and spatial flexibilities of

gluco- and *manno*-configured substrates in the enzymes' active sites, (ii) the subtle spatial differences in the disposition of active site residues and their capacities to form interactions with specific groups of substrates, and (iii) the small variations in charge distributions and shapes of the catalytic sites. We anticipate that the combination of these determinants may collectively drive the key interactions between the active site residues and *gluco*- and *manno*-configured substrates and underlie the enzymes' substrate specificities.

ACKNOWLEDGMENT

J.C., L.C., and J.J.-B. thank J. J. Hernández-Gay for helpful discussions.

SUPPORTING INFORMATION AVAILABLE

Supplementary Tables 1–3. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES

- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The carbohydrate-active enzymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res.* 37, D233–D238.
- Hrmova, M., and Fincher, G. B. (2009) Depolymerizing enzymes. In *Chemistry, Biochemistry and Biology of (1,3)- β -D-Glucans and Related Polysaccharides* (Bacic, T., Fincher, G. B., and Stone, B. A., Eds.) pp 119–170, Academic Press and Elsevier Inc., San Diego.
- Hrmova, M., Harvey, A. J., Wang, J., Shirley, N. J., Jones, G. P., Stone, B. A., Høj, P. B., and Fincher, G. B. (1996) Barley β -D-glucan exohydrolases with β -D-glucosidase activity. Purification and determination of primary structure from a cDNA clone. *J. Biol. Chem.* 271, 5277–5286.
- Hrmova, M., MacGregor, E. A., Biely, P., Stewart, R. J., and Fincher, G. B. (1998) Substrate binding and catalytic mechanism of a barley β -D-glucosidase/(1,4)- β -D-glucan exohydrolase. *J. Biol. Chem.* 273, 11134–11143.
- Hrmova, M., Burton, R. A., Biely, P., Lahnstein, J., and Fincher, G. B. (2006) Hydrolysis of (1,4)- β -D-mannans in barley (*Hordeum vulgare* L.) is mediated by a concerted action of a (1,4)- β -D-mannan endohydrolase and a β -D-mannosidase. *Biochem. J.* 399, 77–90.
- Kuntthom, T., Luang, S., Fincher, G. B., Opasiri, R., Hrmova, M., and Ketudat Cairns, J. R. (2008) Closely related barley and rice family GH1 glycosyl hydrolases with β -D-glucosidase and β -D-mannosidase activities. *Arch. Biochem. Biophys.* 491, 85–95.
- Opasiri, R., Hua, Y., Wara-aswapati, O., Akiyama, T., Svasti, J., Esen, A., and Ketudat Cairns, J. R. (2004) β -Glucosidase, exo- β -glucanase and pyridoxine transglucosylase activities of rice BGlu1. *Biochem. J.* 379, 125–131.
- Marana, S. R. (2006) Molecular basis of substrate specificity in family 1 glycoside hydrolases. *IUMB Life* 58, 6–73.
- Vocadlo, D. J., and Davies, G. J. (2008) Mechanistic insights into glycosidase chemistry. *Curr. Opin. Chem. Biol.* 12, 539–555.
- 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.
- Zechel, D. L., and Withers, S. G. (2001) Dissection of nucleophilic and acid-base catalysis in glycosidases. *Curr. Opin. Chem. Biol.* 5, 643–649.
- Vallmitjana, M., Ferrer-Navarro, M., Planell, R., Abel, M., Querol, E., Planas, A., and Pérez-Pons, J.-A. (2001) Mechanism of the family β -glucosidase from *Streptomyces* sp.: Catalytic residues and kinetic studies. *Biochemistry* 40, 5975–5982.
- Ly, H. D., and Withers, S. G. (1999) Mutagenesis of glycosidases. *Annu. Rev. Biochem.* 68, 487–522.
- Cicek, M., Blanchard, D., Bevan, D. R., and Esen, A. (2000) The aglycone specificity-determining sites are different in 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA)-glucosidase (maize β -glucosidase) and dhurrinase (sorghum β -glucosidase). *J. Biol. Chem.* 275, 20002–20011.

15. Czjzek, M., Cicek, M., Zamboni, V., Bevan, D. R., Henrissat, B., and Esen, A. (2000) The mechanism of substrate (aglycone) specificity in β -glucosidases is revealed by crystal structures of mutant maize β -glucosidase-DIMBOA, -DIMBOAGlc, and -dhurrin complexes. *Proc. Natl. Acad. Sci. U.S.A.* 97, 13555–13560.
16. Chuenchor, W., Pengthaisong, S., Robinson, R. C., Jirundon, Y., Esen, A., Chen, C.-J., Opassiri, R., Svasti, J., and Ketudat Cairns, J. R. (2008) Structural insights into rice BGLu1 β -glucosidase oligosaccharide hydrolysis and transglycosylation. *J. Mol. Biol.* 377, 1200–1215.
17. Street, I. P., Kempton, J. B., and Withers, S. G. (1992) Inactivation of a β -glucosidase through the accumulation of a stable 2-deoxy-2-fluoro- β -D-glucopyranosyl-enzyme intermediate: A detailed investigation. *Biochemistry* 31, 9970–9978.
18. Bause, E., and Legler, G. (1980) Isolation and structure of a tryptic glycopeptide from the active site of β -glucosidase A₃ from *Aspergillus wentii*. *Biochim. Biophys. Acta* 626, 459–465.
19. 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 archaeon *Sulfolobus solfataricus*: Resilience as a key factor in thermostability. *J. Mol. Biol.* 271, 789–802.
20. Verdoucq, L., Moriniere, J., Bevan, D. R., Esen, A., Vasella, A., Henrissat, B., and 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.
21. Sulzenbacher, G., Driguez, H., Henrissat, B., Schülein, M., and Davies, G. J. (1996) Structure of the *Fusarium oxysporum* endoglucanase I with a nonhydrolyzable substrate analogue: Substrate distortion gives rise to the preferred axial orientation for the leaving group. *Biochemistry* 35, 15280–15287.
22. Hrmova, M., De Gori, R., Smith, B. J., Fairweather, J. K., Driguez, H., Varghese, J. N., and Fincher, G. B. (2002) Structural basis for a broad specificity in higher plant β -D-glucan glucohydrolases. *Plant Cell* 14, 1033–1052.
23. Money, V., Cartmell, A., Guerreiro, C. I. P. D., Ducros, V. M.-A., Fontes, C. M. G. A., Gilbert, H. J., and Davies, G. J. (2008) Probing the β -1,3:1,4 glucanase, Ctlc26A, with a thio-oligosaccharide and enzyme variants. *Org. Biomol. Chem.* 6, 851–853.
24. Opassiri, R., Ketudat Cairns, J. R., Akiyama, T., Wara-aswapati, O., Svasti, J., and Esen, A. (2003) Characterization of a rice β -glucosidase highly expressed in flower and germinating shoot. *Plant Sci.* 165, 627–638.
25. Blanc-Muesser, M., Defaye, J., and Driguez, H. (1978) Syntheses stereoselectives de 1-thiogluco-sides. *Carbohydr. Res.* 67, 305–328.
26. Chuenchor, W., Pengthaisong, S., Yuvaniyama, J., Opassiri, R., Svasti, J., and Ketudat Cairns, J. R. (2006) Purification, crystallization and preliminary X-ray analysis of rice BGLu1 β -glucosidase with and without 2-deoxy-2-fluoro- β -D-glucoside inhibitor. *Acta Crystallogr. F* 62, 798–801.
27. Vogther, M., and Peters, T. (2000) Application of NMR based binding assays to identify key hydroxyl groups for intermolecular recognition. *J. Am. Chem. Soc.* 122, 6093–6099.
28. Clavel, C., Canales, A., Gupta, G., Cañada, F. J., Penadés, S., Suriola, A., and Jiménez-Barbero, J. (2007) NMR investigation of the bound conformation of natural and synthetic oligomannosides to banana lectin. *Eur. J. Org. Chem.* 10, 1577–1585.
29. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 24, 4876–4882.
30. Kim, D. E., Chivian, D., and Baker, D. (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* 32 (Suppl. 2), W256–W261.
31. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815.
32. 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.
33. Vriend, G. (1990) WHAT IF: A molecular modeling and drug design program. *J. Mol. Graphics* 8, 52–56.
34. Lüthy, R., Bowie, J. E., and Eisenberg, D. (1992) Assessment of protein models with three-dimensional profiles. *Nature* 356, 83–85.
35. Sippl, M. J. (1993) Recognition of errors in 3-dimensional structures of proteins. *Proteins: Struct., Funct., Bioinf.* 17, 355–362.
36. Sumathi, K., Ananthalakshmi, P., Md Roshan, M. B. A., and Sekar, K. (2006) 3dSS: 3D structural superposition. *Nucleic Acids Res.* 34, W128–W134.
37. DeLano, W. (2009) The PyMOL Molecular Graphics System, DeLano Scientific LLC, San Carlos, CA.
38. Weisel, M., Proschak, E., and Schneider, G. (2007) PocketPicker: Analysis of ligand-binding sites with shape descriptors. *Chem. Cent. J.* 1, 7.
39. MacroModel, version 9.6 (2008) Schrödinger, LLC, New York.
40. Zhao, Y., Schultz, N. E., and Truhlar, D. G. (2005) Exchange-correlation functional with broad accuracy for metallic and nonmetallic compounds, kinetics, and noncovalent interactions. *J. Chem. Phys.* 123, 161103.
41. Zhao, Y., Schultz, N. E., and Truhlar, D. G. (2006) Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. *J. Chem. Theory Comput.* 2, 364–382.
42. Zheng, J., Zhao, Y., and Truhlar, D. G. (2007) Representative benchmark suites for barrier heights of diverse reaction types and assessment of electronic structure methods for thermochemical kinetics. *J. Chem. Theory Comput.* 3, 569–582.
43. Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 118, 11225–11236.
44. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 47, 1739–1749.
45. Tvaroška, I., and Carver, J. P. (1996) *Ab initio* molecular orbital calculation of carbohydrate model compounds. 5. Anomeric, exoanomic, and reverse anomeric effects in C-, N-, and S-glycosyl compounds. *J. Phys. Chem.* 100, 11305–11313.
46. Rempel, B. P., and Withers, S. G. (2008) Covalent inhibitors of glycosidases and their applications in biochemistry and biology. *Glycobiology* 18, 570–586.
47. Blanchard, J. E., and Withers, S. G. (2001) Rapid screening of the aglycone specificity of glycosidases: Applications to enzymatic synthesis of oligosaccharides. *Chem. Biol.* 8, 627–633.
48. Davulcu, O., Flynn, P. F., Chapman, M. S., and Skaliky, J. J. (2009) Intrinsic domain and loop dynamics commensurate with catalytic turnover in an induced-fit enzyme. *Structure* 17, 1356–1367.
49. Ermert, P., Rupitz, K., Vasella, A., Weber, M., and Withers, S. G. (1993) Transition state analogue inhibitors of glycosidases are configurationally selective: A study with nojiritetrazoles, a new class of glycosidase inhibitors. *Carbohydr. Res.* 250, 113–128.
50. Hrmova, M., Streltsov, V. A., Smith, B. J., Vasella, A., Varghese, J. N., and Fincher, G. B. (2005) Structural rationale for low nanomolar binding of transition state mimics to a family GH3 β -D-glucan glucohydrolase from barley. *Biochemistry* 44, 16529–16539.
51. Gloster, T. M., Meloncelli, P., Stick, R. V., Zechel, D., Vasella, A., and Davies, G. J. (2007) Glycosidase inhibition: An assessment of the binding of 18 putative transition-state mimics. *J. Am. Chem. Soc.* 129, 2345–2354.
52. Bras, N. F., Moura-Tameres, A., Fernandes, P. A., and Ramos, M. J. (2008) Mechanistic studies on the formation of glycosidase-substrate and glycosidase-inhibitor covalent intermediates. *J. Comput. Chem.* 29, 2565–2574.
53. Martin-Pastor, M., Vega-Vázquez, M., De Capua, A., Canales, A., Andre, S., Gabius, H. J., and Jiménez-Barbero, J. (2006) Enhanced signal dispersion in saturation transfer difference experiments by conversion to a 1D-STD-homodecoupled spectrum. *J. Biomol. NMR* 36, 103–109.
54. Englebienne, P., Fiaux, H., Kuntz, D. A., Corbeil, C. R., Gerber-Lemaire, S., Rose, D. R., and Moitessier, N. (2007) Evaluation of docking programs for predicting binding of Golgi α -mannosidase II inhibitors: A comparison with crystallography. *Proteins* 69, 160–176.
55. Money, V. A., Smith, N. L., Scaffidi, A., Stick, R. V., Gilbert, H. J., and Davies, G. J. (2006) Substrate distortion by a lichenase highlights the different conformational itineraries harnessed by related glycoside hydrolases. *Angew. Chem., Int. Ed.* 45, 5136–5140.

