



รายงานวิจัยฉบับสมบูรณ์
Final Report

โครงการ โครงการ โครงสร้าง หน้าที่ และการใช้ประโยชน์ ของ บีตาไกลโคซิเดส

Structure, function and application of β -glycosidases

โดย ศ. ดร. เจมส์ เกตุทัต-คาร์นส์ และคณะ
By Prof. Dr. James R. Ketudat-Cairns and Team

สิงหาคม 2559

รายงานวิจัยฉบับสมบูรณ์
Final Report

โครงการ โครงสร้าง หน้าที่ และการใช้ประโยชน์ ของ บีตาไกลโคซิเดส

Structure, function and application of β -glycosidases

คณะผู้วิจัย สังกัด

Research Team and Affiliations

- Prof. Dr. James R. Ketudat-Cairns, Head of Project,
Institute of Science, Suranaree University of Technology
- Dr. Salila Pengthaisong, PhD, Full Time Postdoctoral Researcher, Coinvestigator
Institute of Science, Suranaree University of Technology
- Dr. Supaporn Baiya, PhD, Coinvestigator, Faculty of Science, Kasetsart
University Sriracha Campus.
- Dr. Jaggaiah Naidu Gorantla, PhD, Full Time Postdoctoral Researcher,
Coinvestigator, Institute of Science, Suranaree University
of Technology
- Dr. Ratana Charoenwattanasatien, PhD, Coinvestigator, collaborator,
Synchrotron Light Research Institute PLC, Thailand
- Mr. Akkarawit Prawisut, Ms. Kadsada Sala, Mr. Meng Huang, Ph.D. Students,
Institute of Science, Suranaree University of Technology
- Ms. Chamaiporn Beagbandee, MSc, Research Assistant, Institute of Science,
Suranaree University of Technology

This study was supported by The Thailand Research Fund
(The ideas presented are those of the investigator and are not necessarily
supported by The Thailand Research Fund)

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย
(ความเห็นในรายงานนี้เป็นของผู้วิจัย
สกว. ไม่จำเป็นต้องเห็นด้วยเสมอไป)

Thailand Research Fund Contract Number BRG5960015
Structure, function and application of β -glycosidases
Final Report

Grant Period: 31 August, 2016- 30 August, 2019.

Project Title: **Structure, function and application of β -glycosidases**

(Thai): โครงสร้าง หน้าที่ และการใช้ประโยชน์ ของ บีตาไกลโคซิเดส

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

Abstract

β -Glucosidases play many roles in living organisms and are important industrial enzymes for biomass utilization, and they and related transglycosidases can be used for production of glycosides for various purposes. Our group has nearly 20 years of experience studying β -glucosidase structure and function and in the current work we also demonstrated some new applications for these enzymes. In this project, we studied the structure, function and applications of glycoside hydrolase (GH) family GH1, GH3, GH116 β -glucosidases and transglucosidases. We demonstrated that rice Os4BGlu18 and its close relatives can act as monolignol β -glucosidases to modulate monolignol glucoside levels *in planta*, and solved the X-ray crystal structure of Os4BGlu18, the first for a monolignol β -glucosidase. We also demonstrated that the GH1 enzyme Os12BGlu38 is critical for pollen development. In a mutagenesis study, we identified variants of Os9BGlu31 transglucosidase that had the highest activity for several acceptor substrates, including water for hydrolysis in some cases. Machine learning could identify determinants of the activity on certain substrates, and the variants could be used to produce useful products, such as phytohormone glucoconjugates for further studies. In the GH3 work, two exoglucanases and a β -xylosidase from rice were characterized and found to be similar to previously described enzymes. Mutagenesis of barley HvExoI was combined with structural and computational studies to identify the mechanism by which glucose is released from the active site, which explained the basis for processive oligosaccharide/polysaccharide hydrolysis in a pocket-shaped active site. Continuing on from our previous solution of the first GH116 structure, that of TxGH116, we solved several more structures of acid/base mutants to support the conclusion of a vertical protonation mechanism, which is different from previously reported retaining GH. We also made several mutations of the active site substrate-binding residues and characterized the activities of the enzymes, and additional nucleophile residue mutations. The acid/base and nucleophile mutation variants were found to efficiently transglycosylate azide to make azido- β -D-glucose and azido- α -D-glucose, respectively, which could be used to make various glucosyl triazoles by copper-catalyzed Click chemistry. The β -glucosyl triazoles were found to be efficient inhibitors of bacterial GH116 β -glucosidase, but not human β -glucosidases, while the α -glucosyl triazoles hold promise as α -glucosidase inhibitors. Some of the active site TxGH116 mutations also improved the tolerance to glucose and substrates, which will be important if the enzyme is used for biomass conversion in the future. In addition, we also expressed and characterized the cyanobacterium β -glucosidase *Thermosynechococcus elongatus* TeGH116. This enzyme strongly preferred β -D-glucoside, but had low levels of activity on β -D-galactoside, N-acetyl- β -D-glucosaminide and β -D-mannoside as well, and could hydrolyze oligosaccharides. It may act to help recycle cell wall cellulose-oligosaccharides from the cell wall of the bacterium. We were able to crystallize and solve the structure of this enzyme, which could be the second structure from the GH116 family. This work contributed to the theses of three PhD students, as well as postdoctoral fellows and assistants, and has so far led to the publication of one national and four international journal papers and two patent application submissions. In the future, more papers and applications are coming out, once further development is carried out.

บทคัดย่อ

บีตา-กลูโคซิเดสมีบทบาทมากมายในสิ่งมีชีวิตและเป็นเอนไซม์อุตสาหกรรมที่สำคัญสำหรับการใช้ประโยชน์จากชีวมวล และเอนไซม์เหล่านี้และทรานส์โกลโคซิเดสที่เกี่ยวข้องสามารถใช้สำหรับการผลิตโกลโคไซด์สำหรับวัตถุประสงค์ต่างๆ กลุ่มของเรามีประสบการณ์เกือบ 20 ปีในการศึกษาโครงสร้างและหน้าที่ของบีตา-กลูโคซิเดสและในงานปัจจุบันเรายังได้ประยุกต์ใช้งานใหม่ๆ สำหรับเอนไซม์เหล่านี้ ในโครงการนี้เราได้ศึกษาโครงสร้าง หน้าที่และการประยุกต์ใช้งานของบีตา-กลูโคซิเดสและทรานส์โกลโคซิเดสในตระกูลโกลโคไซด์ไฮโดรเลส (GH) ที่ 3 และ 116 เราแสดงให้เห็นว่าเอนไซม์ Os4BGlu18 จากข้าวและเอนไซม์ที่มีความสัมพันธ์ใกล้ชิดกันสามารถทำหน้าที่เป็นโมโนกลีโคซิลบีตา-กลูโคซิเดสเพื่อรักษาสถิตของโมโนกลีโคซิลโกลโคไซด์ในพืช และหาโครงสร้างผลึกของ Os4BGlu18 ซึ่งเป็นโครงสร้างแรกของโมโนกลีโคซิลบีตา-กลูโคซิเดส นอกจากนี้ เรายังแสดงให้เห็นว่าเอนไซม์ Os12BGlu38 ในตระกูล GH1 นั้นมีความสำคัญต่อการพัฒนาของเรณู ในการศึกษาการกลายพันธุ์ เราได้วิเคราะห์หาเอนไซม์กลายพันธุ์ของทรานส์โกลโคซิเดส Os9BGlu31 ที่มีการทำงานสูงสุดสำหรับสารตั้งต้นตัวรับหลากหลาย รวมทั้งน้ำสำหรับการไฮโดรไลซิสในบางกรณี ซึ่งสามารถใช้เป็นแบบอย่างในการเรียนรู้การทำงานของเอนไซม์กับสารตั้งต้นและเอนไซม์กลายพันธุ์สามารถนำไปใช้ผลิตผลิตภัณฑ์ที่มีประโยชน์ เช่น ไฟโตสอร์บอนที่เป็นอนุพันธ์ของกลูโคส สำหรับการศึกษาเพิ่มเติม ในการศึกษาการทำงานของเอนไซม์ในตระกูล GH3 ได้แก่ เอนไซม์เอกโซกลูคาเนส 2 ชนิด และบีตา-ไฮโลซิเดสจากข้าว พบว่ามีการทำงานคล้ายกับเอนไซม์ที่อธิบายไว้ก่อนหน้านี้ การกลายพันธุ์ของเอนไซม์ HvExoI จากข้าวบาร์เลย์ ร่วมกับการศึกษาโครงสร้างและการวิเคราะห์ด้วยคอมพิวเตอร์เพื่อระบุกลไกที่กลูโคสถูกปล่อยออกมาจากบริเวณเร่ง ได้อธิบายพื้นฐานสำหรับการทำปฏิกิริยาไฮโดรไลซิสของโอลิโกแซ็กคาไรด์หรือโพลีแซ็กคาไรด์ในบริเวณเร่งที่มีรูปทรงกระเปาะ ดำเนินการต่อจากการศึกษาก่อนหน้าของโครงสร้าง GH116 แรกของเอนไซม์ TxGH116 เราได้หาโครงสร้างของเอนไซม์กลายพันธุ์ที่ตำแหน่งกรด/เบสอีกหลายตัวเพื่อรองรับข้อสรุปของกลไกการให้โปรตอนในแนวตั้งซึ่งแตกต่างจากกลไกการเร่งปฏิกิริยาแบบ retaining ของ GH116 ที่รายงานไว้ก่อนหน้านี้ นอกจากนี้เรายังได้ทำการกลายพันธุ์ของบริเวณเร่งที่ตำแหน่งของกรดอะมิโนที่จับกับสารตั้งต้นและศึกษาการทำงานของเอนไซม์กลายพันธุ์ รวมทั้งทำการกลายพันธุ์ของเอนไซม์ที่ตำแหน่งนิวคลีโอไฟล์ พบว่าเอนไซม์กลายพันธุ์ที่ตำแหน่งกรด/เบสและนิวคลีโอไฟล์มีประสิทธิภาพในการเคลื่อนย้ายกลูโคสไปยังอะไซด์ในการสร้าง อะซิโด-บีตา-ดี-กลูโคส และ อะซิโด-แอลฟา-ดี-กลูโคส ตามลำดับ ซึ่งสามารถนำมาใช้เพื่อสังเคราะห์กลูโคซิลไตรอะโซลชนิดต่างๆ ด้วยปฏิกิริยาคลิกเคมีที่เร่งด้วยคอปเปอร์ จากการศึกษามพบว่า บีตา-กลูโคซิลไตรอะโซล เป็นสารยับยั้งที่มีประสิทธิภาพของบีตา-กลูโคซิเดสในตระกูล GH116 จากแบคทีเรีย แต่ไม่ใช่จากคน ในขณะที่ แอลฟา-กลูโคซิลไตรอะโซล ถือเป็นสารยับยั้งแอลฟา-กลูโคซิเดส การกลายพันธุ์ของ TxGH116 ในบริเวณเร่งบางตัวยังช่วยเพิ่มความทนทานต่อกลูโคสและสารตั้งต้นซึ่งจะมีความสำคัญหากเอนไซม์นี้สำหรับการเปลี่ยนชีวมวลในอนาคต นอกจากนี้เรายังทำการแสดงออกของยีนและศึกษาการทำงานของบีตา-กลูโคซิเดส TeGH116 จากไซยาโนแบคทีเรีย *Thermosynechococcus elongatus* เอนไซม์นี้สามารถทำงานได้ดีมากกับสารตั้งต้นบีตา-ดี-กลูโคไซด์ แต่ทำงานได้ไม่ดีในบีตา-ดี-กาแลคโตไซด์ เอน-อะซิโด-บีตา-ดี-กลูโคซามิโนส และ บีตา-ดี-แมนโนไซด์ และสามารถไฮโดรไลซ์โอลิโกแซ็กคาไรด์ได้ เอนไซม์นี้อาจทำหน้าที่ช่วยรีไซเคิลผนังเซลล์ที่เป็นเซลล์โอลิโกแซ็กคาไรด์จากผนังเซลล์ของแบคทีเรีย เราสามารถดผลึกและหาโครงสร้างของเอนไซม์นี้ซึ่งเป็นโครงสร้างที่สองของตระกูล GH116 โครงการนี้มีส่วนเกี่ยวข้องกับวิทยานิพนธ์ของนักศึกษาระดับปริญญาเอก 3 คน นักวิจัยหลังปริญญาเอกและผู้ช่วย และนำไปสู่การตีพิมพ์ในวารสารระดับชาติ 1 ฉบับและนานาชาติ 4 ฉบับและการยื่นคำขอสิทธิบัตร 2 ฉบับ ในอนาคตจะมีตีพิมพ์ในวารสารและการประยุกต์ใช้เพิ่มขึ้นอีกเมื่อมีการพัฒนาเพิ่มเติม

Executive Summary

Importance:

β -Glucosidases (3.2.1.21) have many functions in plants and other organisms and are critical enzymes for biomass break-down and nutritional improvement of feeds, among other applications. Moreover, related enzymes include β -mannosidases, β -galactosidases, β -xylosidases, disaccharidases and transglycosidases that move sugar moieties from one molecule to another, rather than hydrolyzing the bond and releasing them. All of these enzymes have established or potential applications in science and industry.

β -Glucosidases belong to glycoside hydrolase (GH) families GH1, GH2, GH3, GH5, GH9, GH30 and GH116, with few verified activities outside these families (Ketudat Cairns and Esen, 2010; Ketudat Cairns et al., 2015; Lombard et al., 2014). In this study, we considered enzymes from GH1 and GH3 from plants (rice) and GH116 from bacteria. GH1 and GH3 are the most prominent families of plant β -glucosidases and related enzymes, like β -mannosidases and β -xylosidases/ α -L-arabinosidases. In these families, the enzymes exhibit a large range of specificities, especially in GH1, where β -glucosidases with β -mannosidase activity and β -mannosidases with β -glucosidase activity, disaccharidases and transglycosidases have been studied in our laboratory in the past (Chuankhayan et al., 2007; Kuntothom et al., 2009; Tankratok et al., 2013; 2015; Luang et al., 2013; Komvongsa et al., 2015a, 2015b). Aside from playing unique roles in cell wall metabolism of the plant, these enzymes also have potential applications in production of desired and unique glycoconjugates for characterization of their bioactivities, as well as biomass break-down. Thus, we set-out to discover the biological functions of the enzymes, the structural basis for these functions and ways to understand and engineer the enzymatic function of Os9BGlu31 transglucosidase by site-directed mutagenesis.

Among the GH3 plant enzymes, those that have been characterized are mainly from barley (Hrmova et al., 1996, 1998), although some from maize (Lee et al., 2001) and the small Thai eggplant *Solanum torvum* (Arthan et al., 2006) have also been characterized. Most of these enzymes act on cell wall polysaccharides and derived oligosaccharides, although the *S. torvum* enzyme does not hydrolyze these and instead acts on furanoglucosides. When the structure of the HvExoI exoglucanase was solved, it was found to have a glucose molecule left in the active site (Varghese et al., 1999), which was thought to come from the last hydrolysis reaction of the enzyme, which was purified from plants. Crystals of recombinant enzymes do not contain this glucose, unless soaked with glucose or substrate (Streltsov et al., 2019). Thus, it is curious to know how this glucose is removed when the new substrate comes and how this process affects the action of the enzyme on the substrates. Such knowledge could assist in engineering of these enzymes for roles in biomass breakdown and utilization. Another interesting aspect of GH3 is that the plant enzymes can be broken into two clades, one containing β -glucosidases, including endoglucanases, and one containing β -xylosidases and β -xylosidase/ α -L-arabinosidases. So, it is interesting to observe the factors that lead to different substrate glycon specificity.