56. Offen, W. A., Zechel, D. L., Withers, S. G., Gilbert, H. J., and Davies, G. J. (2009) Structure of the Michaelis complex of β -mannosidase, Man2A, provides insight into the conformational itinerary of mannoside hydrolysis. *Chem. Commun.*, 2484–2486.
57. Nam, K. H., Sung, M. W., and Hwang, K. Y. (2010) Structural insights into the substrate specificity of β -glucosidase. *Biochem. Biophys. Res. Commun.* 391, 1131–1135.
58. Tailford, L. E., Ducros, V. M., Flint, J. E., Roberts, S. M., Morland, C., Zechel, D. L., Smith, N., Bjørnvad, M. E., Borchert, T. V., Wilson, K. S., Davies, G. J., and Gilbert, H. J. (2009) Understanding how diverse β -mannanases recognize heterogeneous substrates. *Biochemistry* 48, 7009–7018.
59. Fersht, A. (1999) Structure and Mechanism in Protein Science, pp 1–631, W. H. Freeman and Co., New York.
60. Hrmova, M., and Fincher, G. B. (2009) Functional genomics and structural biology in the definition of gene function. In *Methods in Molecular Biology: Plant Genomics* (Somers, D., Langridge, P., and Gustafson, P., Eds.) Vol. 513, pp 175–198, Humana Press Inc., Totowa, NJ.



Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

The structural basis of oligosaccharide binding by rice BGlu1 beta-glucosidase

Watchalee Chuenchor^{a,1}, Salila Pengthaisong^a, Robert C. Robinson^b, Jirundon Yuvaniyama^c, Jisnusun Svasti^c, James R. Ketudat Cairns^{a,*}

^aSchools of Biochemistry and Chemistry, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand

^bInstitute of Molecular and Cell Biology, 61 Biopolis Dr., Proteos, Singapore 138673, Singapore

^cDepartment of Biochemistry and Center of Excellence in Protein Structure and Function, Faculty of Science, Mahidol University, Bangkok 10400, Thailand

ARTICLE INFO

Article history:

Received 22 June 2010

Received in revised form 27 August 2010

Accepted 22 September 2010

Available online xxxx

Keywords:

Glycoside hydrolase

Acid–base mutant

Cello-oligosaccharides

X-ray crystallography

Enzyme–substrate complex

ABSTRACT

Rice BGlu1 β -glucosidase is an oligosaccharide exoglucosidase that binds to six β -(1 $\rightarrow$ 4)-linked glucosyl residues in its active site cleft. Here, we demonstrate that a BGlu1 E176Q active site mutant can be effectively rescued by small nucleophiles, such as acetate, azide and ascorbate, for hydrolysis of aryl glycosides in a pH-independent manner above pH 5, consistent with the role of E176 as the catalytic acid–base. Cellotriose, cellotetraose, cellopentaose, cellohexaose and laminaribiose are not hydrolyzed by the mutant and instead exhibit competitive inhibition. The structures of the BGlu1 E176Q, its complexes with cello-tetraose, cellopentaose and laminaribiose, and its covalent intermediate with 2-deoxy-2-fluoroglucoside were determined at 1.65, 1.95, 1.80, 2.80, and 1.90 Å resolution, respectively. The Q176 N ϵ was found to hydrogen bond to the glycosidic oxygen of the scissile bond, thereby explaining its high activity. The enzyme interacts with cellooligosaccharides through direct hydrogen bonds to the nonreducing terminal glucosyl residue. However, interaction with the other glucosyl residues is predominantly mediated through water molecules, with the exception of a direct hydrogen bond from N245 to glucosyl residue 3, consistent with the apparent high binding energy at this residue. Hydrophobic interactions with the aromatic sidechain of W358 appear to orient glucosyl residues 2 and 3, while Y341 orients glucosyl residues 4 and 5. In contrast, laminaribiose has its second glucosyl residue positioned to allow direct hydrogen bonding between its O2 and Q176 O ϵ and O1 and N245. These are the first GH1 glycoside hydrolase family structures to show oligosaccharide binding in the hydrolytic configuration.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

β -Glucosidases (E.C. 3.2.1.21, β -D-glucopyranosidases) hydrolyze β -glycosidic bonds to release nonreducing terminal D-glucopyranosyl residues from oligosaccharides and glycosides (Ketudat Cairns and Esen, 2010). As such, they have become much studied due to their potential use in biomass conversion of cellulosic waste, together with cellulases and exoglucanases, which break the long cellulose chains into shorter (1 $\rightarrow$ 4)- β -linked glucooligosaccharides

that are favored substrates for certain β -glucosidases (Sticklin, 2006). Microbial β -glucosidases, which have evolved to take advantage of the vast cellulosic energy pool of plant cell wall materials, are typically used for biomass conversion applications. However, plants have separately evolved their own β -glucosidases for (1 $\rightarrow$ 4)- β -linked oligosaccharides generated in the process of cell wall recycling within the plant. One of the most well-studied of these plant enzymes is rice (*Oryza sativa* L.) BGlu1, systematically named Os3BGlu7, which was identified as a highly expressed isoenzyme in rice seedlings and subsequently noted to be the most abundantly expressed glycoside hydrolase family GH1 (Henrissat, 1991) member in rice (Opassiri et al., 2003; Opassiri, 2006).

GH1 proteins, including rice BGlu1, are members of glycoside hydrolase clan A, which consists of several families of glycoside hydrolases with catalytic domains that have a (β/α)₈ barrel structure and two conserved carboxylic acid residues on β -strands 4 and 7 of the β -barrel, which serve as the catalytic acid–base and nucleophile, respectively (Fig. 1A) (Jenkins et al., 1995; Henrissat et al., 1995). These enzymes are thought to act through a double displacement mechanism (Fig. 1B), whereby the glycosidic oxygen is protonated by the catalytic acid–base and the catalytic

Abbreviations: Abg, *Agrobacterium* sp. β -glucosidase; dNPGlc, 2,4-dinitrophenyl β -D-glucopyranoside; G2F, 2-deoxy-2-fluoro-glucoside; GH1, glycoside hydrolase family 1; Glc, glucosyl residue; pNPGlc, para-nitrophenyl β -D-glucopyranoside.

* Corresponding author. Address: Institute of Science, Suranaree University of Technology, 111 University Avenue, Nakhon Ratchasima 30000, Thailand. Fax: +66 44 224185.

E-mail addresses: watchalee_lee@hotmail.com (W. Chuenchor), salila01@yahoo.com (S. Pengthaisong), rrobinson@icmb.a-star.edu.sg (R.C. Robinson), scjyv@mahidol.ac.th (J. Yuvaniyama), scjsv@mahidol.ac (J. Svasti), cairns@sut.ac.th (James R. Ketudat Cairns).

¹ Present Address: Dept. of Chemistry & Biochemistry, University of Maryland, College Park, MD 20742, USA

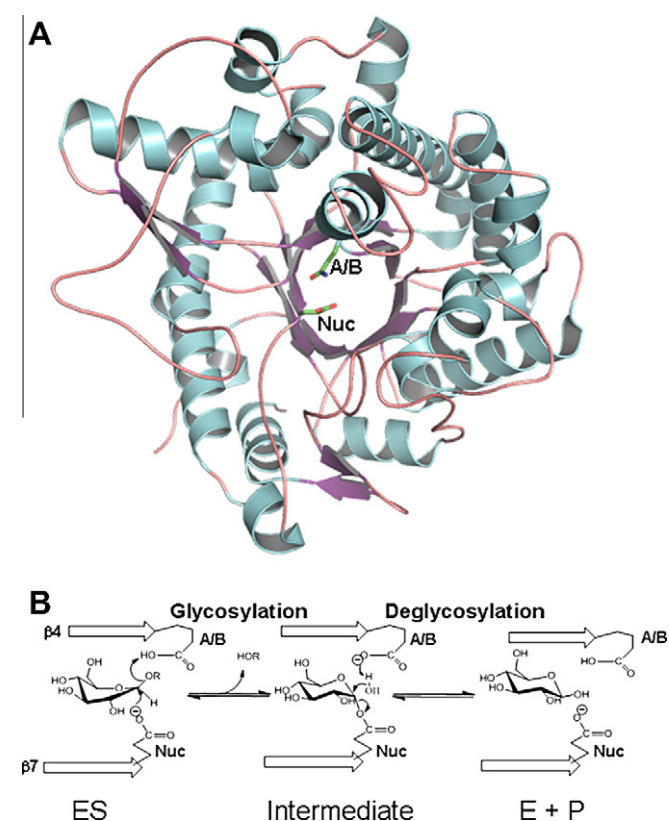


Fig. 1. Overall structure and double displacement mechanism proposed for glycosyl hydrolase family 1 β -D-glucosidases. (A). Cartoon diagram of the rice BGLu1 E176Q structural model (PDB entry 3F4V) with the catalytic acid–base (A/B) and nucleophile (Nuc) shown in stick representation. Note that in this acid/base mutant, the catalytic acid/base has been substituted with Gln, but it should normally be Glu. (B). In the double displacement mechanism, the β -glucoside substrate binds with the glycone in a distorted (1S_3 skew boat) conformation with the anomeric carbon between the acid–base glutamic acid (A/B) and the nucleophile glutamate (Nuc), located on beta-strands β_4 and β_7 of the (β/α)₈ barrel, respectively. Protonation of the leaving group allows the nucleophile to displace the aglycone from the glucosyl residue via an oxocarbenium ion-like transition state to form a covalent intermediate in the glycosylation step. Deglycosylation occurs when the intermediate is hydrolyzed by water, which is activated by extraction of a proton by the catalytic acid–base, essentially reversing the glycosylation step.

nucleophile attacks the anomeric carbon in the glycosylation step to form an α -linked covalent intermediate, followed by deglycosylation, in which a water or another nucleophile attacks at the anomeric carbon with basic assistance from the catalytic acid–base to displace the nucleophile and release the β -D-glucopyranose. Evidence for the formation of the covalent intermediate and the identity of the nucleophile was first provided by the trapping the GH1 *Agrobacterium* sp. β -glucosidase (Abg) linked to a 2-fluoroglucosyl residue, in which the electron withdrawal by the fluorine atom destabilizes the oxocarbenium cation-like transition state, greatly slowing both stages of the reaction (Withers et al., 1990). The use of a highly ionizable 2,4-dinitrophenolate as the leaving group of the glycoside substrate allowed the glycosylation to proceed relatively rapidly, while the deglycosylation half-reaction was slowed to trap the covalent intermediate. This covalent intermediate has since been observed in a number of GH1 enzyme crystal structures (Burmeister et al., 1997; Zechel et al., 2003; Gloster et al., 2004; Isorna et al., 2007; Chuenchor et al., 2008).