The GH116 family was first described to contain human GBA2 glucoceramidase and certain archaeal enzymes, including a β -xylosidase/ β -glucosidase (Cobucci-Ponzano et al., 2010). We were the first to describe the activity of a bacterial enzyme in this family (Sansenya et al., 2015) and subsequently solved the first structure in this family (Charoenwattanasatien et al., 2016). Although the bacterial enzyme from *Thermoanaerobacterium xylanolyticum* does not hydrolyze the human GBA2 substrate glucosylceramide, it did serve as a reasonable model for the sugar-binding part of the active site, in which all the amino acid residues are conserved between TxGH116 and human GBA2. Several unique features were noted in this structure, including an unusual distance

between the catalytic nucleophile and catalytic acid/base, which was wider than seen in other retaining GH. Thus more work was needed to understand the catalytic mechanism and whether it can be manipulated to produce effective glycosynthase or glycothioligase activities (Wang et al., 1994, 1995; Ly and Withers, 1999; Müllegger et al., 2005; Chuenchor et al., 2011; Pengthaisong et al., 2012, 2014), although we had previously demonstrated the production of analytical amounts of glucosyl azides by such mutants of TxGH116 (Charoenwattanasatien et al., 2016). Indeed much remains to be learned about GH116, both in terms of possible utilization of TxGH116 in biomass breakdown and transglycosylation, and in the actions of other enzymes in this family, since only 4 had been functionally characterized and only one structurally characterized at the onset of this project.

Objectives:

The purpose of this project was to investigate the structure and function of beta-glucosidases and related enzymes from bacteria and plants and utilize this information to apply engineered enzymes to production of new and useful products. To achieve this goal, we addressed the following objectives:

1. To determine the structures of new enzymes from GH1 β -glucosidases/transglucosidases in order to explain the structural basis for the unique properties of these enzymes.
2. To determine the thermodynamics of binding of oligosaccharide and other substrates and mechanism-based inhibitors to β -glycosidase and/or transglycosidase enzymes and assess the ligand functional groups and enzyme residues responsible for these interactions.
- 3 To explain the difference between plant GH3 exoglucanases and β -xylosidases.
4. To explain the structural basis of TxGH116 β -glucosidase binding of oligosaccharides to allow engineering of the enzyme for better biomass conversion and to better model substrate-enzyme interactions in this family.
5. To engineer the TxGH116 β -glucosidase to improve glucose tolerance for effective hydrolysis of oligosaccharides in biomass conversion.
6. To functionally characterize a new GH116 enzyme from a eukaryotic or bacterial system for possible structural studies (if significantly different from the TxGH116 enzyme).

Methods and Results:

In this project, we functionally and structurally characterized rice β -glucosidases and engineered transglucosidases from GH1, barley and rice GH3 β -glucosidases and β -xylosidase, and GH116 enzymes from bacteria by recombinant expression, mutagenesis, enzymology and x-ray crystallography.

For GH1, we characterized the function and structure of Os4BGlu18. We had previously characterized the function of Os4BGlu18 expressed in *E. coli* (Baiya et al., 2015), but here we expressed Os4BGlu18, as well as Os4BGlu14 and Os4BGlu16 in *Arabidopsis* that were mutant for one monolignol β -glucosidase. We extracted the plants and measured monolignol and monolignol glucoside levels by triple quadrupole liquid chromatography tandem mass spectrometry (LCMSMS), and found that the monolignol glucoside levels were higher in the knockout plants, but reduced to similar to or lower than wild type in plants expressing Os4BGlu14, Os4BGlu16, and Os4BGlu18 (Baiya et al., 2018). This suggested that these enzymes act as monolignol β -glucosidases *in planta*, and that although Os4BGlu18 was shown to localize to the cell wall, the enzymes could modulate levels of monolignol glucosides, which are thought to be localized to the vacuole. We also utilized the crystals

that we obtained earlier to obtain x-ray crystal structures of Os4BGlu18 alone and with δ -gluconolactone inhibitor in the active site. These structures allowed docking of monolignol structures into the active site to analyze the determinants of monolignol β -glucoside specificity.

In another functional study, we expressed Os12BGlu38 in *E. coli* and yeast and showed it had basic β -glucosidase activity against synthetic substrate and oligosaccharides. This supplemented the work of our collaborator Jong-Seong Jeon's lab, in which they showed that knockout of Os12BGlu38 causes abortive pollen development, with little or no intine cell wall development.

To learn more about what makes Os9BGlu31 a transglucosidase, instead of a glucosidase and improve its usage for synthesizing glucoconjugates, we mutated residue 243 to all possible common amino acids, and combined the previously identified high activity W243N mutation with hydrophilic active site mutations by site-directed mutagenesis. By this approach, we were able to identify several new high activity mutation variants, including W243L and L241D/W243N, which could produce certain substrates better than the high activity W243N variant in enzymatic reactions analyzed by UPLC of the products. Machine learning analysis of the activities of the variants on various substrates could identify factors contributing to the high activity of certain variants on specific substrates (Tran et al., 2019). We were also able to use these enzymes to produce abscisic acid glucose ester (patent pending) and other phytohormone glucoconjugates for use in studying phytohormone β -glucosidases.

Our work on GH3 included following up our previous expression of barley Exo I (HvExoI) in *Pichia pastoris*, with enzymatic and structural studies of active site mutations, and further characterization of rice exoglucanases and β -xylosidase. The structural analysis of HvExoI mutations, together with computational simulations, could explain the exchange of released glucose and incoming substrate in the active site, leading to the proposal of a new progressive mechanism of hydrolysis. Expression of one rice exoglucanase, OsExoII, in *E. coli*, another, OsExoI, in *Pichia pastoris* and a β -xylosidase, OsXylI, in both, allowed us to characterize their enzymatic activities more thoroughly.

For GH116, we continued to characterize TxGH116 β -glucosidase by mutagenesis, enzymology and x-ray crystallography, and cloned, characterized and began to solve the structure of a new β -glucosidase from a cyanobacteria, TeGH116. Characterization of additional acid/base mutants and their structures with oligosaccharides confirmed the unusual protonation geometry in the catalytic mechanism, while quantum mechanic/ molecular mechanic simulations by our collaborators in Carme Rovira's group confirmed the plausibility of this mechanism. We made further nucleophile mutations as well and tested each for transglycosylation activity. This led to the idea that they could be used to make click glycosides (glucosyl triazoles) by enzymatic synthesis of azido- β -D-glucose and azido- α -D-glucose (patents pending). We were able to make gram quantities of azido- β -D-glucose and use it to synthesize 15 glucosyl-1,2,3-triazoles as putative β -glucosidase inhibitors, some of which had high inhibition activity toward TxGH116 (Gorantla et al., 2019). A smaller set of α -glucosyl-1,2,3-triazoles was also synthesized, which we hope to test as α -glucosidase inhibitors in the future. In addition, we have made mutations of all of the active site residues that appeared to be involved in glycon sugar binding in the crystal structure and confirmed that all of these had lower k_{cat}/K_M values than wild type, although a few had higher k_{cat} values, suggesting applications in high substrate situations like biomass conversion.

After several attempts to express animal GH116 enzymes and one plant GH116 enzyme in *E. coli* and yeast, we decided to express a cyanobacterium β -glucosidase, which

may have similar function to a plant GH116 β -glucosidase. We cloned the *Thermosynechococcus elongatus* TeGH116 gene into pET32a, expressed it in *E. coli*, purified and characterized it. The enzyme had activity on 4-nitrophenyl β -D-glucoside with much lower activity (8% relative activity) with β -D-galactoside. Other 4NP-glycosides were hydrolyzed at less than 1% of the relative activity of the β -glucoside. It also hydrolyzed celooligosaccharides and laminarioligosaccharides, suggesting it could be involved in recycling of the cellulose in the cyanobacterium cell wall. We were able to crystallize the protein under various conditions and a few gave diffracting crystals allowing us to collect a 2.5 Å dataset, from which a preliminary structure has been solved.

Output and Benefits:

In this project, we gained considerable insight into the workings of three families of glycoside hydrolases, which resulted in publication of 5 papers, with 4 more in preparation or revision, and submission of two patent applications, as well as contributing to the training of 3 PhD students, 1 postdoc and 1 research assistant. This work provided us the knowledge to produce several new products, including phytohormone glucosides, which we could use for other studies in the laboratory and thereby save considerable expenses, as well as sets of new glucosyl triazoles that can be tested for inhibition of enzymes of interest. This may result in the development of new drugs to help in treating metabolic disorder and other diseases, as these products are taken for testing in the future. Moreover, our approach may inspire others to create new glycosyl compounds. The knowledge gained on the enzymes' mechanisms also will help us to apply these enzymes and engineer them for better use in the future, as well as providing background for people working on related enzymes. Thus, we were able to produce unexpected products generated from application of our basic research, and expect application of these knowledge gained in this project will lead to more products in the future.

BRG5980015

1. Project Name

Thai name: โครงสร้าง หน้าที่ และการใช้ประโยชน์ ของ ปีตากลูโคลิเดสจากพืช และเอนไซม์ที่คล้ายคลึงกัน

English name: **Structure, function and application of β -glycosidases**

2. Project Participants

2.1.1 Name: Mr. James R. Ketudat Cairns

2.1.2 Degree: Ph.D.

2.1.3 Position: 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: Ms. Salila Pengthaisong

2.2.2 Degree: Ph.D

2.2.3 Position: Full time Postdoctoral Researcher

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, kinetic and structural investigation of rice Os3BGlu7 glycosynthase and active site cleft mutants; along with structure refinement, mutagenesis, enzymology and mutant crystallization for TxGH116.

2.3.1 Name: Mr. Jaggaiah Naidu

2.3.2 Degree: Ph.D

2.3.3 Position: Full time Postdoctoral Researcher 2015-2016

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

2.3.5 Position and responsibility: Coinvestigator for organic synthesis of inhibitors and click-products from enzymatically produced glucosyl azides.

2.4.1 Name: Mr. Akkarawit Prawisut

2.4.2 Degree: B.Sc.

2.4.3 Position: Ph.D. Student

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

2.4.5 Position and responsibilities: Student investigating rice GH3 exoglucanase-like enzymes.

2.5.1 Name: Ms. Kadsada Sala

2.5.2 Degree: M.Sc.

2.5.3 Position: Ph.D. student

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

2.5.5 Position and responsibilities: Student investigating GH3 β -xylosidase.

2.6.1 Name: Ms. Ratana Charoenwattanasatien

2.6.2 Degree: Ph.D.

2.6.3 Position: Postgraduate Researcher 2013-2014

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

2.6.5 Position and responsibilities: Coinvestigator, data collection and refinement of
TeGH116

2.7.1 Name: Ms. Supaporn Baiya

2.4.2 Degree: Ph.D.

2.4.3 Position: Instructor

2.4.4 Work address: Faculty of Science at Sriracha, Kasetsart University, Sriracha
Campus, Chonburi 20230

2.4.5 Position and responsibilities: Coinvestigator, characterization and structure
solution of Os4BGlu18.

2.8.1 Name: Mr. Meng Huang

2.8.2 Degree: M.Sc.

2.8.3 Position: Ph.D. Student

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

2.8.5 Position and responsibilities: Student, characterizing TxGH116 mutants.

2.9.1 Name: Ms. Chamaiporn Beagbandee

2.9.2 Degree: M.Sc.

2.9.3 Position: Research Assistant

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

2.9.5 Position and responsibilities: Research Assistant, cloning, expression,
characterization of TeGH116, General molecular cloning support work.

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

4. Problem and Its Importance

Simple and complex carbohydrates play structural, nutritional/energy storage, signaling, and metabolic roles, among others, in living organisms. These roles are modulated by the enzymes that act on carbohydrates, including glycoside hydrolases (GH) and lyases (GL), which break glycosidic linkages between sugar residues and between sugar residues and other moieties, and glycosyltransferases (GT) and transglycosidases (TG), which generate these linkages. Together, these enzymes, along with lytic polysaccharide monooxygenases, carbohydrate esterases and acyl transferases, and enzymes acting on polysaccharide-associated lignin, have been designated “Carbohydrate Active eZYmes” or “cazymes” and are catalogued in protein-sequence similarity-based families in the CAZY online database (www.cazy.org) (Lombard et al., 2014). Cazymes play critical roles in all domains of life, with the greatest expansion of these roles seen in eukaryotes, and particularly in plants, which display the widest range of carbohydrate-containing molecules.

The β -glucosidases that have been characterized are members of GH families GH1, GH2, GH3, GH5, GH9, GH30 and GH116, with the enzymes from plants falling in GH1, GH3, GH5 and

GH116, while those from humans fall in GH1, GH30 and GH116 (Ketudat Cairns and Esen, 2010, Ketudat Cairns et al., 2015, Lombard et al., 2014; Kallemeijin et al., 2014). Within these families, enzymes with related activities, such as β -mannosidase, β -galactosidase, β -xylosidase, disaccharidase, transglucosidase, and transgalactosidase, have also been described. In plants, a wide range of functions for these enzymes has been described, including cell wall recycling, defense against herbivores and microorganisms, activation of phytohormones, release of aromatics, release of glucosyl blocking groups from intermediates in secondary metabolism, and release of monolignols from storage forms, as well as glycolipid recycling, anthocyanin biosynthesis and transglycosylation of various phenolic acids. In humans, the GH1 enzymes appear to play nutritional roles for utilizing lactose and plant glucosides, while the GH30 and GH116 enzymes are involved in glycosphingolipid degradation. Notably, mutations and low expression of the GH1 enzyme lactase phloridzin hydrolase (LPH) result in lactose intolerance, while mutations in the GH30 enzyme GBA1 result in Gaucher disease and contribute to Parkinsonism and mutations in the GH116 enzyme GBA2 result in ataxia or paraplegia with spasticity (Hammer et al., 2013; Martin et al., 2013).

β -Glucosidases and related enzymes are of biotechnological interest as well. β -glucosidases from bacteria, archaea and fungi have been used for degrading oligosaccharides to glucose in biomass conversion (Ketudat Cairns and Esen, 2010). Identifying the functions of the enzymes in plants and the genetic effects of their over or under expression is also critical to developing low lignin and phenolic cell walls for biomass conversion, as well as developing crops with better yield and resistance to biotic and abiotic stress (Chapelle et al., 2012; Ketudat Cairns et al., 2015). In addition, these enzymes can be used to convert less active glycosides to their more active aglycones, such as in the case of isoflavonoids (Chuankhayan et al., 2007), or to improve the solubility and bioavailability of other bioactive compounds (Komvongsa et al., 2015a).

Our group has been studying rice β -glucosidases and related enzymes for nearly 20 years, and has also investigated related plant enzymes and carbohydrate active enzymes from other organisms, including humans, in collaboration with the Biochemistry Laboratory of Chulabhorn Research Institute, under direction of Prof. M.R. Jisnusun Svasti. Aside from studying the activities of the enzymes in order to understand their functions in the plants and potential *in vitro* applications, we have been interested in elucidating the structural basis for their substrate specificities and other enzymatic properties (Chuenchor et al., 2008, 2011; Seshadri et al., 2009; Sansenya et al., 2012; Tankrathok et al., 2013; Pengthaisong et al., 2012, 2014). This structure and functional understanding of the enzymes can allow us to engineer appropriate enzymes for better utilization in synthesis or degradation, as well as helping us to more accurately predict the activities of as yet uncharacterized enzymes from public genomic databases. Thus, in this project, we continued to study the structure and function of enzyme of the GH1, GH3 and GH116 families, and found ways to apply these enzymes to glycosylation of molecules of interest, as well as beginning to improve the properties of one for the depolymerization of biomass.

5. Objectives

The purpose of this project was to investigate the structure and function of beta-glucosidases and related enzymes from bacteria and plants and utilize this information to apply engineered enzymes to production of new and useful products. To achieve this goal, we addressed the following objectives:

1. Determination of the structures of new enzymes from GH1 β -glucosidases/transglucosidases in order to explain the structural basis for the unique properties of these enzymes.
2. Determination of the thermodynamics of binding of oligosaccharide and other substrates and mechanism-based inhibitors to β -glycosidase and/or transglycosidase

enzymes and assess the ligand functional groups and enzyme residues responsible for these interactions.

3 Explanation of the difference between plant GH3 exoglucanases and β -xylosidases.

4. Explanation of the structural basis of TxGH116 β -glucosidase binding of oligosaccharides to allow engineering of the enzyme for better biomass conversion and to better model substrate-enzyme interactions in this family.

5. Engineering of TxGH116 β -glucosidase to improve glucose tolerance for effective hydrolysis of oligosaccharides in biomass conversion.

6. Functional characterization of a new GH116 enzyme from a eukaryotic or bacterial system for possible structural studies (if significantly different from the TxGH116 enzyme).

6. Activities and Methods:

6.1. Structural and Functional Characterization of GH1 enzymes

During the first year, the structure of Os4BGlu18, a rice monolignol β -glucosidase was solved, and the catalytic acid/base was mutated from Glu to Gln, in order to allow production of crystals in complexes with monolignol glucosides, such as coniferin and 4-hydroxycinnaminic acid. The native and mutant proteins were expressed as thioredoxin fusion proteins in Origami B(DE3) *E. coli*, purified, and crystallized. The crystals were soaked with coniferin, coniferyl alcohol, 4-coumaryl alcohol, 4-coumaryl alcohol glucoside, and syringin, as well as the inhibitors 2,4-dinitrophenyl 2-deoxy-2-fluoroglucoside and δ -gluconolactone. These crystals were frozen and data collected at the SPring-8 and NSRRC synchrotrons with 0.9000 and 1.0000 Å X-ray beams, respectively. The data were processed with the HKL-2000 software package and the indexed, refined and scaled datasets were transferred to the CCP4 package for molecular replacement in MolRep and subsequent structure refinement in Refmac5 (Otwinowski and Minor, 1997; Vagin et al., 1997; Murshudov et al., 1999). Rebuilding, manual refinement and quality control were done in coot (Emsley and Cowtan, 2004). During the second year the structures of the native protein and its complexes with δ -gluconolactone and 2,4-dinitrophenyl 2-deoxy-2-fluoroglucoside were refined, essentially to completion. However, some small mistakes were discovered, so the refinement was repeated in the 3rd year. The structures were validated with Molprobtity and PROCHECK (Laskowski et al., 1993; Davies et al., 2007). Furthermore, the Os4BGlu18 structure was used to dock the monolignol glucoside structures, since we were unable to resolve electron density for these substrates in the diffraction data of the putative complexes.

In our earlier effort to verify the biological function of Os4BGlu18, along with Os4BGlu14 and Os4BGlu16, the cDNAs for these genes were cloned into the pGWB502 expression vector, the recombinant vectors transformed into *Agrobacterium* and introduced into *Arabidopsis* (Columbia landrace) with an insertional mutant in the *AtBGLU45* gene. This mutant line was previously shown to have elevated levels of the monolignol glucoside coniferin (Chapelle et al., 2012). After growing and confirming the recombinant plants at Kyung Hee University (S. Korea), the areal parts of wild type Columbia, *atbglu45* mutant, and *atbglu45* mutant rescued with *Os4BGlu14*, *Os4BGlu16* or *Os4BGlu18* were harvested, powdered in liquid nitrogen and extracted with methanol. The relative amounts of coniferin, syringin, and 4-coumaryl alcohol glucoside, as well as their aglycones and control phenolics were determined by triple quadrupole electrospray ionization (ESI) LCMSMS in the multiple reactant monitoring (MRM) mode by comparing the relative abundances of the compounds of interest to standard curves of those compounds. In the second year, we had to extend this work to show the localization of Os4BGlu18 protein by transformation of tobacco epithelial

cells, in order to show that Os4BGlu18 is localized at the cell wall. This data allowed us to complete publication of the paper on this subject (Baiya et al., 2018).

In a previous grant, we characterized Os9BGlu31 as a transglucosidase that moves glucose most effectively between phenolic acids and their glucose esters and established mutated enzymes that had greater activity than wild type (Luang et al., 2013; Komvongsa et al., 2015a, 2015b). In this grant, we continued mutation of the position of residue W243, mutations of which had previously been shown to have higher activity than wild type, by completing the set of 20 common amino acids in this position, and characterizing their activities for glucosylation of a set of acceptor substrate molecules (Tran et al., 2019). We also combined the previously described W243N mutation with other active site mutations to see the effects of double mutants on the enzyme activity and specificity. Since we lacked a crystal structure, homology models were generated with a covalent glucosyl intermediate for each of the Os9BGlu31 variants and docking of known acceptors was attempted. To gain further insight into the structural basis for acceptor substrate specificity, we also submitted the protein and acceptor substrate structures to cheminformatics analysis with machine learning (Tran et al., 2019). Furthermore, the high activity mutants, which seemed more stable, were submitted to crystallization trials, since we had previously failed to get crystals of the wild type Os9BGlu31 enzyme.

In another area, we produced recombinant Os12BGlu38 in *E. coli* and *P. pastoris* cultures to characterize the biochemical activity of the enzyme. The crude and purified enzymes were tested for activity on 4-nitrophenyl beta-D-glucoside (4NPGlc) and natural substrates. This work complemented the genetic studies done with Prof. Jong-Seong Jeon in Kyung Hee University, Korea, which showed that knock-out of this gene results in male-sterile rice plants with abortive pollen development.

Further GH1 work was done on generating new crystals of Os3BGlu7 with mannoimidazole and collecting a new dataset, since the previous ones only had glucoimidazole contaminant or no ligand in the active site, as well as crystallizing Os4BGlu13 for a clearer explanation of why it is better at gibberellin glucose ester hydrolysis than the closely related Os4BGlu12. The structures of Os3BGlu7 with mannose have been submitted to computational chemists for QMMM analysis to explain the relatively higher activity on glucosides, compared to mannosides.

6.2. Characterization of GH3 enzymes from rice.

Throughout the year, the OsExoII exoglucanase was expressed as a thioredoxin fusion protein in *E. coli*, purification optimized, and the protein characterized. The protein purification that gave the best results was immobilized metal affinity chromatography on nickel-equilibrated resin, followed by phenyl sepharose and S200 gel filtration chromatography. The obtained protein was characterized for pH optimum in citrate-phosphate universal buffers, temperature optimum, relative activities toward 4-nitrophenyl (4NP) glycosides at 1 mM concentration, and kinetics for hydrolysis of 4NP- β -D-glucopyranoside (4NPGlc), cellobiose and laminaribiose. The OsExoI isoenzyme was also expressed, but in *Pichia pastoris*, and was purified by IMAC and characterized for relative activities toward 4NP glycosides, oligosaccharides and polysaccharides, similar to OsExoII.

For GH3 β -xylosidase, although we were able to express the OsXyl1 β -xylosidase in various *E. coli* strains from pET32a, pSY7 and pSY9 vectors, as reported in our previous work (BRG5680012), optimization in these strains gave relatively low levels of expression. So, the cDNA was cloned into pPICZ α BNH8 and expressed in *Pichia pastoris* SMD1168H. The protein was purified by IMAC and the purified protein deglycosylated with endoglycosidase H to show that the deglycosylated protein was essentially a single band

on SDS-PAGE. The enzyme was tested for hydrolysis of various 4NP-glycosides and kinetics for 4NP- β -D-xylopyranoside determined.

We had previously expressed and characterized mutations of barley HvExoI, according to our previously described method (Luang et al., 2010a). Some of these mutations were crystallized (Luang et al., 2010b) and structures solved before and after soaking with glucose. Those structures were used as the basis for QMMM calculations of the catalytic and glucose exit mechanism, in collaboration with several computational chemists (Streletsov et al., 2019). Meanwhile, we also made site-directed mutations of HvExoI to make it mutate active site residues to those of GH3 β -xylosidase. Relative activities of some of those mutations toward 4NPGlc and 4-NP- β -D-xylopyranoside were determined.

6.3. Structural studies and engineering of TxGH116 β -glucosidase

For TxGH116, we have produced many datasets of the TxGH116 enzyme and its mutants, including the D593A and D593N acid/base mutants and E441A nucleophile mutant complexed with oligosaccharides, the refinements of which were continued or completed. In addition, we completed the set of known human pathogenic mutations (Votsi et al., 2014) by site-directed mutagenesis of the TxGH116 β -glucosidase and characterized the kinetics of 4NPGlc hydrolysis and temperature optima. We also mutated the active site cleft residue Y69 to check its influence of substrate binding and hydrolysis. The TxGH116 D508H (human mutation replication), D508N, R786H (human mutation replication), and datasets were collected for D508N and R786H alone and after soaking with 2,4-dinitrophenyl 2-deoxy-2-fluoroglucoside (G2F), along with additional crystals of catalytic acid/base and nucleophile mutants in complexes with cello- and laminari-oligosaccharides. Additional mutations were made to exchange the lengths of the catalytic nucleophile and catalytic acid/base (E441D, D593E and E441D/D593E to evaluate their effects on catalysis and substrate specificity.

To evaluate the unusual acid/base position, the pH profiles of the kinetic parameters of the wild type and D593N and D593A mutants were determined with 4NPGlc. Evaluation was also made of the pH optimum with cellobiose, a substrate with greater requirement for acid assistance during hydrolysis.

In order to assess the unusual protonation trajectory of TxGH116, we refined the structures of TxGH116 D593N in complexes with cellobiose and laminaribiose to completeness. These structures were then sent to Prof. Carme Rovira at the University of Barcelona, whose student ran molecular dynamics simulations to show that the horizontal protonation hypothesized from the structures is consistent with the energetics of the system. We subsequently solved the structure of complexes of the TxGH116 D593A mutant with cellobiose and laminaribiose as well, to show that the glycosidic bond position was not affected by hydrogen bonding to the D593 residue.

We further evaluated the use of acid/base (D593A, D593N) “glycothioligase” mutants and nucleophile (E441A, E441G, E441D) for their rescue by and transglycosylation of various nucleophiles. The use of the most efficient of these mutants for synthesis of beta- and α -glucosyl azides was then assessed, as a way to produce precursor for click synthesis of putative inhibitors. Fifteen β -glucosyl-triazole inhibitors were then produced from the β -glucosyl azide and a smaller set of α -glucosyl triazoles was also synthesized. The β -glucosyl-triazole inhibitors were then assessed for their inhibition of purified TxGH116 and TeGH116, as well as human β -glucosidase activity in crude cell extracts of Cos7 cells overexpressing human GBA1 and GBA2 (in collaboration with Dr. Lukana Ngiwsara of the Biochemistry Laboratory of Chulabhorn Research Institute).

6.4. Expression of new GH116 enzymes

Initial work was done on animal GH116 enzyme expression in *E. coli*. Optimized cDNA encoding human GBA2 and its deep sea vent tube worm homologue (supposed thermophilic GBA2) were obtained and cloned into a number of vectors, including pET30, pET32, pSY5, pSY7, pSY9, and pFATT2 and expression conducted in *E. coli* strains BL21(DE3) and Origami B(DE3) at low temperature (20° C) and 37° C, with induction with isopropyl-thio-beta-galactoside (IPTG) at concentrations ranging from 0 to 0.5 mM, as well as production in autoinduction media for some constructs. The cells were broken open and activity against 5 or 10 mM 4NPGlc tested. When activity was detected, the extract was purified by IMAC, although no active protein was recovered. SDS-PAGE was used to assess protein in the soluble and insoluble fractions. Insoluble thioredoxin-wormGBA2 fusion protein from expression from pET32 in BL21(DE3) was solubilized in 8 M urea and purified by IMAC in the solution, followed by elution in 250 mM imidazole in 8 M urea, 20 mM Tris, pH 8.0, 150 mM NaCl and dilution into buffer without urea. One rice GH116 gene was also obtained, but no protein could be expressed from this gene.

To further investigate the role of GH116 enzymes, a gene was obtained for a GH116 from a cyanobacterium, *Thermosynechococcus elongatus*, as a possible model of GH116 action in the chloroplast. This gene was cloned into pET32a and initial expression attempted in LB media at 20 °C and 37 °C. We then purified the protein and determined its pH and temperature optima. Furthermore, the relative activity against 4-nitrophenyl-glycosides and oligosaccharides were determined. The kinetic parameters for hydrolysis of 4NPGlc were determined along with the competitive inhibition constant of glucose. Furthermore, we assessed the inhibition of this enzyme by the inhibitors mentioned above and determined the IC₅₀ for inhibition of hydrolysis of 1 mM pNPGlc for the best of these. The TeGH116 was also submitted to crystallization trials and crystals obtained under various conditions. Optimization was done to obtain crystals of sufficient quality for diffraction, which were finally obtained by Ratana Charoenwattanasatien during a visit to Prof. Genji Kurisu's laboratory. The crystals were used for x-ray diffraction at the SPring-8 and NSRRC synchrotrons and the best dataset obtained with a nominal resolution of 2.3 to 2.5 Å.

7. Results:

7.1. Structural analysis of Os4BGlu18.

Os4BGlu18 expressed in *E. coli* has been characterized to show high activity on monolignol glucosides (Baiya et al., 2014), suggesting it could be a useful structural model for these proteins. The protein with the N-terminal thioredoxin and His-tag removed was purified and the pure protein could be crystallized in the condition of 25% polyethylene glycol (PEG) 3350, Tris-HCl, pH 8.5. The optimized crystals could diffract x-rays and showed unit cells in the $P2_12_12_1$ space group with unit cell parameters of $a = 163.8$, $b = 167.4$ and $c = 240.1$ Å. A crystal produced in the presence of G2F was found with a similar unit cell size in the $P4_22_12$ space group. Further optimization gave crystals in MES, pH 6.0 in PEG or PEG5000 MME with the unit cell parameters of $a=52.4$, $b=84.8$, and $c=207.9$ in the $P2_12_12_1$ space group and crystallization of the wild type and E194Q mutant alone and in complex with the inhibitor and possible substrates, products (monolignol alcohols) and inhibitors gave data sets with nominal resolution in the 1.8 to 2.5 Å range. However, we could not identify electron density for the coniferin, coniferyl alcohol, 4-coumaryl alcohol glucoside, 4-coumaryl alcohol or syringin (substrates and aglycon products of hydrolysis) in the active site of structures calculated from these data sets. One dataset of native enzyme was refined to near completeness at a resolution of 1.50 Å, R value of 17% and R_{free} of 20%. As shown in Figure 1, this structure had two protein molecules packed against one another with their active sites facing nearly opposite directions in the asymmetric unit, although gel filtration chromatography indicated the protein is predominantly a monomer in solution. The active site contained a Tris molecule from the buffer.