Mutagenesis to replace the catalytic nucleophile with small non-nucleophilic residues changed the GH1 mechanism to allow direct attack on β -linked substrates by small nucleophiles to form α -linked products or the use of α -fluoroglucoside as a surrogate for

the covalent intermediate in transglycosylation reactions (Mackenzie et al., 1998). Similarly, mutations of the catalytic acid–base validate its identity by providing catalysis when substrates with good leaving groups, like 2,4-dinitrophenolate, that do not require acid-assistance are used. These substrates allow the glycosylation step to proceed, and small anionic nucleophiles and sulfhydryls are provided to displace the enzyme from the glucosyl group in the deglycosylation step (Wang et al., 1995). The use of such acid–base mutants has allowed the trapping of unmodified glucosyl residues on the enzyme, as recently shown for the GH1 human cytoplasmic β -glucosidase E165Q mutant (Noguchi et al., 2008). Interestingly, a natural example of the rescue of an acid–base mutant occurs in the GH1 myrosinases (β -D-thioglucosidases), where the acid–base is replaced with glutamine and ascorbate seems to serve in the catalytic base function (Burmeister et al., 2000).

Recently, crystal structures have been determined for bacterial, fungal and plant GH1 enzymes that act on cellooligosaccharides (Isorna et al., 2007; Nijkken et al., 2007; Chuenchor et al., 2008). Rice BGLu1 has evolved separately from the microbial enzymes and is more closely related to plant enzymes that act on other substrates than to the microbial enzymes that act on cellooligosaccharides. X-ray crystal structures have been determined for a number of plant GH1 enzymes (Barrett et al., 1995; Burmeister et al., 1997; Czjzek et al., 2000, 2001; Verdoucq et al., 2004; Sue et al., 2006; Barleben et al., 2007; Seshadri et al., 2009), but none of the plant GH1 β -glucosidases with known structures is known to hydrolyze cellooligosaccharides efficiently, other than rice BGLu1. The aglycone specificity of the maize, sorghum and wheat β -glucosidases has been extensively studied by mutagenesis and structural studies of mutants with substrates or inhibitors (Czjzek et al., 2000, 2001; Zouhar et al., 2001; Verdoucq et al., 2003, 2004; Sue et al., 2006; Dopitova et al., 2008). These studies suggest that a conserved Trp, corresponding to rice BGLu1 W358, and several variable residues around the active site pocket are important for substrate aglycone binding. However, no general rule for how to identify the substrate specificity from the active site residues has emerged.

BGLu1, like a closely related β -mannosidase/ β -glucosidase from barley and its closely related rice isoenzymes Os3BGLu8 and Os7BGLu26, has been shown to hydrolyze cellooligosaccharides with up to 6 glucosyl residues in length with increasing efficiency, although the barley enzyme hydrolyzed cellobiose more rapidly than cellotriose (Hrmova et al., 1996; Hrmova et al., 1998; Opassiri et al., 2004; Kuntothom et al., 2009). Kinetic subsite mapping suggested that this difference was primarily due to the relative affinities at subsites +1 and +2 (which bind cellooligosaccharide glucosyl residues 2 and 3, respectively). Both enzymes hydrolyze laminaribiose with high efficiency, but show decreasing rates of hydrolysis of longer β -(1 $\rightarrow$ 3)-linked oligosaccharides as the degree of polymerization increases. The structures of BGLu1 and its covalent complex with 2-fluoroglucose showed a wide slot running out of the active site, which could act to bind oligosaccharides (Chuenchor et al., 2008). Mutagenesis of I179, N190 and N245, which line this slot affected the efficiency of hydrolysis of cellobiose and cellotriose, but computational docking of a cellotriose molecule into the active site was not precise enough to allow the interactions involved to be defined. To better understand the basis for hydrolysis of oligosaccharides, we have identified a catalytic acid–base mutant of rice BGLu1 (E176Q) that has relatively high activity for *p*-nitrophenyl- β -D-glucopyranoside (pNPGlc) substrate compared to other catalytic residue mutants, but cannot hydrolyze oligosaccharides. This mutant enzyme allowed binding of oligosaccharides to be studied by inhibition kinetics and X-ray crystallography.

2. Materials and methods

2.1. Mutagenesis and protein expression

The E176Q mutation of the rice *bglu1* cDNA was done by the overlap extension method (Wurch et al., 1998) with the T7 promoter (5'-TAATACGACTCACTATAGGG-3'), E176Qr (5'-CCTTGGCTGATTAATGTAAA-3'), E176Qf (5'-TTTAATCAGCCAAGGATAGTA-3'), and EcoRIr (5'-TTCGATGAATTCTCAGTGCTT-3') primers. The cDNA was then subcloned into pET32a(+) as previously described (Opassiri et al., 2003). All other mutations of the rice *bglu1* cDNA were made in the pET32/Bglu1 expression plasmid by the Quikchange method (Stratagene), as described by the supplier. The oligonucleotides for mutagenesis had the following sequences and their reverse complements, for E176A: 5'-CACTGGTTTACATTTAATGCGCAAGGATAGTAGCAG-3', for E176D: 5'-GCACTGGTTTACATTTAATGACCAAGGATAGTAGCA-5', for E386D: 5'-GACAGTCGTCATACTGACAACGGAATGGATCAAC-5', and for E386Q: 5'-CCGACAGTCGTCATACTCAGACGGAATGGATCAACCT-5'. All mutations were verified by sequencing the full cDNA to ensure that each had only the desired mutation. The proteins were expressed and purified as previously described (Opassiri et al., 2003; Chuenchor et al., 2006).

Enzyme rates were determined in triplicate reactions, as previously described (Chuenchor et al., 2008), with a time course to ensure linearity. The pH dependence of the wildtype Bglu1 and the E176 mutants was determined with *p*-nitrophenol β -D-glucopyranoside (pNPGlc) as a substrate in universal pH buffer (0.2 M boric acid, 0.05 M citric acid mixed with 0.05 M tri-sodium phosphate to achieve the pH values). Relative activities of wildtype Bglu1 and Bglu1 E176 mutants were determined in 50 mM buffer: sodium acetate, MES, phosphate or universal buffer, and 50 mM nucleophiles after adjustment of the reaction solution (without enzyme and substrate) to pH 5.0. Kinetic parameters (K_m and apparent k_{cat}) of wildtype Bglu1 and Bglu1 E176 mutants in the presence of 50 mM sodium azide were calculated from nonlinear regression of Michaelis–Menten curves with Grafit 5.0 (Leatherbarrow et al., 2001). The competitive inhibition constants (K_i) were determined from the plots of the apparent K_m/V_{max} vs. the inhibitor (oligosaccharide) concentration in the presence of 50 mM sodium azide, and were taken to be the apparent dissociation constants (K_D) of the inhibitors. The association constants were calculated as $K_A = 1/K_D$; and Gibbs free energies of binding were calculated as $\Delta G_{binding} = -RT \ln K_A$, where R is the gas constant (8.314 J K⁻¹ mol⁻¹) and T is the absolute temperature (303 K). Relative inhibitor potencies were calculated by setting that of cellobiose at 100%, dividing the association constants of the oligosaccharides by that of cellobiose and multiplying by 100%.

2.2. X-ray crystallography

Prior to crystallization trials, Bglu1 E176Q protein was released from its N-terminal thioredoxin tag by enterokinase digestion and further purified by adsorption of the tag to immobilized cobalt resin, followed by S200 gel filtration chromatography of the tag-free Bglu1 E176Q protein, as previously described for wildtype Bglu1 (Chuenchor et al., 2006, 2008). The protein was crystallized alone and in the presence of 1 mM 2,4-dinitrophenyl 2-deoxy-2-fluorogalactoside, up to 100 mM cellobiose, 2 mM cellotetraose, 1 mM cellopentaose or 0.5 mM cellobiose under conditions similar to those reported for wildtype Bglu1 (2 mg/mL protein, 20–23% PEG 5000 MME, 0.18–0.22 M (NH₄)₂SO₄, 0.1 M MES, pH 6.7, with microseeding). Apo Bglu1 E176Q crystals were also soaked with 1 and 10 mM laminaribiose (crystals soaked with 1 mM laminaribiose diffracted to 1.37 Å resolution, but had low occupancy of laminaribiose, while those soaked with 10 mM diffracted to 2.8 Å and had

higher occupancy, so it was used for the ligand structure building). The crystals were flash vitrified in liquid nitrogen after soaking in precipitant solution supplemented by 18% glycerol with the same concentrations of ligands. Data were collected with the crystals cooled by a 105 K nitrogen stream at the BL13B and BL13C beamlines of the National Synchrotron Radiation Research Center, Hsinchu, Taiwan, with 1.0 Å wavelength X-rays and ADSC Quantum 210 (BL13C) and Quantum 315 (BL13B) detectors. Data were indexed, refined and scaled with HKL2000 (Otwinowski and Minor, 1997).

The crystals of Bglu1 E176Q were isomorphous with those of previously reported wildtype Bglu1 crystals, allowing the structures to be solved by simple rigid body refinement with the wildtype structure (PDB code 2RGL) in Refmac 5.0 (Collaborative Computing Project Number 4, 1994; Murshudov et al., 1999). The structures were refined by restrained refinement with the same program and rebuilt with O (Jones et al., 1991). The occupancy of two possible modes of binding of cellobiose in the active site was refined by setting the total occupancy at 1.0 and testing different fractions of each ligand position to determine which gave the lowest ligand temperature (B) factor and R_{free} value. In doing so, a glycerol molecule was included at either the –1 or +5 subsite, for the +1 to +5 and –1 to +4 binding modes, respectively. The structural parameters of the final models were validated with PROCHECK (Laskowski et al., 1993), and are shown in Tables 3 and 4. Superimposed structures were viewed and structural figures were drawn with Pymol (Schrodinger, LLC). Pymol was also used to measure sugar torsion angles, while Cremer–Pople analysis of sugar puckering to identify the ring form (Cremer and Pople, 1975) was done via the Cremer–Pople calculator website of Shinya Fushinobu (Univ. of Tokyo).