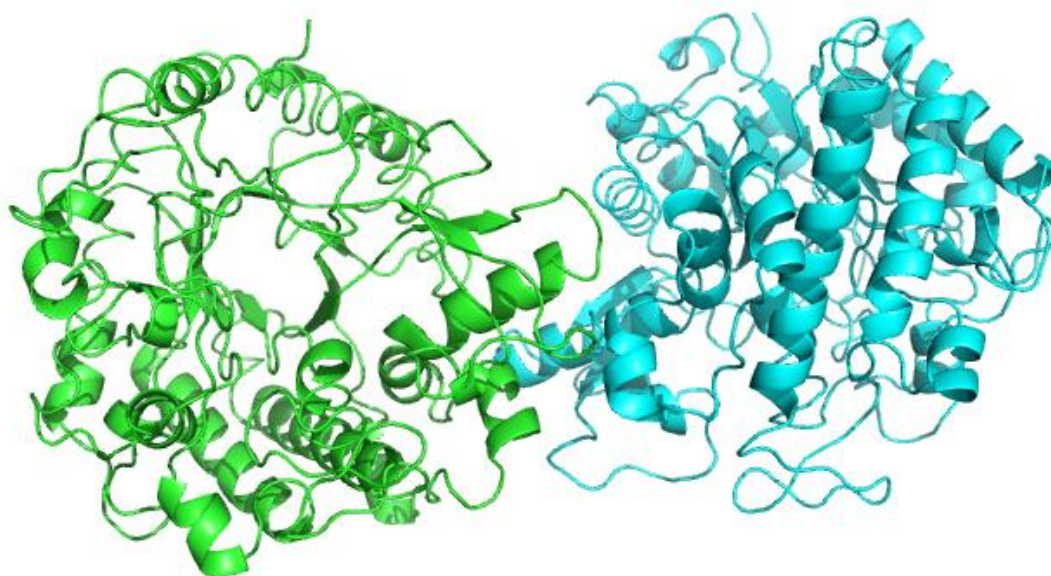


Figure 1: Two molecules of Os4BGlu18 found in the crystallographic asymmetric unit. The typical GH1 (β/α)₈ barrel structures are shown in cartoon representation colored by molecule.

We also mutated the catalytic acid base to make the Os4BGlu18 E194Q mutant and crystallized it, then soaked in monolignol glucosides. Unfortunately, none of the complexes with monolignols had sufficient density for the expected ligands in the active site when the diffraction data was evaluated. Nonetheless, we have completed the refinement of the native enzyme (which has a Tris molecule in the active site), along with complexes with 2,4-

dinitropheyl 2-deoxy-2-fluoro-glucoside and δ -gluconolactone in their active sites. The structures of coniferin, syringin and 4-hydroxy-coumaryl alcohol were docked into the active site. In order to get the sugar to go into the right position, it has to be distorted into a 3S_1 skew boat, as previously seen for Os3BGlu7 (rice BGlu1, Chuenchor et al., 2008). The structure of Os4BGlu18 with coniferin docked in the active site is shown in Figure 2.

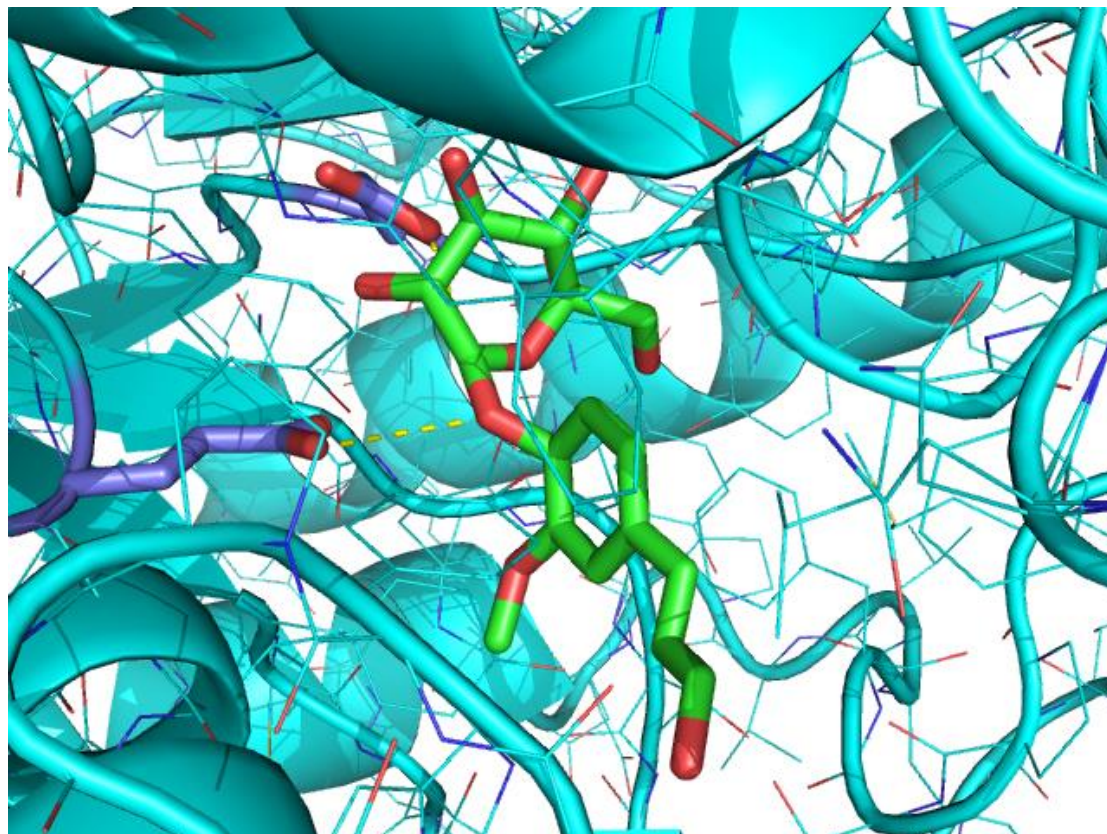


Figure 2: Coniferin docked into the active site of the Os4BGlu18 X-ray crystal structure. The coniferin (with carbons in green) and catalytic residues (with carbons in darker blue) are shown in stick representation, while the other residues are shown in lines superimposed with a cartoon of the backbone.

7.2. Demonstration of Os4BGlu14, Os4BGlu16 and Os4BGlu18 monolignol β -glucosidase activity *in planta*.

Although we previously showed the Os4BGlu18 expressed in *E. coli* had high activity toward monolignol glucosides (Baiya et al., 2014), this activity has not been demonstrated *in planta*. Therefore, we produced plasmids to produce Os4BGlu14, Os4BGlu16 and Os4BGlu18 in plants under control of the TMV 35S promoter and our collaborators in Prof. Jong-Seong Jeon's laboratory, Kyung Hee University, Suwon, Rep. Korea, transformed them into *Arabidopsis* plants that were mutated to lack BGLU45 and have shown elevated coniferin levels. We analyzed the monolignol and monolignol glucoside levels in the above ground part of the plants by extraction and LCMSMS at our laboratory. As shown in Figure 3, the levels of the monolignol glucosides coniferin, 4-coumaryl alcohol glucoside and syringin were all found to be elevated in the mutant plants in our MRM mode LCMSMS analysis, and expression of any of the three putative rice monolignol β -glucosidases lowered

those levels to near those of the wild type *Arabidopsis* plants. This result demonstrated that all three proteins can breakdown the monolignol glucosides in the plant.

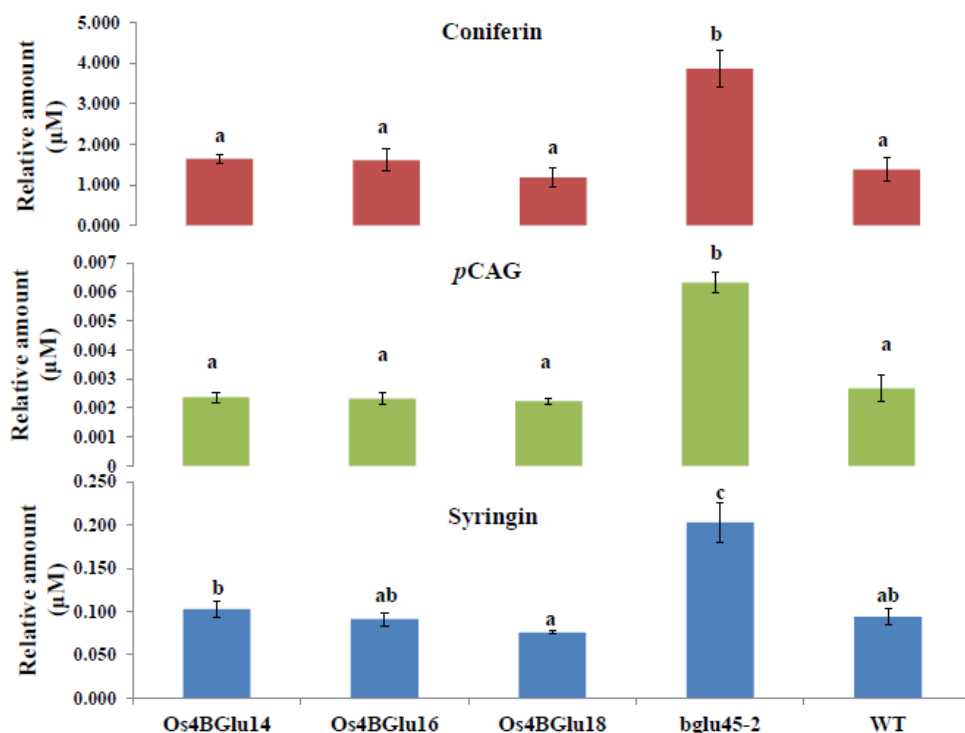


Figure 3. Levels of monolignol glucosides in *Arabidopsis* plants mutant for BGLU45 with and without overexpression of the rice monolignol β -glucosidase homologues Os4BGlu14, Os4BGlu16, and Os4BGLU18, compared to those in wild type plants (Baiya et al., 2018).

Since the monolignol glucosides are thought to primarily reside in the vacuole, a reviewer asked us to check the localization of the rice monolignol β -glucosidases. When we expressed a C-terminally GFP-tagged Os4BGlu18 in tobacco epithelial cells by *Agrobacterium* infiltration mediated gene transfer, the GFP could be seen to partition between the cell wall and plasma membrane in the apoplast (Figure 4), demonstrating the extracellular localization. This result suggests that monolignol glucosides may be sent to the extracellular space, rather than always remaining in the vacuole, or that the extracellular enzymes may sometimes be taken into the vacuole.

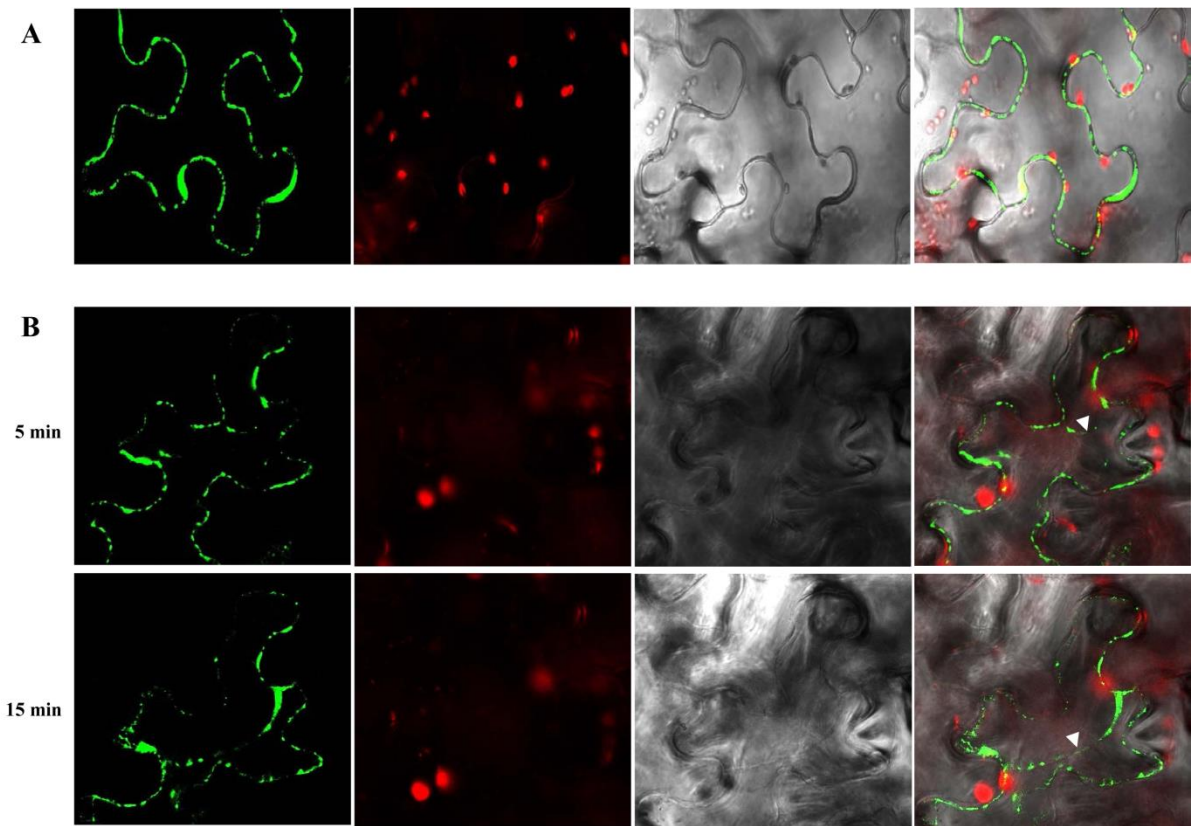


Figure 4. Localization of GFP-tagged Os4BGlu18 is observed at the cell wall and plasma membrane. A. The localization seen by confocal microscopy 3 days after transformation of tobacco leaf epithelial cells with Os4BGlu18-GFP expression plasmid (frame 1), the autofluorescence of chloroplasts (frame 2), phase contrast image of the leaf epithelium (frame 3) and superposition of the three images (frame 4). B. Localization 5 and 15 min after addition of 0.5 M mannitol osmolyte to shrink the cells and pull the plasma membrane away from the cell walls. The frames are as described in A. Figure from Baiya et al., 2018.

7.3. Functional and structural analysis of Os9BGlu31 transglucosidase mutants.

We previously characterized Os9BGlu31 as a transglucosidase with highest activity on glucose esters of phenolic acids, including some phytohormones (Luang et al., 2013). We had further showed that mutations of the active site cleft did not have a large affect on the hydrolysis vs. transglycosylation ratio, but one mutation at residue 243, W243N, showed significantly higher activity and promiscuity than the wild type enzyme (Komvongsa et al., 2015a). Here, we completed the set of 20 amino acid types at residue 243 and combined the W243N mutant with other mutants of the active site that had been expected to favour water binding for hydrolysis (Tran et al., 2019). We found that the W243L mutant had high activity, similar to W243N, but with somewhat different specificity, so that it was best glycosylation of certain substrates (Figure 5). We also found that the destabilizing mutation L241D, which was not previously characterized due to its poor stability and solubility and low activity, could be complemented in the presence of W243N, such that the L241D/W243N mutant had even higher activity than W243N on several substrates. This sign anisotropy confirmed the close interaction of these neighboring residues in the active site cleft and generated a new useful tool for glycosylation of certain substrates. We also found that the double mutants tended to have higher hydrolysis to transglycosylation ratios than the wild type and single mutants, supporting the idea that

making the active site more polar may support hydrolysis while a less polar active site may promote transglycosylation.

Previous attempts to crystallize Os9BGlu31 and Os9BGlu31 W243N did not produce useful crystals, so we used homology models to try to explain the substrate specificity of Os9BGlu31. Docking into these structures did not give reasonable complexes for the acceptors in the glucosyl-enzyme intermediate, so we instead resorted to machine learning algorithms to correlate substrate and enzyme properties (Tran et al., 2019). These programs could indeed correlate several factors for transglycosylation, although the derived correlations were not obvious to implement for enzyme engineering. A further attempt to crystallize Os9BGlu31 succeeded, and structure solution is underway, in the hope that it may lead to a clearer explanation of the basis for transglycosylation vs. hydrolysis and allow for engineering of enzymes for glycosylation of substrates of interest.

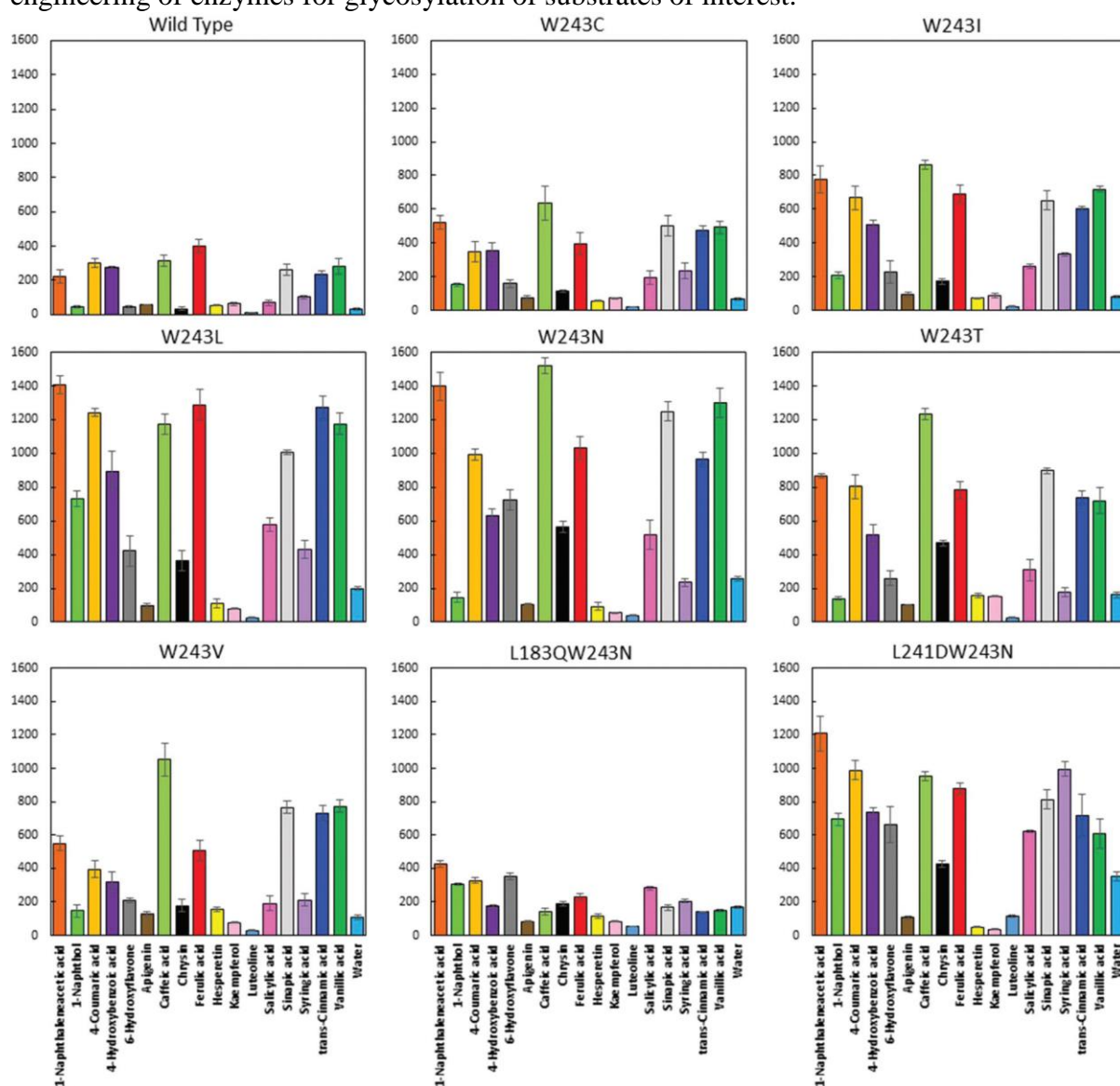


Figure 5. Activity of highly active mutants of Os9BGlu31 on selected receptor substrates relative to wildtype. 4NPGlc was used as the donor substrate and activity quantified as 4NP released. Note that L243D/W243N showed the highest activity for hydrolysis and transglycosylation of Syringic acid and salicylic acid.

7.4. Characterization of GH3 enzymes

Improved expression of OsExoII in *E. coli* was found upon induction with autoinduction media at 15 °C, although the yields were still low. This allowed purification to apparent homogeneity of SDS-PAGE (Figure 6) via one nickel IMAC step, followed by phenyl sepharose and S200 gel filtration chromatography, although results of this purification procedure with still variable. The purified protein had a pH optimum of 5.0, and kinetic parameters at 30° C and pH 5.0 were determined to be $K_M = 0.22$ mM, $k_{cat} = 0.64$ s⁻¹, and $k_{cat}/K_M = 2.86$ for 4NPGlc, $K_M = 0.015$, $k_{cat} = 0.38$ s⁻¹, and $k_{cat}/K_M = 24.6$ mM⁻¹s⁻¹ for cellobiose, and $K_M = 0.39$, $k_{cat} = 0.55$ s⁻¹, and $k_{cat}/K_M = 13.9$ mM⁻¹s⁻¹ for laminaribiose. Several compounds, including glucose, were also found to inhibit the enzyme. Although the enzyme had a temperature optimum of 40 °C in a 15 min assay, it was only stable at temperatures up to 30 °C (when stability was tested at 10 degree intervals).

In addition to OsExoII, OsExoI and OsExoIII genes for *P. pastoris* expression were cloned into pPICZα and transformed into *P. pastoris* SMD1168H. Screening of initial clones indicated relatively high activity only in the OsExoI clones. So, the best OsExoI expressing clone was grown the protein purified from the media by IMAC and preliminary characterization done. The protein hydrolyzed similar substrates as OsExoII, as shown in Table 1. Similarly, it had pH and temperature optima of pH 5.0 and 50 °C, but was only stable up to 30 °C. Deglycosylation with Endo H gave a single band, instead of the disperse protein band without digestion (Figure 7), but this deglycosylation had no effect on temperature stability.

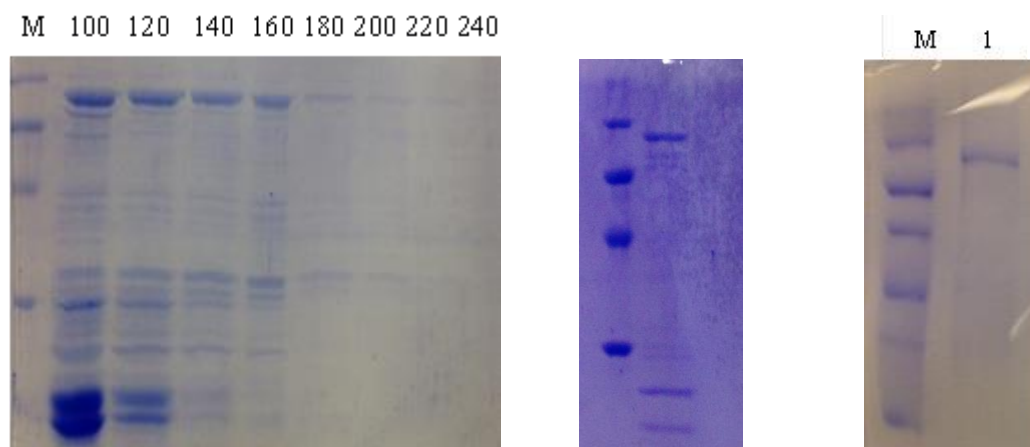


Figure 6. OsExoII protein after purification by IMAC, hydrophobic interaction, and gel filtration chromatography. Left gel, IMAC fractions eluted with the indicated mM concentrations of imidazole values noted at the top; middle gel, marker and pool after phenyl sepharose chromatography; right gel, marker and pool after S200 gel filtration chromatography.

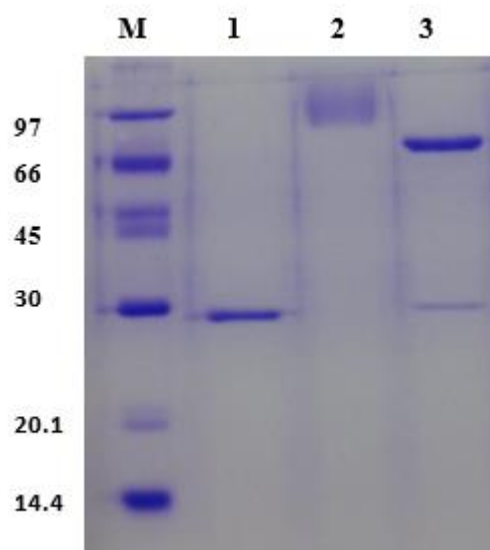


Figure 7 SDS-PAGE analysis of OsExoI expressed in *Pichia pastoris*. Lane M, standard protein marker; Lane 1, endoglycosidase H; Lane 2, OsExo1 protein before deglycosylation; Lane 4, OsExo1 protein after deglycosylation with endoglycosidase H.

For barley ExoI (HvExoI), we continued to characterize the mutation that could be made in the active site following our successful expression in *P. pastoris* (Luang et al., 2010). The active site residues involved in glucose binding and exchange in the active site were mutated and the mutant enzyme found to have little or no activity. Structures of some of the enzymes showed a loss of glucose binding in the active site. Quantum Mechanics/Molecular Mechanics simulations showed that the enzyme could release glucose through a transiently opened channel in the side of the active site, and suggested that coordination of this exit with binding of the next substrate residue could lead to progressive catalysis in the pocket-shaped active site of GH3 enzymes (Steltsov et al., 2019).

Table 1: Comparison of relative activities of rice and barley exonucleases toward various substrates.

Substrate	Relative Activity (%)			
	OsExoI	OsExoII	HvExoI ^a	HvExoII ^a
<i>4-nitrophenyl glycosides</i>				
4NP-β-D-glucopyranoside	100	100	10	56
4NP-β-D-galactopyranoside	NA	2.3	NA	NA
4NP-α-L-arabinopyranoside	NA	2.0	NA	NA
4NP-β-D-xylopyranoside	6.1	1.9	NA	NA
4NP-β-D-fucopyranoside	NA	1.7	NA	NA
4NP-β-D-cellobioside	44.3	NM	NA	NA
<i>Oligosaccharides</i>				
Laminaribiose	100	100	70	75
Laminaritriose	97.2	98.3	NM	NM
Laminaritetraose	96.0	97.1	NM	NM
Laminaripentaose	97.0	98.3	NM	NM
Cellobiose	95.4	91.2	14	20
Cellotriose	93.1	83.8	NM	NM
Cellotetraose	73.7	57.4	NM	NM
Cellopentaose	71.5	69.4	NM	NM
<i>Polysaccharides</i>				
Laminarin (<i>L. digitata</i>)	100	100	100	100
Barley(1,3;1,4)-β-D-glucan	63.2	65.2	10	14
Lichenan (<i>C. islandica</i>)	70.4	71.0	18	49

^aThese data are from Hrmova and Fincher, 1998.

NA means “No activity detected”; NM means “Not measured.”

Although the GH3 β-xylosidase/α-L-arabinosidase OsXyl1 was also expressed in *E. coli*, expression levels were low and it was not purified. Instead, the cDNA was cloned into the *Pichia pastoris* expression vector pPICZaBNH8 (Toonkool et al., 2006) and expressed in *Pichia pastoris* strain SMD1168H. When media from the best clone was used to purify enzyme by immobilized metal affinity chromatography, the protein could be seen as a high molecular weight smear, which reduced to a single major band after digestion with endoglycosidase H, which cleaves high-mannose glycan of glycoproteins after the first (asparagine-linked) N-acetylglucosamine residue (Figure 8). Assessment of the pH and temperature optima indicated that the enzyme has a rather low pH optimum of 4.0 and a high temperature optimum of 60 °C in a 30 min assay at pH 4.0. When the clones were reassessed under these conditions, a second clone was found to have slightly higher expression than the initially identified clone and was used for further analysis. The protein was purified and tested on different synthetic glycosides, as shown in Table 2, which showed that it had highest activity on 4NP-β-D-xylopyranoside, with 8.5% of that maximal activity level with α-L-arabinopyranoside and 5.9% of that activity with β-D-mannopyranoside.

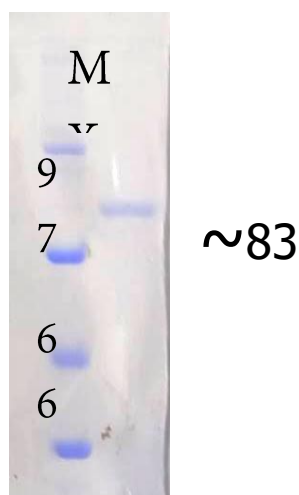


Figure 8 SDS-PAGE analysis of OsXyl1 produced in *P. pastoris* after purification and digestion with Endoglycosidase H.

Table 2: Relative activity of OsXyl1 on various nitrophenyl glycosides

Substrate	Relative activity (Percent of activity on 4NPXyl)
4-NP β -D-xylopyranoside	100
4-NP α -L-arabinopyranoside	8.5
4-NP β -D-mannopyranoside	5.9
2-NP β -D-glucopyranoside	3.3
4-NP β -D-glucopyranoside	2.6
4-NP α -D-galactopyranoside	1.5
4-NP β -D-maltoside	1.3
4-NP α -L-fucopyranoside	1.3
4-NP N-acetyl- β -D-glucosaminide	1.2
4-NP β -D-galactopyranoside	1.2
4-NP β -D-fucopyranoside	1.1
2-NP β -D-galactopyranoside	0.9
4-NP α -D-glucopyranoside	0.6
p-NP β -D-cellobioside	0.7
p-NP α -D-mannopyranoside	0.7

7.5. Characterization of TxGH116 mutants.

During the first year, we collected data and refined structures for mutants and of the thermophilic family GH116 enzyme TxGH116, the structure of which we previously used to explain pathogenic mutations in the human GH116 member GBA2 (Charoenwattanasatien et al., 2016). When investigating the substrate binding cleft, it was noted that the residue Y62 was in a position to interact with oligosaccharides in the active site and its position seemed to depend on the presence or absence of such substrates. To see

if it was important for substrate binding, it was mutated to alanine (Y62A), and kinetic parameters for hydrolysis of 4NPGlc (Table 1). Although it had a minor effect on 4NPGlc hydrolysis, it is anticipated that this position might affect oligosaccharides more, but it was found to have relatively minor effects on these as well

Aside from this, 3 new human mutations were identified after we had done the mutagenenic analysis of their effects in TxGH116 for the previous paper. So, we made the corresponding mutations for two of them, F347V and G599R, in the TxGH116 protein. The third new mutation was a nonsense mutation, which would lack the catalytic domain and therefore not have activity. The F347V and G599R mutants had little or no effect on kinetics compared to the wild type. In fact, this was expected, since these mutations are not in the active site and were expected to affect the stability of the enzyme, rather than its kinetic parameters. In the past, we have found that such mutations, e.g. R544W, had less effect on the TxGH116 protein than human GBA2, due to the greater intrinsic stability of TxGH116.

Since the lower Michaelis constant (K_M) of the R786H mutant compared to wild type TxGH116 suggested it might bind to glucose less tightly than wild type, we compared the competitive inhibition constants (K_i) of TxGH116 to its R786H and Y62A mutants. Unexpectedly, the competitive inhibition constants, corresponding to the apparent dissociation constants (K_D) for glucose for both R786H (1.48 mM) and Y62A (1.08 mM) were both somewhat lower than wild type (3.95 mM), so that these mutants were not the candidates for decreasing glucose inhibition for which we hoped. Therefore, we produced mutations of all active site residues to see their effects on substrate hydrolysis and glucose inhibition, and the mutations were confirmed. Several proteins have been produced and their kinetic constants for 4NPGlc determined, as shown in Table 4. Since many had very low activity, we could not determine the glucose inhibition for most of them, but we did so for the high activity mutants, none of which showed greatly higher competitive K_i , indicating reduced glucose inhibition. For instance, another mutation at the R786 residue, R786A, did show a slightly improved competitive K_i for glucose (6.5 mM), and seemed considerably more efficient at hydrolyzing cellobiose and cellotriose.

Table 3 Kinetics of 4NPGlc hydrolysis by TxGH116 mutants.