2.3. Accession numbers

Coordinates and structure factors have been deposited in the Protein Data Bank under the accession numbers 3F4 V (apo Bglu1 E176Q), 3AHV (Bglu1 E176Q with G2F), 3F5 J (Bglu1 E176Q with cellotetraose), 3F5 K (Bglu1 E176Q with cellopentaose), and 3AHT and 3F5L (Bglu1 E176Q with laminaribiose).

3. Results

3.1. Catalytic residue mutation activities

In order to identify rice Bglu1 mutants appropriate for producing structures in complex with oligosaccharides, three mutations of the putative acid–base, E176A, E176D and E176Q were generated, along with two mutations of the catalytic nucleophile, E386D and E386Q. The mutant enzymes were produced as soluble proteins and their circular dichroism spectra were nearly identical to wildtype (data not shown). Under standard assay conditions, the activity of the nucleophile mutants was undetectable, and assays at increased enzyme concentrations and times showed that E386D and E386Q mutants were at least 3000- and 60,000-fold less active than wildtype Bglu1 in pNPGlc hydrolysis, respectively.

The acid–base mutations also had decreased activity, however, for E176Q, hydrolysis of pNPGlc was still significant in the standard assay. As acid–base mutants of retaining β -glycosidases are expected to be rescued by small nucleophiles (Wang et al., 1995), the activity of the enzymes in the presence of several buffers and nucleophiles was tested, as shown in Table 1. The E176Q basal activity in MES buffer was over 3-fold and 6-fold higher than the E176D and E176A mutants, respectively, and the E176Q activity increased greater than 7-fold and 13-fold in the presence of acetate and azide, respectively. The E176A activity also increased 2–3-fold in the presence of acetate and azide, but E176D was not rescued by

Table 3

Data collection statistics.

	3F4 V	3AHV	3F5 J	3F5 K	3AHT
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameter (Å, °)	$a = 79.8$ $b = 101.0$ $c = 127.6$ $\alpha = \beta = \gamma = 90^\circ$	$a = 79.2$ $b = 100.4$ $c = 127.4$ $\alpha = \beta = \gamma = 90^\circ$	$a = 79.7$ $b = 100.8$ $c = 127.5$ $\alpha = \beta = \gamma = 90^\circ$	$a = 79.8$ $b = 101.2$ $c = 127.4$ $\alpha = \beta = \gamma = 90^\circ$	$a = 79.7$ $b = 101.5$ $c = 128.3$ $\alpha = \beta = \gamma = 90^\circ$
No. of molecules per ASU	2	2	2	2	2
Resolution range (Å)	30–1.65 (1.71–1.65)	30–1.90 (1.97–1.90)	30–1.95 (2.02–1.95)	30–1.80 (1.86–1.80)	30–2.80 (2.90–2.80)
No. of observations	636,554	560,709	462,300	429,830	158,410
No. of unique observations	124,448	79,635	73,603	94,774	24,681
Redundancy	5.1 (5.0)	7.0 (6.7)	6.3 (5.8)	4.5 (3.6)	6.1 (6.3)
Completeness (%)	99.9 (99.9)	99.1 (98.9)	97.4 (93.7)	98.9 (93.3)	99.8 (99.8)
$I/\sigma(I)$	24.5 (4.5)	23.3 (6.1)	19.2 (6.9)	17.9 (2.7)	16.1 (5.8)
R_{sym} (%) ^a	6.2 (36.5)	8.3 (35.2)	8.3 (29.2)	8.3 (45.6)	11.7 (34.1)

Numbers in parentheses are outer shell parameters.

^a $R_{\text{sym}} = \sum_{\text{hkl}} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle| / \sum_{\text{hkl}} \sum_i I_i(\text{hkl})$.**Table 4**

Refinement statistics.

	3F4 V	3F5I	3F5 J	3F5 K	3AHT
Resolution range (Å)	30–1.65	26.29–1.90	30–1.95	30–1.80	30–2.80
No. of amino-acid residues	944	944	944	944	942
No. of protein atoms	7618	7618	7618	7618	7602
No. of water molecules	901	668	727	801	71
Refined carbohydrate	None	G2F	Cellotetraose	Cellopentaose	Laminaribiose
No. of carbohydrate atoms	None	22	90	112	46
No. of other hetero atoms	53	53	35	47	35
R_{factor} (%) ^b	17.5	18.6	18.0	17.8	20.7
R_{free} (%) ^c	19.6	18.9	21.4	21.0	24.8
Ramachandran statistics					
Favored region (%)	97.4	97.6	97.6	98.0	95.6
Allowed region (%)	2.6	2.4	2.4	2.0	4.4
Outlier region (%)	0	0	0	0	0
R.m.s.d. from ideality					
Bond distances (Å)	0.011	0.014	0.018	0.015	0.008
Bond angle (°)	1.357	1.461	1.585	1.512	1.057
Mean B factors (Å²)					
All protein atoms	12.03	14.43	9.15	14.74	21.874
Waters	24.43	27.35	19.10	26.88	31.87
Hetero atoms	19.54	28.44	20.59	34.49	46.97
Carbohydrate atoms	None	9.84	24.29	32.71	43.16
Subsite –1 (A/B)	None	9.67/10.01	20.32/19.58	26.09/23.99	41.20/41.93
Subsite +1 (A/B)	None	None	21.73/21.57	29.75/29.06	44.69/44.59
Subsite +2 (A/B)	None	None	21.94/21.82	26.97/26.94	None
Subsite +3 (A/B)	None	None	32.87/32.90	32.91/32.66	None
Subsite +4 (A/B)	None	None	None	49.01/46.49	None

^b $R_{\text{factor}} = (\sum |F_o| - |F_c|) / \sum |F_o|$.^c Based on 5% of the unique observations not included in the refinement.

any anionic nucleophiles. Formate and cyanate could also slightly rescue E176Q activity. Ascorbate was found in the active site of white mustard myrosinase, which also has Gln in the place of the catalytic acid–base (Burmeister et al., 2000), so rescue with 50 mM ascorbate was attempted. The E176 activity increased 15–18-fold in the presence of ascorbate, while E176A activity was increased 3–4-fold, E176D was not affected, and wildtype BGlu1 appeared to be inhibited. However, due to the instability of ascorbate, azide was used for further kinetic analysis.

In the presence of 50 mM azide, the apparent k_{cat} values of the E176Q and E176A mutants were nearly independent of the pH from pH 5–10, while those of the wildtype and E176D enzymes decreased 30- and 8-fold, respectively, as the pH increased from 5 to 10 (Fig. 2). The pH dependence over the acidic range of 3–5 was maintained in the mutants and wildtype enzyme. These data confirmed that E176 is the catalytic acid–base, and suggest that E176Q is unusually active with a high capacity to transglycosylate anionic

nucleophiles, whereas E176D appears to maintain an acid catalyst that is weakened by its greater distance from the substrate due to the shortened sidechain.

While wildtype BGlu1 hydrolyzed 2,4-dinitrophenyl- β -D-glucoside (dNPGlc), which has an aglycone with a pK_a of 3.96, with an apparent k_{cat} of $9.1 \pm 0.7 \text{ s}^{-1}$ compared to $4.7 \pm 0.3 \text{ s}^{-1}$ for pNPGlc (pK_a 7.18), the BGlu1 E176Q mutant had essentially the same apparent k_{cat} for both substrates in the presence of azide ($1.32 \pm 0.01 \text{ s}^{-1}$ for dNPGlc with 0.2 M azide and $1.84 \pm 0.05 \text{ s}^{-1}$ for pNPGlc with 0.32 M azide). In contrast, hydrolysis of oligosaccharides by BGlu1 E176Q was not observed.

3.2. Oligosaccharide binding by BGlu1 E176Q

The availability of a semi-active mutant that could be efficiently rescued by small nucleophiles for aryl glycosides, but not oligosaccharides, allowed kinetic investigation of the binding of

Table 1

Relative activities of wildtype and acid–base mutants of BGlu1 in the presence and absence of nucleophiles.

Buffer	Specific activity ($\mu\text{mol}/\text{mg}/\text{min}$)			
	WT	E176Q	E176A	E176D
MES	n.d.	0.070	0.011	0.022
Acetate/MES	6.8	0.51	0.028	0.029
Azide/MES	7.2	0.89	0.031	0.028
Ascorbate/MES	5.6	1.06	0.040	0.029
Formate/MES	7.7	0.111	0.014	0.026
TFA/MES	7.5	0.046	0.005	0.028
KF/MES	5.8	0.072	0.008	0.022
Cyanate/MES	7.2	0.036	0.007	0.027
Acetate/UB	7.3	0.33	0.021	0.024
Azide/UB	7.3	0.84	0.030	0.025
Ascorbate/UB	5.1	0.95	0.036	0.026
UB	7.9	0.052	0.013	0.028

* All anionic nucleophiles and MES buffer, were prepared 50 mM, pH 5.0. Universal buffer (UB) was 0.2 M boric acid, 0.05 mM citric acid mixed with 0.05 M tri-sodium phosphate to attain pH 5.0. Assays were performed with 1 mM pNPGlc, as described in the methods. “n.d.” means not determined.

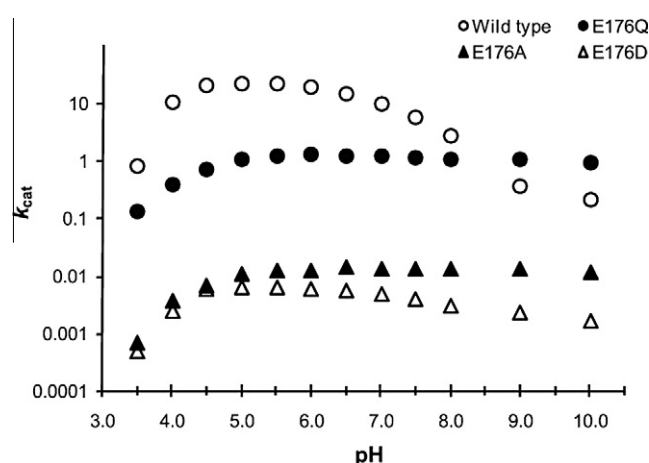


Fig. 2. Comparison of pH dependence of the BGlu1 wild type, E176Q, E176A, and E176D enzymes in 50 mM azide in universal pH buffer. The semi-log plot of the k_{cat} against the pH for each of the four enzymes to show that the E176Q and E176A mutants show little dependence of the rate on the pH between six and ten, whereas the wildtype and E176D mutant proteins have 10–100-fold decreases in k_{cat} over this range.

oligosaccharides to the BGlu1 E176Q active site by inhibition studies. Cellotriose, cellotetraose, cellopentaose, cellohexaose, and laminaribiose gave predominantly competitive inhibition (Table 2). From the competitive inhibition constants, apparent association constants for each oligosaccharide encompassing all possible binding modes could be calculated, along with the apparent overall

energy of binding. However, cellobiose was a very weak inhibitor that showed mixed inhibition, so only maximal values could be estimated for the competitive inhibition constant and association constant.