TxGH116 variant	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M ($mM^{-1}s^{-1}$)	Temperature optimum ($^{\circ}C$)
Wild type*	0.18 ± 0.008	49.0 ± 0.8	272	75
Y62A	0.29 ± 0.006	56.0 ± 0.3	193	N.D.
F347V	0.17 ± 0.007	49.0 ± 0.65	288	65
G599R	0.17 ± 0.008	42.3 ± 0.65	249	70-75
R544W*	0.26 ± 0.011	48.2 ± 0.5	186	60-65
R786H*	15.7 ± 0.5	194 ± 2.4	12.3	70

N.D. is not determined (no change in temperature optimum expected). *Note that the last two mutations and wild type were previously published in Charoenwattanasatien et al., 2006.

To understand the effects of the human mutations better, the TxGH116 mutants R544W and R786H were crystalized and their structures solved alone and with G2F covalent inhibitor. R544W had minor effects on the packing in the middle of the protein and no obvious effects in the active site, but Figure 9 shows that the R786H mutation (magenta) had two possible positions for the H786 side chain that were not similar to that of the R786 position and did not hydrogen bond to the glucosyl 6 hydroxyl group. The neighboring tyrosine, Y455, was also displaced in comparison to the wild type enzyme.

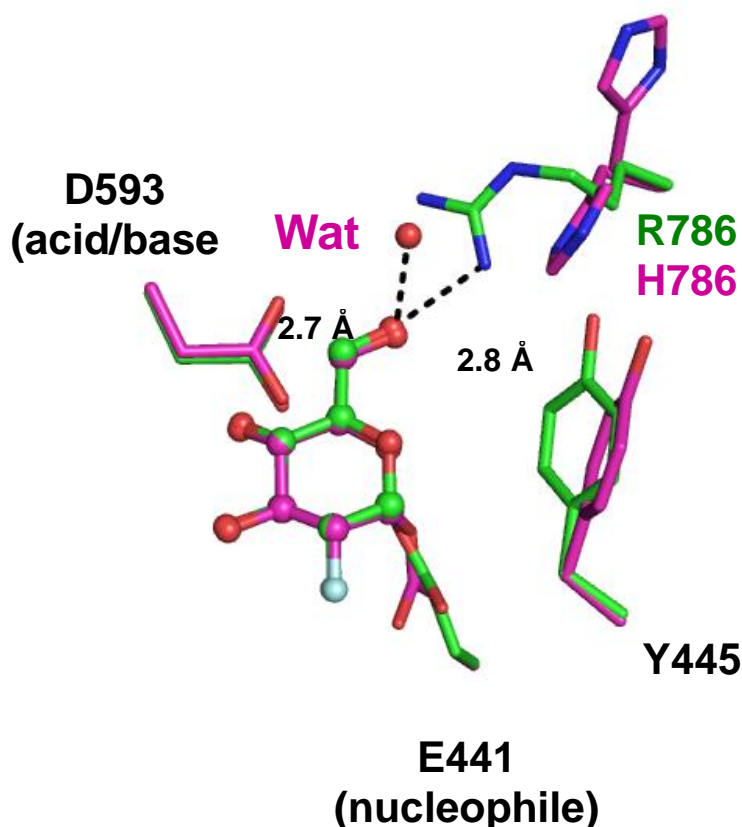


Figure 9. Structural superposition of the active site of TxGH116 in complex with G2F.(green) with TxGH116 R786H in complex with G2F (magenta). Note that the H786 residue is moved away from the position of R786 and the neighboring tyrosine residue, Y445, has also moved.

Additional work was done to further characterize the acid/base residue, due to its unusual position related to other retaining β -glucosidases (Heightman and Vasella, 1999, Charoenwattanasatien et al., 2016). Therefore, the pH dependence of the kinetic parameters were determined. The k_{cat}/K_M showed a sharp optimum at 5.0 and inflection points at pH 3.8 and 6.2 corresponding to the pKa of the catalytic nucleophile and acid/base respectively (Figure 10). When the D593N acid/base mutant was characterized in the same way, it maintained the peak at pH 5, but the basic part of the curve did not decrease and instead rose to a higher peak at pH 7.5 (Figure 11). It should be noted that this peak was still at a k_{cat}/K_M value about 3000-fold lower than the wild type, however. This was not quite as expected, since the activity against 1 mM 4NPGlc follows the expected plateau at values above pH 5 until 7.5 or so, but the kinetic parameters are affected by a higher K_M and k_{cat} near pH 7.5.

Since we noted that the acid/base mutants were relatively highly active in the production of beta-glucosylazide (1-azido- β -D-glucose), we conjectured that this reaction could be used to efficiently prepare substrate for click glycosylation reactions. Screening showed that the D593A and D593N mutants gave similar production of 1-azido- β -D-glucose, but the D593A mutant seemed slightly better for use. So, we generated medium scale reactions (up to 30 mL) and were able to produce and purify over 1 g of 1-azido- β -D-glucose in one optimized reaction. This substrate could then be applied to copper-catalyzed Click chemistry to prepare β -D-glucose-1,2,3-triazoles with a variety of functional groups, without risk of breakdown upon deprotection, since no deprotection was required (Gorantla et al., 2019).

Table 4: Effects of other active site residues mutations on TxGH116 catalytic properties.

Protein	$K_M(\text{mM})$	$k_{\text{cat}}(\text{s}^{-1})$	k_{cat}/K_M ($\text{mM}^{-1}\text{s}^{-1}$)	Optimum temperature (°C)	Optimum pH
WT	0.16	40.4	252.5	-	5.5
D452A	27.8	0.12	0.0043	-	5.0
D452N	8.81	0.0074	0.0008	-	4.5
H507A	10.5	0.28	0.027	70	5.0
H507E	9.9	0.01	0.0010	60	5.0
H507Q	47.3	14.1	0.30	60	5.0
T591A	12.8	3.65	0.29	-	4.5
E730A	3.79	80.8	21.3	-	4.0
W732F	0.47	62.35	132.7	-	5.0
E777A	29.7	0.12	0.0040	-	5.5
E777Q	7.0	0.02	0.0029	70	3.5
R786A	5.6	129.0	23.04	-	4.5
R786K	7.3	88.7	12.15	70	4.5
R792A	4.53	0.0083	0.0018	-	5.0

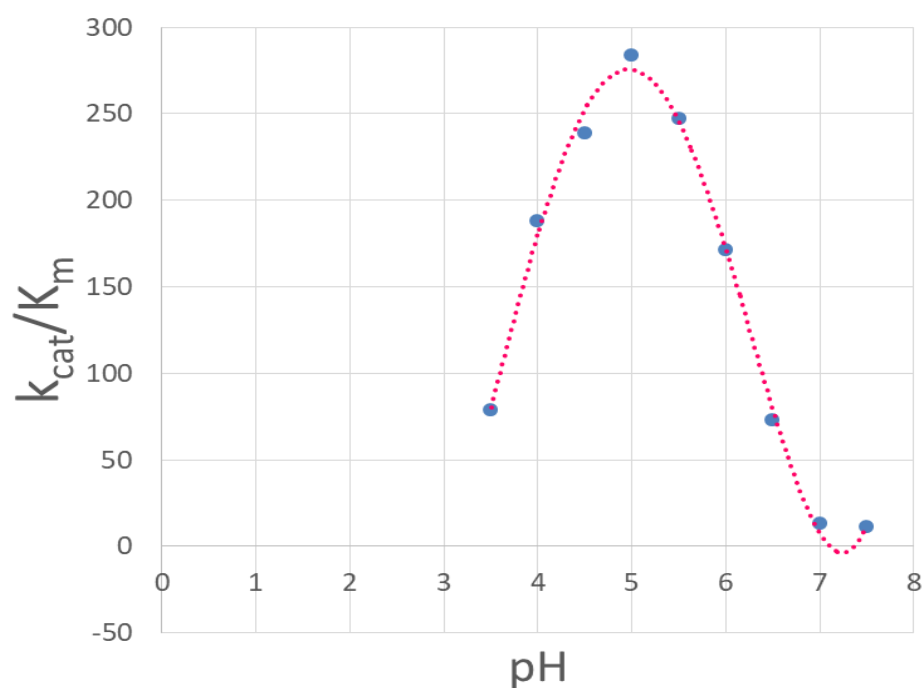


Figure 10. The k_{cat}/K_M dependence on the pH for wild type TxGH116.

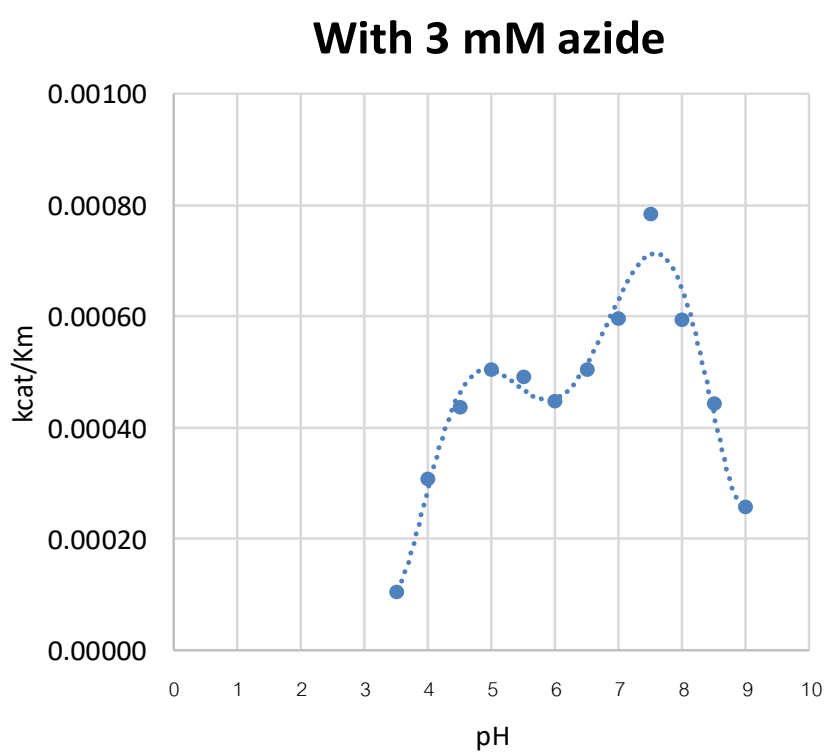
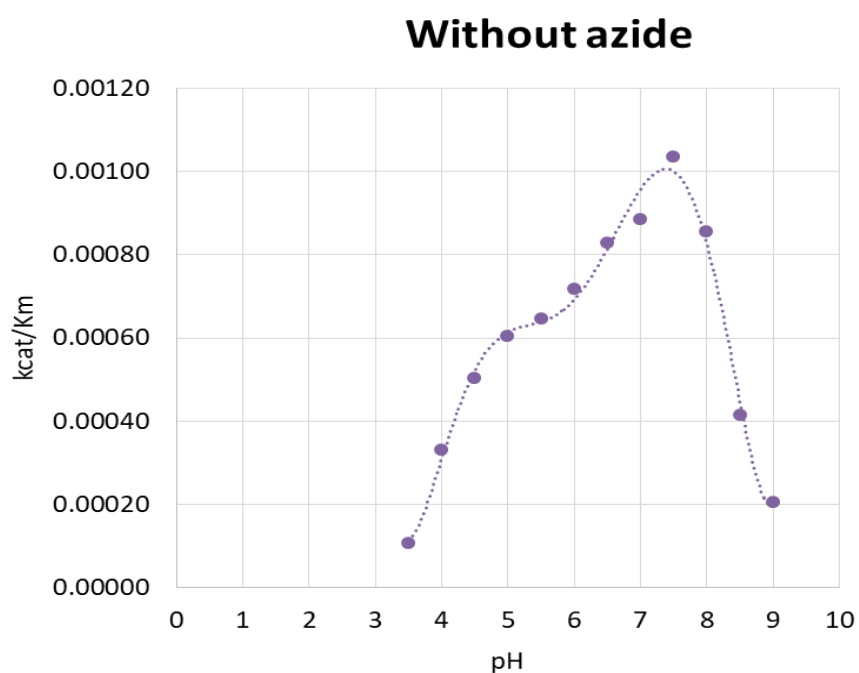


Figure 11 The pH dependence of the TxGH116 D593N mutant in the absence (top) and presence (bottom) of 3 mM azide. In the top condition, hydrolysis predominates, while in the lower condition, transglycosylation occurs.

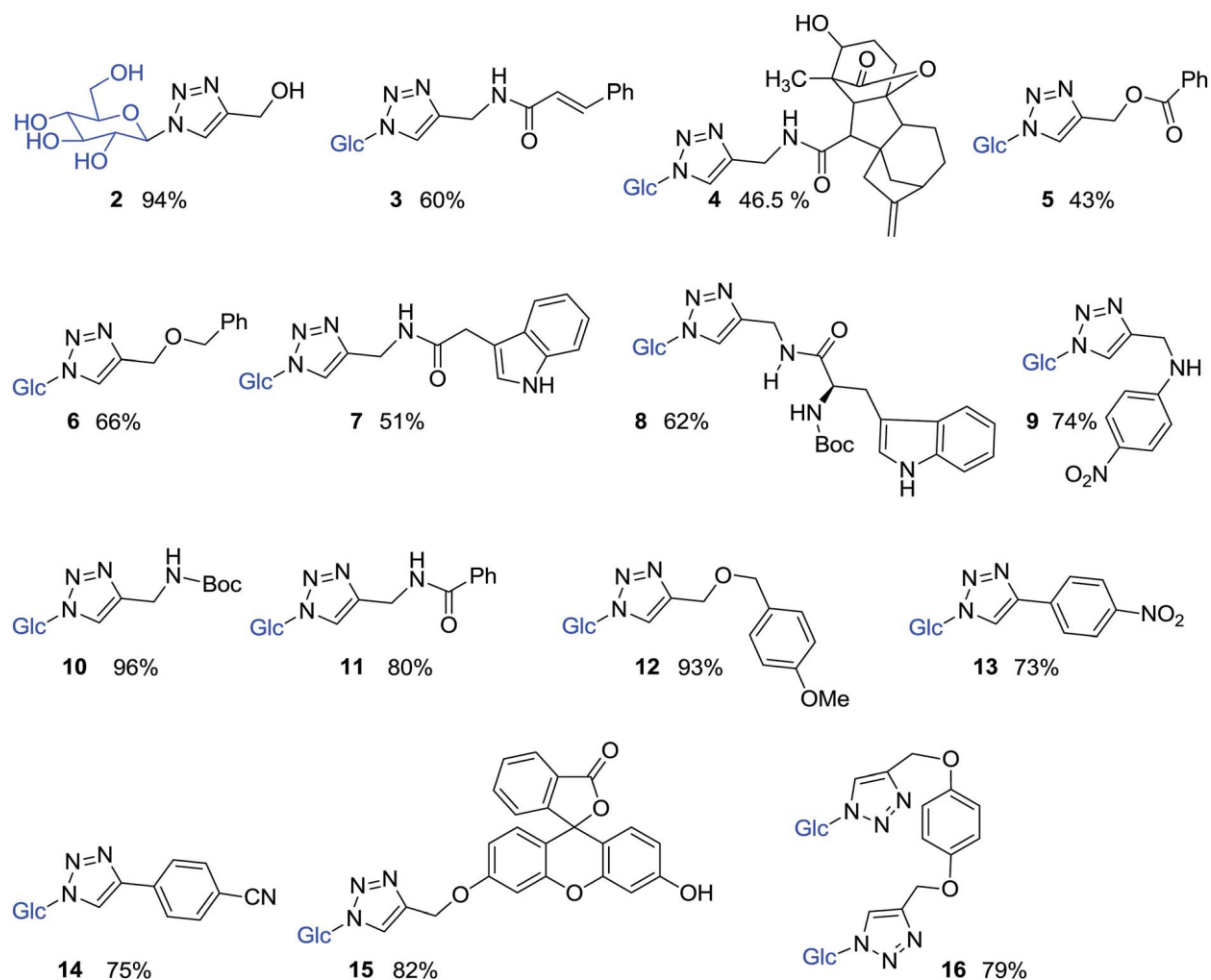


Figure 12 Structures of β -D-glucosyl-1,2,3-triazoles synthesized from enzymatically produced 1-azido- β -D-glucose. Percentages indicate the yields in the click reactions and purification. The compounds were designated GT2-16, according to the numbers shown, with 1-azido- β -D-glucose designated GT1. The figure is from Gorantla et al. (2019).