3.3. X-ray crystallographic structures of E176Q complexes

Binding of BGlu1 E176Q to oligosaccharides was also investigated by X-ray crystallography. Five structures were determined, including the apo BGlu1 E176Q mutant protein (PDB code 3F4V), its covalent complex with 2-fluoroglucoside (G2F, PDB code 3AHV), and its noncovalent complexes with cellotetraose (3F5 J), cellopentaose (3F5 K), and laminaribiose (3AHT). Attempts to cocrystallize or soak the crystals with cellobiose concentrations up to 100 mM resulted in no significant density in the active site and crystals with cellohexaose could not be obtained by co-crystallization. Diffraction and model parameters for the 5 structures can be found in Tables 3 and 4.

The apo BGlu1 E176Q (3F4 V) and BGlu1 E176Q covalent intermediate with G2F (3AHV) structures were nearly identical to the structures of wildtype BGlu1 (2RGL) and its complex with G2F (2RGM) (12), with an RMSD of 2RGL to 3F4 V of 0.163 Å over 926 C α atoms and for 2RGM to 3AHV of 0.245 Å over 925 atoms. For comparison, the wildtype structures (2RGL and 2RGM) had RMSD of 0.121 over 881 atoms, and the E176Q mutant structures (3F4 V and 3AHV) also had RMSD of 0.121 over 881 atoms.

3.4. Structural analysis of oligosaccharide binding

Fig. 3 shows the electron density omit maps (Fo–Fc maps calculated with the ligands removed) for the active site glycerol in the apo enzyme (A), and α -2-deoxy-2-F-glucoside (B), laminaribiose (C) cellotetraose (D) and cellopentaose (E) in their complexes with BGlu1 E176Q. Although the electron density for the reducing and nonreducing end glucosyl residues in the cellotetraose and cellopentaose complexes is less well defined, that for residues 2, 3 and 4 (in the cellopentaose structure) is clear. The nonreducing terminal residue was clearly found in the –1 subsite, since the density was more extensive than the glycerol in the apo enzyme active site, which superimposes on C2, O2, C3, O3, and C4, but not O4, C5, O5, C6 and O6 (Fig. 3 and Supplementary Fig. 1). The density for C1 was very weak, but the nonreducing glucosyl residue could be refined best when a glucosyl residue in a 1S_3 skew boat conformation was inserted and the C2–C1–O5–C5 dihedral angles refined to values of +3.5° for laminaribiose, +29° for cellotetraose and +24° for cellopentaose. This puts one O ϵ of E386 nearly directly in line with the glycosidic bond for its nucleophilic attack, at a distance of 3.15 Å in the cellopentaose structure, while the other E386 O ϵ makes a strong H-bond (2.53 Å) with the neighboring C2 OH (Fig 5 A). The cellopentaose nonreducing glucosyl residue with a C2–C1–O5–C5 angle of +24° appears to be somewhere between the

Table 2

Inhibition of BGlu1 E176Q hydrolysis of pNPGlc by gluco-oligosaccharides.

Oligosaccharide	Type of inhibition	Apparent K_i (mM)	Apparent K_A (M^{-1})	Apparent $\Delta G_{\text{binding}}$ *	Relative inhibitor potency (percent)
Cellobiose	Mixed	44 + 3	23	–7.9**	0.04
Cellotriose	Competitive	0.48 + 0.03	2,100	–19.3	3.8
Cellotetraose	Competitive	0.052 + 0.006	19,200	–24.8	34.6
Cellopentaose	Competitive	0.024 + 0.002	41,700	–26.8	75.0
Cellohexaose	Competitive	0.018 + 0.001	56,000	–27.5	100
Laminaribiose	Competitive	0.191 + 0.02	5,200	–21.6	9.4

* The apparent K_A and $\Delta G_{\text{binding}}$ are calculated based on the assumption of purely competitive inhibition. As such, they represent maximal values. The equations used to calculate these values were: $K_A = 1/K_D = 1/K_i$; $\Delta G_{\text{binding}} = -RT \ln K_A$, where R is the gas constant ($8.314 \text{ J K}^{-1} \text{ mol}^{-1}$) and T is the absolute temperature (303 K). The relative inhibitor potency was calculated by dividing the apparent K_A by that of the most potent inhibitor (cellohexaose) and multiplying by 100%.

** Since cellobiose gave mixed inhibition, the apparent K_A is expected to be an overestimate of the maximal value. The $\Delta G_{\text{binding}}$ is therefore more energetically favorable than is realistic.

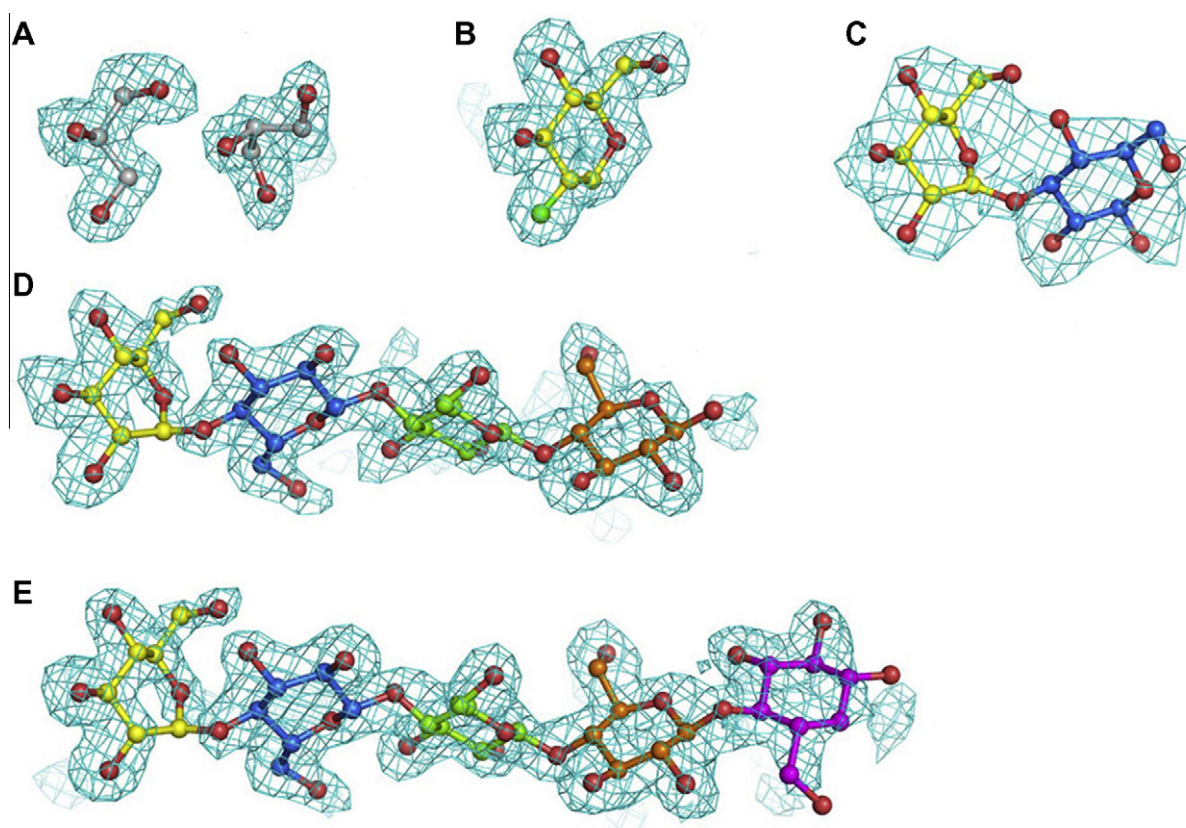


Fig. 3. The electron densities in the active sites of BGLu1 E176Q and its ligand complexes. The electron density (Fo–Fc) maps were calculated for each structure with all heteroatoms omitted from the active site, and are shown in blue mesh at $+2\sigma$. The ligands from the final structures are superimposed on these electron density maps and shown in ball and stick representation. (A) shows the BGLu1 E176Q apo enzyme to which no ligands were added, but in which 2 glycerol (carbons in silver) are seen in the active site. (B) shows the BGLu1 E176Q complex with 2-deoxy-2-fluoroglucoside, which is found in an alpha-linkage with the catalytic nucleophile. (C) shows the BGLu1 E176Q complex with laminaribiose, (D) shows the complex with cellotetraose, and (E) shows the complex with cellopentaose. The ligand glucosyl residues are color coded with carbons in yellow for the nonreducing glucosyl residue, blue for the second, green for the third, orange for the fourth and pink for the fifth. Oxygen is shown in red and fluoride in green. Direct overlays of the glycerol densities on those of cellopentaose and laminaribiose are shown in [Supplementary Fig. 1](#).

1S_3 skew boat (C2–C1–O5–C5 dihedral angle of $+45^\circ$) (Biarnes et al., 2006) and the expected 4H_3 half chair transition state or the closely related 4E envelope (C2–C1–O5–C5 dihedral angle of 0°), whereas this dihedral angle in other glucosyl residues was close to the relaxed 4C_1 chair value of -45° . Calculation of the Cremer–Pople parameters (Cremer and Pople, 1975) for these glucosyl rings gave $\varphi = 240^\circ$, $\theta = 58^\circ$ and $Q = 0.605$ for laminaribiose, $\varphi = 227^\circ$, $\theta = 70^\circ$ and $Q = 0.723$ for cellotetraose, and $\varphi = 231^\circ$, $\theta = 58^\circ$ and $Q = 0.675$ for cellopentaose, which were all closest to the values for a 4E envelope, though with higher puckering amplitudes (Q). This suggested the initially inserted 1S_3 skew boats had been substantially moved toward a 4E structure, which is closely related to the 4H_3 half chair and might act as a transition state, by the refinement. The weak density for C1 is consistent with movement between these ring structures in the complex.

The structure of BGLu1 with cellotetraose was close to that with cellopentaose with an RMSD of 0.078 Å (911 residues). The glucosyl residues of cellotetraose and cellopentaose bound to BGLu1 E176Q in the same positions for the first four residues (Fig. 4A), in line with the subsite model of oligosaccharide binding. The interactions between the five glucosyl residues of cellopentaose with the surrounding amino acids are mapped on the interaction diagram shown in Fig. 5A. Aside from the interaction of the catalytic nucleophile, mentioned above, several other direct interactions with Glc1 (the nonreducing terminal glucosyl residue) are found, in common with other GH1 structures (Sanz-Aparicio et al., 1998). No water molecules in position to interact with Glc1 were found in the -1 subsite, in contrast to all other subsites,

where nearly all polar interactions are mediated by water molecules. As seen in Fig. 5A, the only direct hydrogen bond between cellopentaose (and cellotetraose) Glc2 to Glc5 and BGLu1 E176Q is that between N δ of N245 and O2 of Glc3, and the Glc3 O3 atom is also 3.3 Å from this nitrogen, suggesting it may also form a hydrogen bond to N245. These interactions may help explain the strong binding at this subsite (Opassiri et al., 2004) (demonstrated in Table 2). Glucose residues Glc2 and Glc3 are aligned to stack onto the indole ring of W358, while the oligosaccharide twists nearly 90° between Glc3 and Glc4 to stack Glc4 and Glc5 along the face of the Y341 phenol ring (Fig. 4B). It is notable that in the apo enzyme structure Y341 and its neighbors in the flexible C loop, V342 and F343, are found in two possible positions, while they are stabilized in the one position by interactions with either cellotetraose or cellopentaose.

The poor density of the nonreducing end glucosyl residue of cellopentaose in the -1 subsite, together with some diffuse density in the apparent $+5$ site of the cellopentaose structure (Fig. 3E), led us to suppose that the cellopentaose was found in two binding modes in the crystal. Therefore, the cellopentaose model was built in both the productive position occupying subsites -1 to $+4$ and in the unproductive position occupying subsites $+1$ to $+5$, with a glycerol in the -1 subsite, as seen in the apo protein crystal structure. However, when the occupancy was refined to find the percent binding in each position that gave the lowest R_{free} and ligand temperature (B) factor, the best solution was one in which the occupancy in the productive position was at least 95% (with 5% binding from the $+1$ to $+5$ subsites). In fact, the density in the $+5$ site could be fit to a

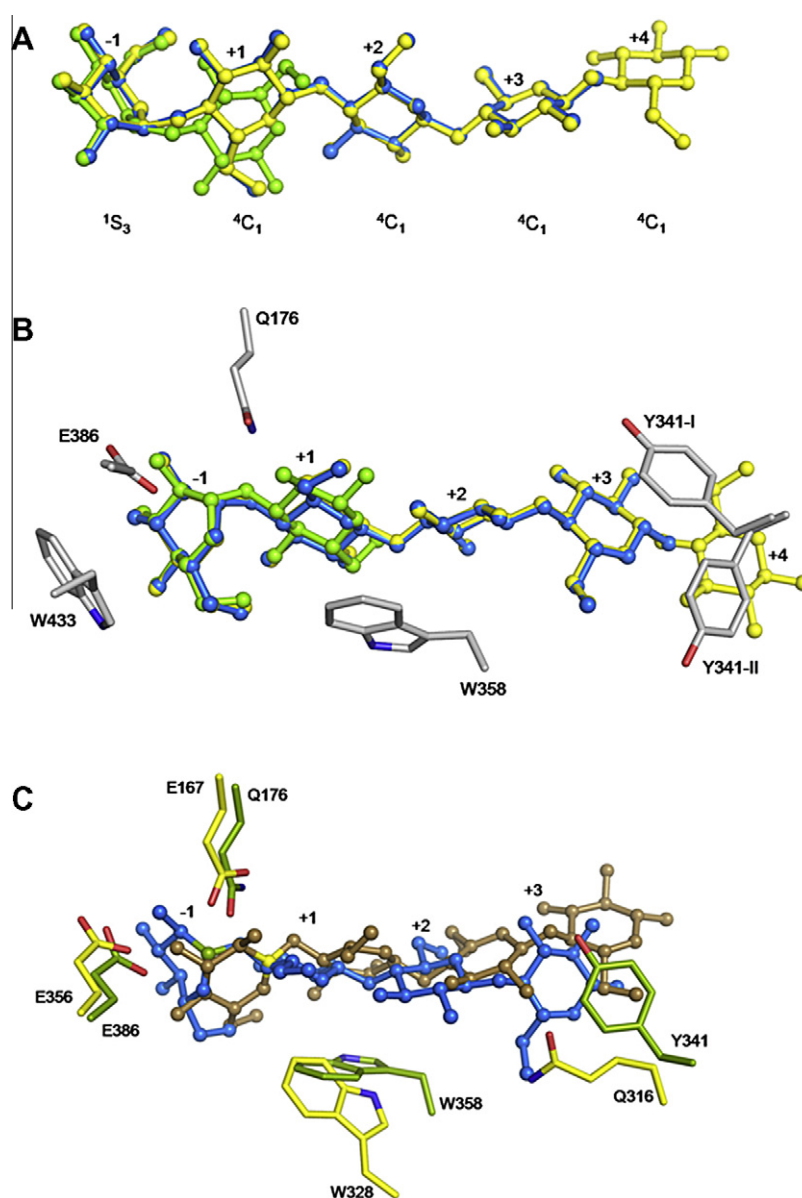


Fig. 4. Superposition of oligosaccharides in the active site. (A). Relative positions of cellotetraose, cellopentaose, and laminaribiose in the superimposed active sites of their E176Q mutant complexes. The sugar conformations fit into the density for each residue (4C_1 chair or 3S_1 skew boat) and the C2–C1–O5–C5 dihedral angles for glucosyl residues 1 and 2 from the nonreducing end in cellotetraose (Cel4), cellopentaose (Cel5) and laminaribiose (Lam2) are shown below that residue. Cremer–Pople analysis indicated that the refinement moved all 3S_1 skew boats to be best fit to 4E envelope rings in the final structures, while those residues in relaxed chairs stayed in that conformation. Cellotetraose (blue) and cellopentaose (yellow) were nearly completely superimposed, while the laminaribiose (red) Glc2 was flipped to lie on the opposite face, in addition to being connected to Glc1 C1 by O3 instead of O4. (B) Aromatic–sugar interactions most critical for stacking the sugars in their places. The Cel4, Cel5 and Lam2 are drawn as in A, while the residues W433 upon which the –1 subsite glucosyl residue stacks, W358, upon which the +1 subsite, and to a lesser extent +2 subsite, glucosyl residues stack, and Tyr341, upon which the +3 subsite, and to a lesser extent +4 subsite, glucosyl residues stack, along with the catalytic residues, are drawn with gray carbons. Two alternate conformations of Y341 (I and II) were seen in the electron densities generated from the high resolution data of the apo enzyme and its complexes with G2F and low occupancy Lam2, but Y341 is locked into position I by interactions with cellotetraose and cellopentaose. The +1 subsite glucosyl residue of Lam2 is flipped 180° compared to the cellooligosaccharides, thereby allowing the C1, C3, and C5 hydrogens to interact with the indole ring of W358. (C) Superimposition of rice BGlul (green) and *Paenibacillus polymyxa* β -glucosidase B (PpBGlulB, yellow) structures in complex with cellotetraose. Cellotetraose in complex with rice BGlul is colored blue, while it is sand colored in its complex from PpBGlulB. It can be seen that, whereas BGlul E386 is 3.2 Å from the C1 (colored green) of the nonreducing glucosyl residue of cellotetraose in its complex, the catalytic nucleophile of PpBGlulB, E356, is 6.3 Å away and blocked from making a direct attack on the anomeric carbon (C1, colored yellow) by C3. In addition to the catalytic acid–base and nucleophile, the aromatic amino acid side chains that stack the 2nd and 3rd glucosyl residues (Trp258) and 4th and 5th glucosyl residues (Tyr341) in BGlul and the corresponding residues in PpBGlulB are shown. The glucosyl residue-binding subsites (–1 to +3) are marked above the sugars.

glycerol better than a glucosyl residue, so little evidence was seen for unproductive binding. Therefore, it was deduced that, under the conditions in the crystals at the time of flash cooling, at least 95% of the cellopentaose bound in the –1 to +4 position. Cellotetraose was similarly largely in the productive position.

In the complex of BGlul E176Q with laminaribiose, the nonreducing sugar is in nearly the same position as in those with cello-

tetraose and cellopentaose, with the largest differences occurring at the positions of O5 and C1. The positions of these atoms and its glycosidic linkage to the C3 of the reducing glucosyl residue placed this reducing end residue in a distinct position (Fig. 4A) and the ring was flipped 180° compared to the Glc2 of the cellooligosaccharides. This orientation positioned the C1, C3 and C5 protons, which are expected to have a weak positive dipole due to

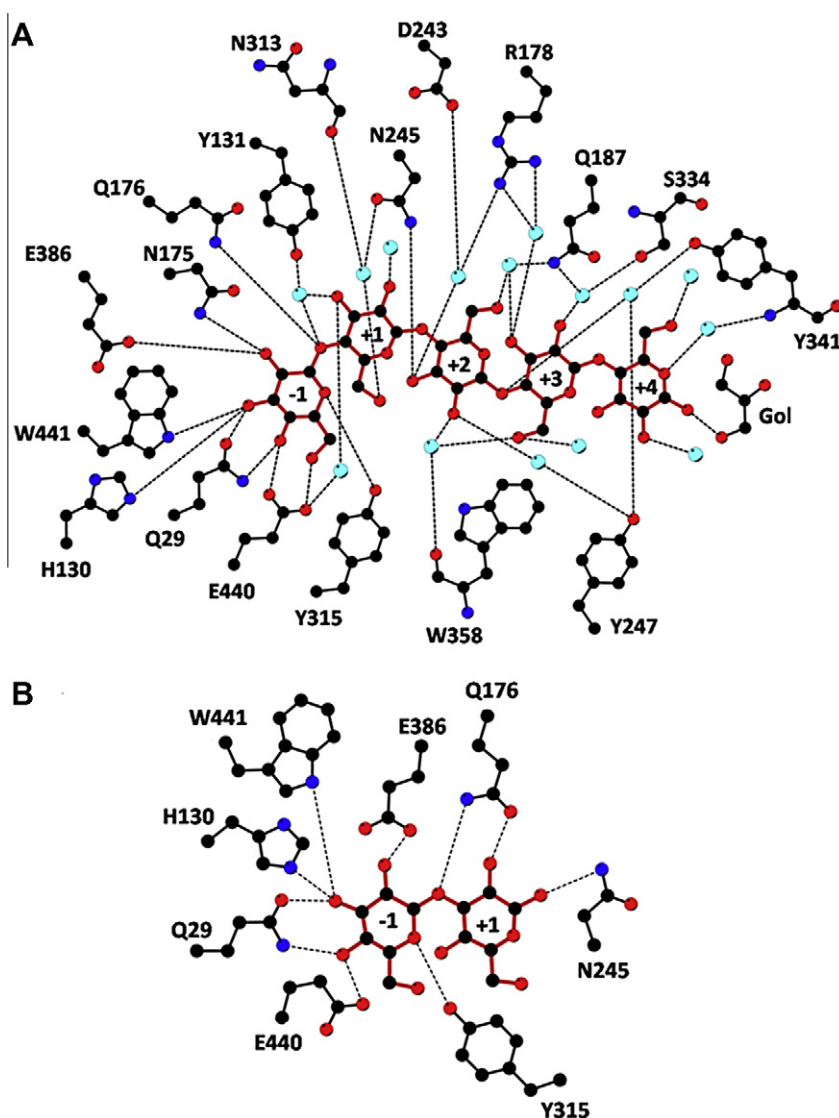


Fig. 5. Polar interactions between the BGluc1 E176Q mutant active site residues and cellopentaose (A) and laminaribiose (B). The diagram in A shows that all apparent hydrogen bonds (dotted lines shown where the distance is <3.2 Å) are directly between the cellopentaose Glc1 and the protein, but are mediated through water for all other residues, except for one direct hydrogen bond between N245 and Glc3 O3. In contrast, (B) shows that, in addition to water-mediated hydrogen bonds, laminaribiose forms three direct hydrogen bonds to the protein at Glc2, while Glc1 makes the same direct hydrogen bonds to the protein as does cellopentaose Glc1.

the sigma withdrawal of electrons to the sugar oxygen atoms, to make weakly polar interactions (sometimes called soft hydrogen bonds (Weiss et al., 2001)) with the aromatic electron cloud of the Trp358 indole ring to stack this sugar on that ring (Fig. 4B). The β -conformation required for this positioning of the C1 hydrogen, also facilitated the formation of a direct hydrogen bond between the C1 OH and the N245 sidechain, as well as a water-mediated hydrogen bond between the Glc2 O1 and the same sidechain (Fig. 5B).