The β -D-glucosyl-1,2,3-triazoles are potential inhibitors of β -glucosidases. They were tested for inhibition of human β -glucosidases in crude cell lysates, but little inhibition was seen. However, several showed strong inhibition of TxGH116, as shown in Figure 12. The strongest inhibitor of TxGH116 was GT13, which had an IC_{50} value of 40 nM for TxGH116 hydrolysis of 1 mM 4NPGlc, while the second best was GT14, which had an IC_{50} value of 108 nM in the same assay (due to the substrate concentration of 5 times the K_i , these IC_{50} values should be approximately 6 times the K_i values, based on simple competitive inhibition). It seems that having another aromatic ring within 1 bond of the triazole was favorable for inhibition of TxGH116. These inhibitors were soaked into TxGH116 crystals and the data collected to obtain their structures, which are still in refinement.

Given that the TxGH116 acid/base mutants could efficiently be used to produce β -D-glucosyl-1,2,3-triazoles, we explored the use of the TxGH116 nucleophile mutants E441A and E441G for synthesis of 1-azido- α -D-glucose and its use to make α -D-glucosyl triazoles, which are potential inhibitors of α -amylase and α -glucosidases, and thereby may be potential drugs to lower the glycemic index of foods for diabetics. Although high concentrations of enzyme were needed and the yields were not as high, due to the low activity of the nucleophile mutant, 1-azido- α -D-glucose was still produced more efficiently than by

synthetic chemistry, and it was used to synthesize several α -D-glucosyl-1,2,3-triazoles, which will be tested for biological activity and enzyme inhibition in the future.

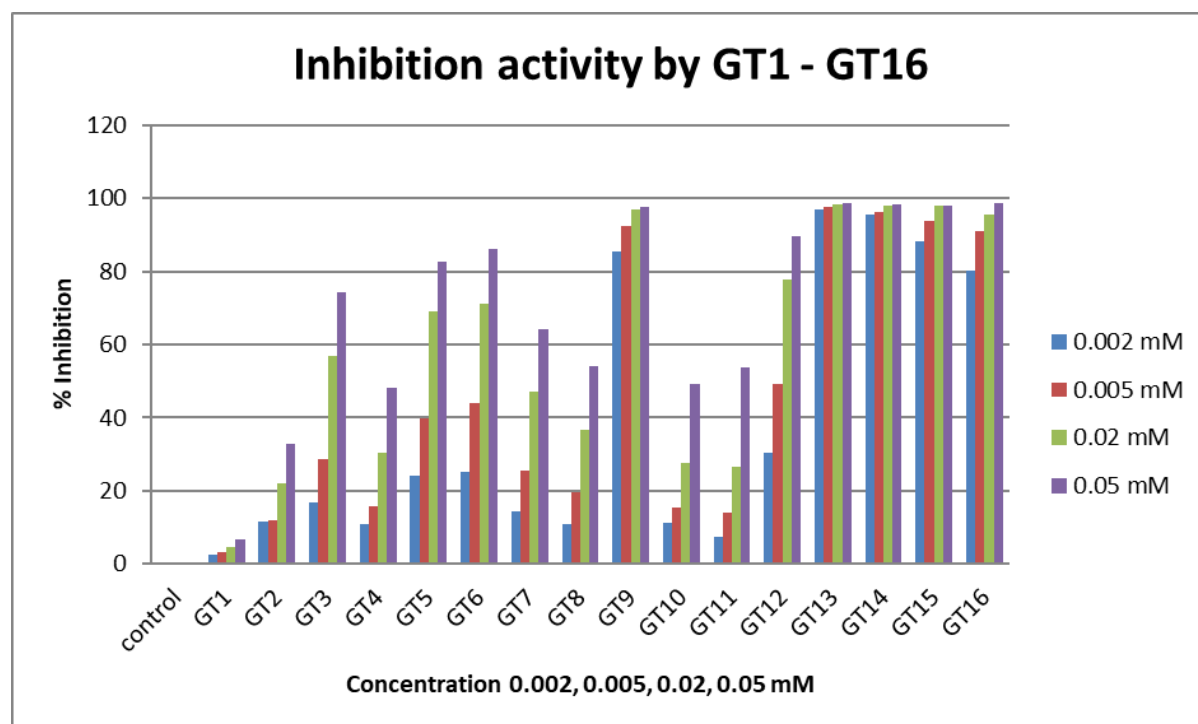


Figure 13 Percent inhibition of TxGH116 by different concentrations of β -D-glucosyl-1,2,3-triazoles.

7.5. Expression of other GH116 enzymes.

Although much work was done on expression of human GBA2 and a putative thermophilic GBA2 from a deep sea tubeworm (worm GBA2), only low and inconsistent β -glucosidase activities were detected in the soluble fractions or refolded human and worm GBA2. In another project, we could produce human GBA2 as a GFP fusion in tobacco leaves, as judged by fluorescence localization, but western blots gave poor results, suggesting degradation. We learned that others could produce active human GBA2 in baculovirus insect-cell expression systems, but the solubility led to lack of crystallization. Since other groups were ahead of us in GBA2 expression, we turned to characterizing other GH116 enzymes.

Due to the difficulty in expressing animal GH116 proteins, we tried to expression on GH116 protein from a cyanobacterium. Since the genus *Synechococcus* is closely related to plant chloroplasts, a gene from a thermophilic relative, *Thermosynechococcus elongatus* (TeGH116) was obtained and cloned into the pET32a expression vector for an initial attempt. This protein could be obtained as a thioredoxin fusion protein in this system and the initial test showed that it has β -glucosidase activity and can be purified by IMAC (Figure 14). The purified TeGH116 had a pH optimum of 5.5 and temperature optimum at 55 °C. We went on to test the TeGH116 enzyme with several substrates. The enzyme was quite specific to β -D-glucoside, when comparing 4-nitrophenolate substrates, with only 2.7% relative activity with β -D-galactopyranoside, 0.8% with β -D-N-acetylglucosaminide, 0.6% with β -D-mannopyranoside and less than 0.5% with other 4NP glycosides tested, compared to 4NPGlc. It also hydrolyzed β -1,3- and β -1,4- linked glucooligosaccharides, with highest activity

toward cellotriose. When screened with the β -glucosyl 1,2,3-triazoles produced above, TeGH116 was found to be less inhibited than TxGH116 with the best inhibition by the gibberellin-derived GT4.

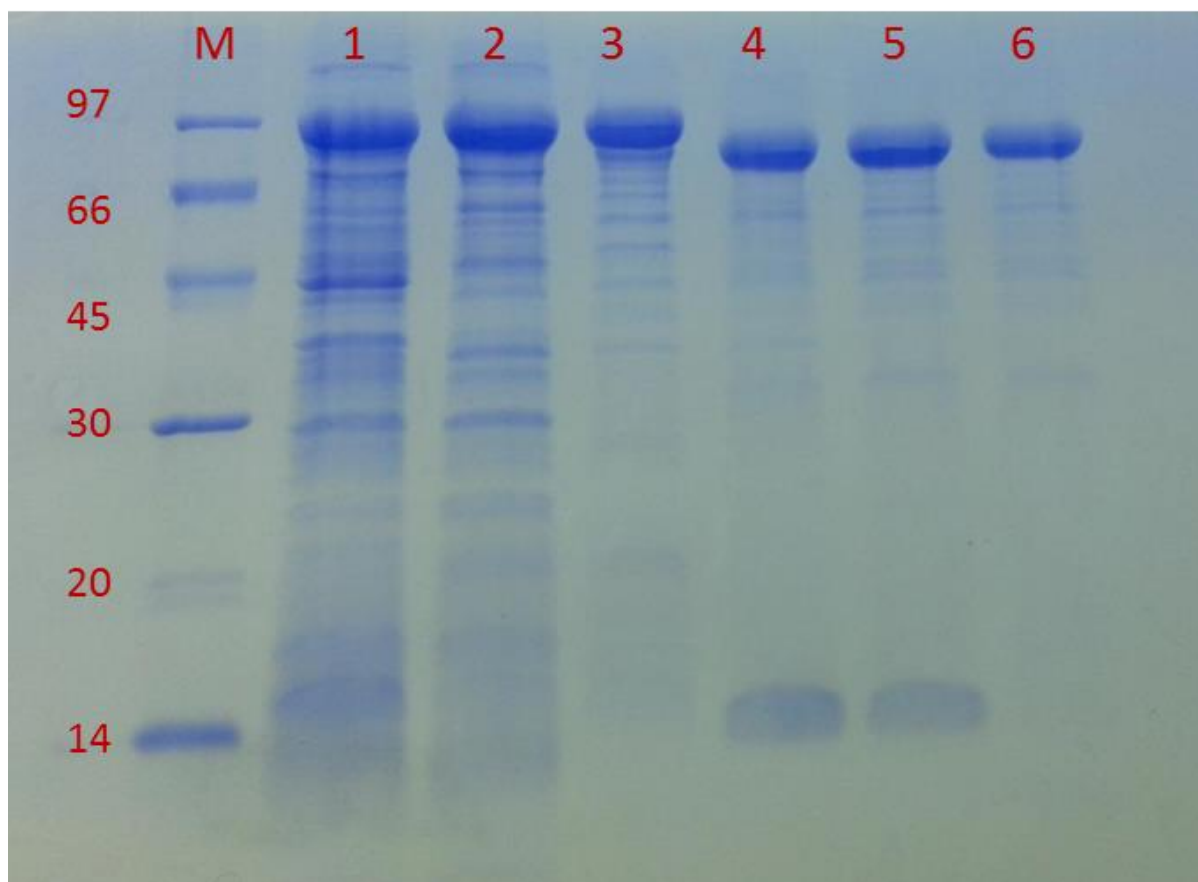


Figure 14 SDS-PAGE analysis of *Thermosynechococcus elongatus* TeGH116 purification after *E. coli* expression. Lane M, Molecular weight marker; lane 1, crude soluble extract of *E. coli* after expression; lane 2, crude soluble lysate after heating at 60 °C for 30 min; lane 3, IMAC purification fraction pool; lane 4, Enterokinase digest of 1 IMAC pool; lane 5, Pool after second round of IMAC; lane 6, Pool after S-200 gel filtration chromatography of 2nd IMAC pool.

Since the TeGH116 enzyme is the first GH116 enzyme to be characterized from a cyanobacterium and showed somewhat different properties from TxGH116, we crystallized the protein and it was found to crystalize in a variety of conditions. One set of crystals was able to diffract to reasonable resolution, and a dataset was collected to around 2.5 Å resolution. The data refined to the I2 spacegroup and the structure could be tentatively solved by molecular replacement with the TxGH116 structure. The TeGH116 structure is currently under refinement, although the residuals are still a bit high (R factor of 0.23 and R_{free} of 0.29). The current model for the TeGH116 structure is shown in Figure 15. More crystals have been generated and soaked with inhibitors and glucose to generate models for analysis of substrate interaction.

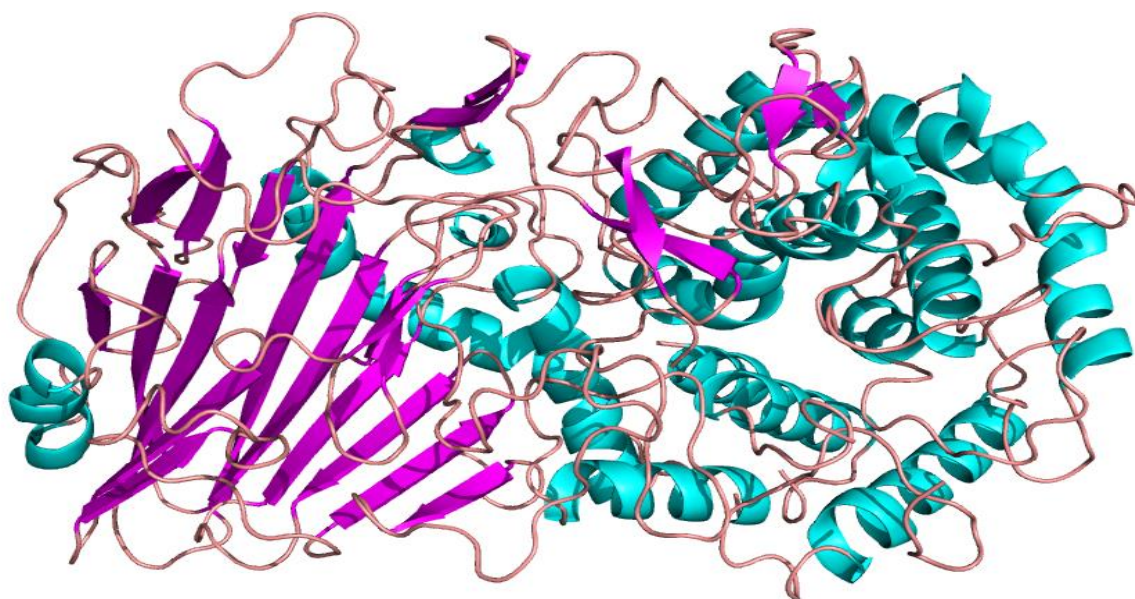


Figure 15 Structural model of TeGH116. The structure is shown as a cartoon depiction with alpha-helices in cyan, beta-strands in violet and loops and turns in salmon. Further refinement of the model is in process.

8. Conclusions and Benefits

8.1. Scientific Conclusions and Benefits

During this grant, we continued our work on structural and functional characterization of β -glucosidases, transglucosidases and β -xylosidases from GH families GH1, GH3 and GH116. This work gave us a better understanding of how β -glucosidases from different families function at the atomic and molecular levels and what differentiates enzymes with different substrate and product specificities. It also allowed us to produce new products using the enzymes as transglucosidases to add the glucose directly to the aglycones of interest or to produce azides that could then be added to molecules of interest by click chemistry.

For the GH1 enzymes, we were able to solve the structure of Os4BGlu18, which is the first example of a monolignol β -glucosidase with a known structure. This allowed us to dock monolignols into the structure to see what bonding interactions result in specificity for monolignols. Moreover, we were able to demonstrate that Os4BGlu18 and its near relatives Os4BGlu14 and Os4BGlu16 could really affect the levels of monolignol glucosides in plants, despite the fact that the enzymes and glucosides are thought to reside in different compartments in the plant. This shows the need for a more nuanced understanding of how these enzymes and their substrates are localized in the plant and a possible dynamic aspect of this localization.

We were also able to demonstrate that the Os12BGlu38 β -glucosidase is critical for pollen development in rice. However, we could only express a weakly active enzyme in a recombinant system with some activity to oligosaccharides and certain glycosides that suggests it could play a role in cell wall development, but was less conclusive than desired. Nonetheless, it shows a critical new role for a β -glucosidase in development.

Testing of different mutations of Os9BGlu31 was able to identify new highly active mutants of this transglucosidase which are better for production of certain glucosides than the previously identified mutants. We also found that double mutants that increased the active site hydrophilicity tend to have higher hydrolysis activity, suggesting the overall hydrophilicity/hydrophobicity of the active site does contribute to the transglycosylation vs. hydrolysis determination, although a single mutation cannot show this. Furthermore, combination of two mutations (L241D/W243N) was able to increase the activity of the enzyme, despite the fact that one of those mutations alone (L241D) resulted in much worse activity than wildtype. This points to the importance of considering amino acid interactions in the active site in engineering enzymes. We were also able by machine learning to identify factors that lead to the transglycosylation of certain molecules, which may be used in the future to predict which substrates may be glycosylated by our engineered enzymes, thus improving the use of these enzymes. However, due to the need for more precision in predicting the enzyme substrate interactions, we are currently solving the x-ray crystal structure of Os9BGlu31.