4. Discussion

4.1. Analysis of catalytic residue mutants

The identity of the catalytic acid–base and nucleophile were already known, due to their conservation with those of other GH1 β -glucosidases and their positions in the structure. Additional evidence for the catalytic nucleophile was provided by the glycosynthase activity of E386A, E386G and E386S mutants (Hommalai et al., 2007), and the covalent labeling of the E386 with G2F

(Chuenchor et al., 2008), so loss of activity in the E386D and E386Q mutants was expected. The loss of activity in the E176 mutants, the rescue of E176A and E176Q by anionic nucleophiles, and the loss of the basic shoulder of the pH-activity profiles of these two mutants provided support for the identification of E176 as the acid–base catalyst, as noted previously for the corresponding *Agrobacterium* sp. β -glucosidase Abg mutants (Wang et al., 1995; Mülleger et al., 2005). It is notable that these mutations were most highly active in the presence of ascorbate, a cofactor for the hydrolysis of thioglucosides by myrosinase, which also has a Gln in the position corresponding to the GH1 catalytic acid–base (Burmeister et al., 2000). The hydrolysis of thioglucoside by almond β -glucosidase has also been shown to be promoted by low concentrations of ascorbate, but inhibited by high concentrations (Shen and Byers, 2007). Wildtype BGluc1 also shows inhibition by ascorbate (Table 1) and higher concentrations of ascorbate inhibited BGluc1 E176Q, but hydrolysis of pNP- β -D-thioglucoside was not detected under standard assay conditions (data not shown).

A similarly high activity to that seen with the BGluc1 E176Q mutant was seen with the corresponding mutation of Abg,

although that enzyme appeared to show more rapid hydrolysis of dNPGlc ($k_{\text{cat}} = 16 \text{ s}^{-1}$) than pNPGlc ($k_{\text{cat}} = 1 \text{ s}^{-1}$) (Mülegger et al., 2005). It has been noted previously that the *Agrobacterium* enzyme seemed to have a higher requirement for acid-assistance than other enzymes, such as *Cellulomas fimi* Cex (Wang et al., 1995), and, as shown here, rice BGlu1. The occurrence of the G2F covalent intermediate confirms that the 2,4-dinitrophenolate does not require acid-assistance in its departure, since glycosylation takes place despite the destabilization of the presumed oxycarbenium cation-like transition state by the electron withdrawal by the fluorine at C2. Recently, a similar covalent intermediate with an unmodified glucosyl residue (without a fluorine atom) with the acid–base E to Q mutant of human cytosolic β -glucosidase has been reported (Noguchi et al., 2008). Attempts by our group to produce a similar covalent intermediate with BGlu1 E176Q have so far failed to show noticeable density in the active site, perhaps due to the unusually high activity of this mutant of BGlu1.

Mülegger et al. (2005) suggested that the high activity of the Abg acid–base E to Q mutant may have resulted from a hydrogen bond made by the glutamine. Indeed, the Q176 N ϵ and the glycosidic oxygen of the scissile bond are within hydrogen bonding distance in both the covalent G2F intermediate and Michaelis-complex-like oligosaccharide complexes with BGlu1 E176Q (Fig. 4), as was seen the corresponding mutants of human cytoplasmic and strictosidine β -glucosidases (Noguchi et al., 2008; Barleben et al., 2007). In fact, despite not being hydrolyzed, the oligosaccharide complexes showed distortion of the nonreducing glucosyl residue between the $^1\text{S}_3$ skew boat and a planar transition state most similar to the ^4E envelope, based on the Cremer–Pople parameters and C2–C1–O5–C5 torsion angles for the refined structures. Movement between these structures may account for the poor electron density for the anomeric carbon (C1).

4.2. Comparison of substrate binding to other GH1 structures

The glycone moiety is bound by the conserved residues found in other GH1 β -glucosidases (Czjzek et al., 2000; Verdoucq et al., 2004; Sanz-Aparicio et al., 1998), as seen in the G2F-labeled wild-type enzyme (Chuenchor et al., 2008), but the aglycone-binding residues are variable among these enzymes, reflecting their different substrate specificities. Previously, the conserved Trp corresponding to W358 has been noted to be critical for binding of aglycones, and variability in its position has been noted to reflect substrate specificity (Czjzek et al., 2000, 2001; Zouhar et al., 2001; Verdoucq et al., 2003, 2004; Barleben et al., 2007). Other residues found to interact with the aglycone in maize Glu1 included F198, F205, and F466, which correspond to L183, N190, and L442 in rice BGlu1. In sorghum SbDhr1, V196 and L203, corresponding to the former two of these residues, formed hydrophobic interactions with the dhurriin phenolic ring, but S462, corresponding BGlu1 L442, was most critical in forming a water-mediated hydrogen bond to the aglycone phenolic oxygen (Czjzek et al., 2000; Verdoucq et al., 2004). Similar to SbDhr1, rice BGlu1 residue L183 appears to form a hydrophobic interaction with Glc residue 3. However, N190 is over 6 Å away and interacts only indirectly, while the substrate turns away from L442. Mutagenesis of L442 had no effect on catalytic activity (Chuenchor et al., 2008). F261 in SbDhr1 appears to restrict the active site, while M251 in this position has a similar effect on the rice Os3BGlu6 active site (Seshadri et al., 2009). In contrast, the corresponding BGlu1 N245 forms two hydrogen bonds to Glc residue 3, and its polar sidechain has been shown to contribute to hydrolysis of pNPGlc, as well as cellooligosaccharides (Chuenchor et al., 2008). Since the cellooligosaccharide substrates of BGlu1 are longer, the BGlu1 active site is more extensive than the other plant enzymes and includes residues further

out, such as Y341 and S334, which interact with the last two glucosyl residues.

4.3. Analysis of oligosaccharide binding

The increase in the value of the apparent association constant (K_A) for up to 6 residues, supports the previous prediction of 6 glucosyl residue-binding subsites based on subsite mapping by kinetic analysis of hydrolysis of oligosaccharides by BGlu1 (Opassiri et al., 2004). Since cellotetraose and cellopentaose appeared to bind predominantly in the productive positions with the four common glucosyl residues in the same positions, it is likely that the differences in binding energies for, at least, the $\Delta G_{\text{binding,cellohexaose}} - \Delta G_{\text{binding,cellopentaose}}$ and $\Delta G_{\text{binding,cellopentaose}} - \Delta G_{\text{binding,cellotetraose}}$ are reasonable estimates of the energy of binding subsites +5 and +4, respectively. In general, the changes of free energy of binding upon addition of a glucosyl residue correlate well with the number of direct interactions and water-mediated hydrogen bonds at the subsite to which that residue binds when the nonreducing end is bound to subsite –1. If the assumption is made that the binding of cellotriose to cellohexaose is overwhelmingly in the position that would be productive for hydrolysis in the wildtype enzyme with the nonreducing glucosyl residue in the –1 site, then subsite affinities of >11.4 kJ/mol for the +2 subsite, 4.5 kJ/mol for the +3 subsite, 2 kJ/mol for the +4 subsite and 0.7 kJ/mol for the +5 subsite can be calculated for binding of the (1 $\rightarrow$ 4)- β -linked D-glucosyl residues of cellooligosaccharides. The relative energies of interaction for the different subsites are similar to those calculated from subsite mapping (Opassiri et al., 2004).

The strong interactions between the reducing end glucosyl residue and N245 may explain why the binding of laminaribiose is more than 2 kJ/mol more energetically favorable than cellotriose. These interactions could only be formed for the β -anomer of the reducing glucosyl residue, demonstrating an apparent preference for the chirality of this residue. An initial high resolution (1.37 Å) dataset from a crystal soaked with 1 mM laminaribiose showed partial occupancy for the laminaribiose (PDB code 3F5L). Since the protein without ligands (3F4V) had 2 glycerol molecules in subsites –1 and +1 that overlapped with the laminaribiose position, their strong partial occupancies may have biased the laminaribiose placement. However, the dataset from a crystal that had been soaked in 10 mM laminaribiose collected to 2.8 Å resolution (PDB code 3AHT) had more substantial laminaribiose density.

4.4. Comparison to the complex of cellotetraose with *Paaenobacillus polymyxa* β -glucosidase B

Isorna et al. (2007) reported the structure of *P. polymyxa* β -glucosidase B with cellotetraose bound into the active site (PDB code 3Z1S), based on the partial electron density observed in the active site after a short incubation and flash freezing. In that case, the enzyme was fully active, unlike BGlu1 E176Q, which is active toward aryl β -glucosides with good leaving groups, but inactive toward oligosaccharides. In that structure, the catalytic nucleophile was 6.3 Å away from the anomeric carbon of the nonreducing sugar, which was in a relaxed chair conformation and out of position for a direct nucleophilic attack (Fig 4C). In contrast, in the BGlu1 E176Q oligosaccharide complexes, E386 is within 3.2 Å of the anomeric carbon, and directly opposite and inline with the glycosidic bond due to the distorted conformation of the nonreducing glucosyl residue. The position and distortion of cellotetraose and cellopentaose in the BGlu1 E176Q complexes appear to be indicative of a Michaelis complex, while the cellotetraose in the *P. polymyxa* BgluB model appears to bind with the glucosyl residues half way between the productive binding subsites, which is likely why it could be observed in an enzyme that is competent to hydrolyze

it. The fact that it could be observed may indicate that binding in this position is somewhat stable, which suggests that it may be a staging site for binding of substrates before insertion into the catalytic pocket and may have implications for the mechanism of such insertion of oligosaccharides. We have yet to observe such a staging position in rice BGLu1, but it may occur transiently in BGLu1 and other GH1 β -glucosidases.

5. Conclusions

The complexes of rice BGLu1 E176Q with gluco-oligosaccharide appear to be the first for a family GH1 hydrolase with oligosaccharides in productive positions for hydrolysis. The determination of the glucosyl residue binding sites in BGLu1 is critical for understanding how plant β -glucosidases have adapted to the job of cellobiosaccharide binding and hydrolysis. The demonstration that BGLu1 binds to the nonreducing terminal residue directly, yet interacts with the remaining glucosyl residues predominantly through water-mediated hydrogen bonds and weakly polar stacking interactions between aromatic residues and glucosyl CH groups, indicates the interaction involves dehydration of Glc1, while other glucosyl residues remain hydrated. It is possible that the transition from hydrated to hydrogen bonded requires a bound intermediate, which may have been observed in *P. polymyxa* BgluB (Isorna et al., 2007), but further investigation is needed to see if such an intermediate can be observed in plant family 1 glycoside hydrolases.

Acknowledgements

Chartchai Krittanai and Apichai Bourchookarn are thanked for assistance in circular dichroism of the wildtype and mutant enzymes. Chun-Jung Chen, Rodjana Opasiri, Phimonphan Chuankhayan, Worrapoj Oonant, Catleya Rojviriyi, and the staff of the NSRRC PX beamlines are thanked for helpful suggestions and assistance in data collection, processing and interpretation. Shinya Fushinobu is thanked for the use of his website for Cremer–Pople calculations. This work was supported by the Thailand Research Fund grant BRG5080007 and Suranaree University of Technology. R.C.R. was supported by A*STAR, Singapore. Portions of this research were carried out at the National Synchrotron Radiation Research Center, a national user facility supported by the National Science Council of Taiwan, ROC. The Synchrotron Radiation Protein Crystallography Facility is supported by the National Research Program for Genomic Medicine.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jsb.2010.09.021](https://doi.org/10.1016/j.jsb.2010.09.021).

References

- Barleben, L., Panjikar, S., Ruppert, M., Koepke, J., Stockigt, J., 2007. Molecular and structural characterization of hexameric β -D-glucosidases in wheat and rye. *Plant Cell* 19, 2886–2897.
- Barrett, T., Suresh, C.G., Tolley, S.P., Dodson, E.J., Hughes, M.A., 1995. The crystal structure of a cyanogenic β -glucosidase from white clover, a family 1 glycosyl hydrolase. *Structure* 3, 951–960.
- Biarnes, X., Nieto, J., Planas, A., Rovira, C., 2006. Substrate distortion in the Michaelis complex of *Bacillus* 1, 3–1, 4- β -glucanase: Insight from first principles molecular dynamics simulations. *J. Biol. Chem.* 281, 1432–1441.
- Burmeister, W.P., Cottaz, S., Driguez, H., Palmieri, S., 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.
- Burmeister, W.P., 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.
- Chuenchor, W., Pengthaisong, S., Yuvaniyama, J., Opasiri, R., Svasti, J., Ketudat Cairns, J.R., 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. Sect. F* 62, 798–801.
- Chuenchor, W., Pengthaisong, S., Robinson, R.C., Yuvaniyama, J., Oonant, W., Bevan, D.R., Esen, A., Chen, C.-J., Opasiri, R., Svasti, J., Ketudat Cairns, J.R., 2008. Structural insights into rice BGLu1 β -glucosidase oligosaccharide hydrolysis and transglycosylation. *J. Mol. Biol.* 377, 1200–1215.
- Collaborative Computing Project Number 4, 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D* 50, 760–763.
- Cremer, D., Pople, J.A., 1975. A general definition of ring puckering coordinates. *J. Am. Chem. Soc.* 97, 1354–1358.
- Czjzek, M., Cicek, M., Zamboni, V., Bevan, D.R., Henrissat, B., Esen, A., 2000. The mechanism of substrate (aglycone) specificity in β -glucosidases 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, W.P., Bevan, D.R., Henrissat, B., Esen, A., 2001. Crystal structure of a monocotyledon (maize ZMGLu1) β -glucosidase and a model of its complex with *p*-nitrophenyl β -D-thiogluconate. *Biochem. J.* 354, 37–46.
- Dopitová, R., Mazura, P., Janda, L., Chaloupková, R., Jeřábek, P., Damborský, J., Filipi, T., Kiran, N.S., Bryzobohatý, B., 2008. Functional analysis of the aglycone-binding site of maize β -glucosidase Zm-p60.1. *FEBS J.* 275, 6123–6135.
- Gloster, T.M., Roberts, S., Ducros, V.M., Perugini, G., Rossi, M., Hoos, R., Moracci, M., Vasella, A., Davies, G.J., 2004. Structural studies of the beta-glycosidase from *Sulfolobus solfataricus* in complex with covalently and noncovalently bound inhibitors. *Biochemistry* 43, 6101–6109.
- 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., Mornon, J.P., 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.
- Hommel, G., Withers, S.G., Chuenchor, W., Ketudat Cairns, J., Svasti, J., 2007. Enzymatic synthesis of cellobio-oligosaccharides by mutated rice β -glucosidases. *Glycobiology* 17, 744–753.
- Hrmova, M., Harvey, A.J., Wang, J., Shirley, N.J., Jones, G.P., Stone, B.A., Høj, P.B., Fincher, G.B., 1996. Barley β -D-glucan exohydrolases with β -D-glucosidase activity. *J. Biol. Chem.* 271, 5277–5286.
- Hrmova, M., MacGregor, E.A., Biely, P., Stewart, R.J., Fincher, G.B., 1998. Substrate binding and catalytic mechanism of a barley β -D-glucosidase/(1, 4)- β -D-glucan exohydrolase. *J. Biol. Chem.* 273, 11134–11143.
- Isorna, P., Polaina, J., Latorre-Garcia, L., Canada, F.J., Gonzalez, B., Sanz-Aparicio, J., 2007. Crystal structures of *Paenibacillus polymyxa* β -glucosidase B complexes reveal the molecular basis of substrate specificity and give new insights into the catalytic machinery of family I glycosidases. *J. Mol. Biol.* 371, 1204–1218.
- Jenkins, J., Lo Leggio, L., Harris, G., Pickersgill, R., 1995. β -glucosidase, β -galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes 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.
- Jones, T.A., Zou, J.Y., Cowan, S.W., Kjeldgaard, M., 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. Sect. A* 47, 110–119.
- Ketudat Cairns, J.R., Esen, A., 2010. β -Glucosidases. *Cell. Mol. Life Sci.* doi:10.1007/s00018-010-0399-2.
- Kuntothom, T., Luang, S., Harvey, A.J., Opasiri, R., Fincher, G.B., Hrmova, M., Ketudat Cairns, J.R., 2009. Rice family GH1 glycoside hydrolases with β -D-glucosidase and β -D-mannosidase activities. *Arch. Biochem. Biophys.* 491, 85–95.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283–291.
- Leatherbarrow, R.J., 2001. GraFit, 5th ed. Erithacus Software Ltd, Horley, UK.
- Mackenzie, L.F., Wang, Q., Warren, R.A.J., Withers, S.G., 1998. Glycosynthases: mutant glycosidases for oligosaccharide synthesis. *J. Am. Chem. Soc.* 120, 5583–5584.
- Mülegger, J., Jahn, M., Chen, H.-M., Warren, R.A.J., Withers, S.G., 2005. Engineering of a thioglycosylase: randomized mutagenesis of the acid-base residue leads to improved catalysts. *Prot. Engineer. Design Select.* 18, 33–40.
- Murshudov, G.N., Lebedev, A., Vagin, A.A., Wilson, K.S., Dodson, E.J., 1999. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. Sect. D* 55, 247–255.
- Nijkken, 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.
- Noguchi, J., Hayashi, Y., Baba, Y., Zkino, O., Kimura, M., 2008. Crystal structure of the covalent intermediate of human cytosolic beta-glucosidase. *Biochem. Biophys. Res. Comm.* 374, 549–552.
- Opasiri, R., Ponthong, B., Onkokoong, T., Akiyama, T., Esen, A., Ketudat Cairns, J.R., 2006. Analysis of rice glycosyl hydrolase family I and expression of Os4bglu12 β -glucosidase. *BMC Plant Biol.* 6, 33.
- Opasiri, R., Ketudat Cairns, J.R., Akiyama, T., Wara-Aswapati, O., Svasti, J., Esen, A., 2003. Characterization of a rice β -glucosidase highly expressed in flower and germinating shoot. *Plant Sci.* 165, 627–638.

- Opassiri, R., Hua, Y., Wara-Aswapati, O., Akiyama, T., Svasti, J., Esen, A., Ketudat Cairns, J.R., 2004. β -Glucosidase, exo- β -glucanase and pyridoxine transglucosylase activities of rice BGlu1. *Biochem. J.* 379, 125–131.
- Otwinowski, Z., Minor, W., 1997. Processing of X-ray diffraction data collected in oscillation mode. In: Carter, C.W., Sweet, R.M. (Eds.), *Meth. Enzymol.*, Vol. 276. Academic Press, New York, pp. 307–326.
- Sanz-Aparicio, J., Hermoso, J.A., Martínez-Ripoll, M., Lequerica, J.L., Polaina, J., 1998. Crystal structure of β -glucosidase A from *Bacillus polymyxa*: insights into the catalytic activity in family1 glycosyl hydrolases. *J. Mol. Biol.* 275, 491–502.
- Seshadri, S., Akiyama, T., Opassiri, R., Kuaprasert, B., Ketudat Cairns, J.R., 2009. Structural and enzymatic characterization of Os3BGlu6, a rice β -glucosidase hydrolyzing hydrophobic glycosides and (1 $\rightarrow$ 3)- and (1 $\rightarrow$ 2)-linked disaccharides. *Plant Physiol.* 151, 47–58.
- Shen, H., Byers, L.D., 2007. Thioglycoside hydrolysis catalyzed by β -glucosidase. *Biochem. Biophys. Res. Commun.* 362, 717–720.
- Sticklin, M., 2006. Plant genetic engineering to improve biomass characteristics for biofuels. *Curr. Opin. Biotechnol.* 17, 315–319.
- 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.
- Verdoucq, L., Czjzek, M., Moriniere, J., Bevan, D.R., Esen, A., 2003. Mutational and structural analysis of aglycone specificity in maize and sorghum β -glucosidases. *J. Biol. Chem.* 278, 25055–25062.
- Verdoucq, L., Moriniere, J., Bevan, D.R., 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.
- Wang, Q., Trimbur, D., Graham, R., Warren, R.A.J., Withers, S.G., 1995. Identification of the acid/base catalyst in *Agrobacterium faecalis* beta-glucosidase by kinetic analysis of mutants. *Biochemistry* 34, 14554–14562.
- Weiss, M.S., Brandl, M., Sühnel, J., Pal, D., Hilgenfeld, R., 2001. More hydrogen bonds for the (structural) biologist. *Trends Biochem. Sci.* 26, 521–523.
- Withers, S.G., Warren, R.A.J., Street, I.P., Rupitz, K., Kempton, J.B., 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.
- Wurch, T., Lestienne, F., Pauwels, P.J., 1998. A modified overlap extension PCR method to create chimeric genes in the absence of restriction enzymes. *Biotechnol. Techniq.* 12, 653–657.
- Zechel, D.L., Boraston, A.B., Gloster, T., Boraston, C.M., Macdonald, J.M., Tilbrook, D.M., Stick, R.V., Davies, G.J., 2003. Iminosugar glycosidase inhibitors: structural and thermodynamic dissection of the binding of isofagomine and 1-deoxynojirimycin to β -glucosidases. *J. Am. Chem. Soc.* 125, 14313–14323.
- Zouhar, J., Vévodová, J., Marek, J., Damborský, J., Su, X.-D., Bryzobohatý, B., 2001. Insights into the functional architecture of the catalytic center of a maize beta-glucosidase Zm-p60.1. *Plant Physiol.* 127, 973–985.