We were able to improve the production and characterization of rice GH3 exoglucanases and β -xylosidase, as well as finally complete and publish our characterization of glucose exit from the active site of barley HvExoI exoglucanase. This results in a better understanding of the difference between the glucosidase and xylosidase enzymes for the rice enzymes and developed a new understanding of how an enzyme with a pocket-shaped active site can progressively hydrolyze polysaccharides, in the case of HvExoI.

We continued to characterize mutations of TxGH116 β -glucosidase acid/base and nucleophile, as well as active site residue mutations and mutations mimicking the disease-causing mutations found in human GBA2. This work resulted in a better understanding of the activity of this enzyme, and particularly the acid/base residue, although we still need to publish this work. Moreover, it allowed us to screen the different acid/base and nucleophile

mutants for production of 1-azido- β -D-glucoside and 1-azido- α -D-glucoside respectively. Production of these compounds allowed us to synthesize new sets of glucosyl 1, 2, 3-triazoles, which can be screened for inhibition of enzymes of interest. Although the initial set of β -glucosyl 1,2,3-triazoles seemed to have little effect on human enzymes, some were quite effective in inhibiting TxGH116, so they may be effective against some bacterial enzymes of interest. The structures of TxGH116 with these compounds are still in process, but may allow us to better understand their inhibition of TxGH116 and why it appears different from the interaction with human GBA2. The mutations of the other residues that interact with glucose in the active site of TxGH116 show the importance of these residues and are likely to identify mutations that have higher activity at high substrate and product concentrations, which are relevant to biomass conversion processes, although these mutants are still being characterized.

Characterization of the cyanobacterium β -glucosidase TeGH116, represents the second characterization of a bacterial enzyme from the GH116 family. Although some aspects of its substrate specificity look similar to TxGH116, it is likely to have a different role in the bacteria, since it appears to be a cytoplasmic protein, while TxGH116 seems likely to be periplasmic. The substrate specificity was subtly different and the thermostability of TeGH116 lower than for TxGH116. Thus, it was useful to determine the structure of TeGH116 to help understand these differences. The availability of more GH116 structures may allow better modeling of human GBA2 and other GH116 enzymes of interest, in addition to helping with future engineering of this family of enzymes.

8.2. Other benefits.

As noted in the output, this work has resulted in the publication of 4 international papers and 1 national paper, with several more papers in process of revising or writing for publication. It has also provided experience to 3 PhD students and one MSc student. Our research assistant and postdoctoral fellows also gained valuable experience working on this project. Although we did not immediately expect patents, we have submitted two patent applications based on the application of our enzymes to production of glucosyl triazoles, which may be useful as enzyme inhibitors in the future.

8.3. Future benefits.

As noted, we have several papers in progress, which will add to the output of this work and the knowledge and products produced may have several applications in the future. We are able to produce modified variants of GH1 transglycosidase and GH116 β -glucosidase that can be applied to producing new glycosides, as noted. These may have many as yet untested bioactivities, which will be tested in the future. For example, we have hopes that the α -glucosyl-1,2,3-triazoles may act as inhibitors of human α -glucosidase and α -amylase, thereby serving as potential anti-diabetic drugs by lowering the glycemic index of starch-based foods. We will need to test this in the future. In addition, this work will bring a new method to producing 1- α -glucosyl-azide and α -glucosyl-1,2,3-triazoles, which are difficult to produce by synthetic organic chemistry. In the future, we may optimize the economics of this process by enzyme immobilization, so that such products can be produced at relatively low cost. We will also test their potential as insecticides and in other potential applications.

The understanding and application of the rice GH1 proteins also has several applications. We know that breeding of rice for low or no production of certain enzymes may be beneficial for producing certain products. For instance, rice with the Os12BGlu38 loss of function could be used in producing hybrid rice, since it is male sterile. The understanding of rice monolignol glucoside metabolism can help to understand whether they

have subtle roles in cell wall lignification or more important roles in lignification induced by pathogen or insect attack. Finally, production of glycosides of interest with Os9BGlu31 mutants can allow us to characterize the functions of these compounds. For instance, production of certain phytohormone glucoconjugates, such as ABA-glucose ester (Ketudat Cairns et al., patent request submitted), has allowed us to characterize enzymes that hydrolyze these substrates *in vitro* and *in planta*. (M. Kongdin et al., in preparation). Other glycosides of potential drugs can be tested for improved solubility and pharmacological properties. Thus, many products could potentially be produced with these engineered enzymes in the future.

9. References:

- Baiya, S., Hua, Y., Ekkhara, W., Ketudat Cairns, J.R. (2014) Expression and enzymatic properties of rice (*Oryza sativa* L.) monolignol beta-glucosidases. *Plant Sci.* 227, 101-109.
- Chapelle, A., Morreel, K., Vanholme, R., Le-Bris, P., Morin, H., Lapierre, C., Boerjan, W., Jouanin, L., Demont-Caulet, N. (2012) Impact of the absence of stem-specific β -glucosidases in lignin and monolignols. *Plant Physiol.* 160, 1204-1217.
- Charoenwattanasatien R, Pengthaisong S, Breen I, Mutoh R, Sansenya S, Hua Y, Tankrathok A, Wu L, Songsiriritthigul C, Tanaka H, Williams SJ, Davies GJ, Kurisu G, Ketudat Cairns JR. (2016) Bacterial β -glucosidase reveals the structural and functional basis of genetic defects in human glucocerebrosidase 2 (GBA2) *ACS Chem. Biol.* 11, 1891-1900.
- Chuankhayan 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.
- Chuenchor W, Pengthaisong S, Robinson RC, Yuvaniyama J, Oonasant 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. (2011) The structural basis of oligosaccharide binding by rice BGlu1 beta-glucosidase. *J. Struct. Biol.* **173**, 169-179.
- Cobucci-Ponzano, B. et al. (2010) A new archaeal beta-glycosidase from *Sulfolobus solfataricus*: seeding a novel retaining beta-glycan-specific glycoside hydrolase family along with the human non-lysosomal glucosylceramidase GBA2. *J Biol Chem* **285**, 20691-20703.
- Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB III, Snoeyink J, Richardson JS, Richardson DC. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375-W383.

- Emsley P. and Cowtan K. 2004. Coot: model-building tools for molecular graphics, *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126-2132.
- Hammer MB, Eleuch-Fayache G, Schottlaender LV, Nehdi H, Gibbs JR, Arepalli SK, Chong SB, Hernandez DG, Sailer A, Liu G, Mistry PK, Cai H, Shrader G, Sassi C, Bouhlal Y, Houlden H, Hentati F, Amouri R, Singleton AB. (2013) Mutations in GBA2 cause autosomal-recessive cerebellar ataxia with spasticity. *Am J Hum Genet* **92**, 245-251.
- Heightman TD, Vasella A. (1999) Recent insights into inhibition, structure, and mechanism of configuration-retaining glycosidases. *Angew. Chem. Int. Ed.* **38**, 750-770.
- Kallemeyn WW, Witte MD, Voorn-Brouwer TM, Walvoort MT, Li KY, Codee JD, van der Marel GA, Boot RG, Overkleeft HS, Aerts JM (2014) A sensitive gel-based method combining distinct cyclophellitol-based probes for the identification of acid/base residues in human retaining β -glucosidases, *J Biol Chem* **289**, 35351-35362.
- Ketudat Cairns JR, Mahong B, Baiya S, Jeon J-S. (2015) β -Glucosidases: Multitasking, moonlighting or simply misunderstood? *Plant Science* **241**, 246-259.
- Ketudat Cairns JR, Esen A. (2010) Beta-Glucosidases. *Cell Mol. Life Sci.* **67**, 3389-3405.
- Komvongsa J, Luang S, Marques JV, Phasai K, Davin LB, Lewis NG, Ketudat Cairns JR. (2015) Active site cleft mutants of Os9BGlu31 transglucosidase modify acceptor substrate specificity and allow production of multiple kaempferol glycosides. *Biochim Biophys Acta, Gen Subj.* **1850**, 1405–1414
- 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.
- 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.
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. (2014) The Carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490–D495.
- 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. *Intl J. Mol. Sci.* **11**, 2759-2769.
- Luang S, Cho J-I, Mahong B, Opassiri R, Akiyama T, Phasai K, Komvongsa J, Sasaki N, Hua Y, Matsuba Y, Ozeki Y, Jeon J-S, Ketudat Cairns JR. 2013. Os9BGlu31 is a transglucosidase with the capacity to equilibrate phenolpropenoid, flavonoid and phytohormone glycoconjugates. *J. Biol. Chem.* **288**, 10111- 10123.

- Ly, H. D., and Withers, S. G. (1999) Mutagenesis of glycosidases. *Annu. Rev. Biochem.* **68**, 487-522.
- Martin E, Schule R, Smets K, Rastetter A, Boukhris A, Loureiro JL, Gonzalez MA, Mundwiller E, Deconinck T, Wessner M, Jornea L, Oteyza AC, Durr A, Martin, JJ, Schols L, Mhiri C, Lamari F, Zuchner S, De Jonghe P, Kabashi E, Brice A, Stevanin G. (2013) Loss of function of glucocerebrosidase GBA2 is responsible for motor neuron defects in hereditary spastic paraplegia. *Am. J. Hum. Genet.* **92**, 238-244,
- Murshudov GN, Lebedev A, Vagin AA, Wilson KS, Dodson EJ. 1999. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta. Crystallogr. D.* **55**, 247-255.
- Müllegger J, Jahn M, Chen HM, Warren RAJ, Withers SG. 2005. Engineering of a thioglycoligase: randomized mutagenesis of the acid–base residue leads to the identification of improved catalysts. *Protein Eng Des Sel.* **18**, 33-40.
- 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.
- Pengthaisong S, Withers SG, Kuaprasert B, Svasti J, Ketudat Cairns JR. (2012a) The role of the oligosaccharide binding cleft of rice BGlu1 in hydrolysis of cellooligosaccharides and in their synthesis by rice BGlu1 glycosynthase. *Protein Sci.* **21**, 362-372.
- Pengthaisong S, Chen C-F, Withers SG, Kuaprasert B, Ketudat Cairns JR. (2012b) Rice BGlu1 glycosynthase and wild type transglycosylation activities distinguished by cyclophellitol inhibition *Carbohydr. Res.* **352**, 51-59.
- Pengthaisong S, Ketudat Cairns JR (2014) Effects of active site cleft residues on oligosaccharide binding, hydrolysis, and glycosynthase activities of rice BGlu1 and its mutants. *Protein Sci.* **23**, 1738-1752.
- Sansenya S, Opassiri R, Kuaprasert B, Chen C-J, Ketudat Cairns JR. (2011) The crystal structure of rice (*Oryza sativa* L.) Os4BGlu12, an oligosaccharide and tuberonic acid glucoside-hydrolyzing beta-glucosidase with significant thioglucohydrolase activity. *Arch. Biochem. Biophys.* **510**, 62-72.
- Seshadri S, Akiyama T, Opassiri R, Kuaprasert B, and 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.
- Tankrathok A, Iglesias-Fernández J, Luang S, Robinson R, Kimura A, Rovira C, Hrmova M, Ketudat Cairns J. (2013) Structural analysis and insights into glycon specificity of the rice GH1 Os7BGlu26 β -D-mannosidase. *Acta Crystallographica Section D Biological Crystallography* **D69**, 2124-2135.
- Tankrathok A, Iglesias-Fernández J, Williams RJ, Pengthaisong S, Baiya S, Hakki Z, Robinson RC, Hrmova M, Rovira C, Williams SJ, Ketudat Cairns JR. (2015) A single

- glycosidase harnesses different pyranoside ring transition state conformations for hydrolysis of mannosides and glucosides. *ACS Catalysis* **5**, 6041-6051.
- 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 α -glucosidase from *Dalbergia cochinchinensis* Pierre, in yeast and bacterial hosts. *Protein Express. Purific.* **48**, 195-204,
- Vagin A, Teplyakov A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022-1025.
- Varghese JN, Hrmova M, Fincher GB. (1999) Three-dimensional structure of a barley α -D-glucan exohydrolase, a family 3 glycosyl hydrolase. *Structure* **7**, 179-190.
- Votsi C, Zamba-Papanicolaou E, Middleton LT, Pantzaris M, Christodoulou K. (2014) A novel GBA2 gene missense mutation in spastic ataxia. *Ann. Hum. Genet.* **78**, 13-22.
- Wang Q, Graham RW, Trimbur D, Warren RAJ, Withers SG. (1994) Changing enzymatic reaction mechanisms by mutagenesis: Conversion of a retaining glucosidase to an inverting enzyme. *J. Am. Chem. Soc.* **116**, 11594-11595.
- Wang Q, Trimbur D, Graham RW, Warren RAJ, Withers SG. (1995) Identification of the acid/base catalyst in *Agrobacterium faecalis* β -glucosidase by kinetic analysis of mutants. *Biochemistry* **34**, 14554-14562.

10. Output from grant project

10.1. International Journal Publications

10.1.1. Published papers

10.1.1.1. Baiya, S., Mahong, B., Lee, S.K., Jeon, J.K., Ketudat Cairns, J.R. (2018) Demonstration of monolignol β -glucosidase activity of rice Os4BGlu14, Os4BGlu16 and Os4BGlu18 in *Arabidopsis thaliana* bglu45 mutant. *Plant Physiol. Biochem.* **127**: 223-230. (June, 2018 issue) DOI:10.1016/j.plaphy.2018.03.026 (Accepted 23 March, 2018). (ISI IF 2018 3.404)

10.1.1.2. Gorantla JN, Pengthaisong S, Choknud S, Kaewpuang T, Manyum T, Promarak V, Ketudat Cairns JR. (2019) Gram scale production of 1-azido- β -D-glucose via enzyme catalysis for the synthesis of 1,2,3-triazole-glucosides. *RSC Advances* **9**: 6211-6220. DOI: 10.1039/c9ra00736a (Accepted 14 February, 2019; ISI JCR 2018 IF 3.049).

10.1.1.3. Streltsov VA, Luang S, Peisley A, Varghese JN, Ketudat Cairns JR, Fort S, Hijnen M, Tvaroška I, Ardá A, Jiménez-Barbero J, Alfonso-Prieto M, Rovira C, Mendoza F, Tiessler-Sala L, Sánchez-Aparicio S-E, Rodríguez-Guerra J, Lluch JM, Maréchal J-D, Masgrau L, Hrmova M (2019) Discovery of processive catalysis by an exo-hydrolase with a pocket-shaped active site. *Nature Communications* **10**: 2222 [ISI JCR 2018 IF=11.878].

10.1.1.4. Tran LT, Blay V, Luang S, Eurtivong C, Choknud S, González-Díaz H, Ketudat Cairns JR 2019. Engineering faster transglycosidases and their acceptor specificity. *Green Chemistry* **21**: 2823-2836. (Accepted 24 April, 2019, DOI: 10.1039/C9GC00621D; ISI JCR 2018: 9.405).

10.1.2. Papers under revision & in preparation

10.1.2.1. Mahong B, Kongdin M, Lee SK, Shima SH, Lee C, Kim YJ, Zhang D, Ketudat Cairns JR and Jeon JS. Rice β -glucosidase Os12BGlu38 is required for synthesis of intine cell wall and pollen fertility. (last rejected from *The Plant Cell* 2018 IF 8.228).

10.1.2.2. Prawisut A, Choknud S, Ketudat Cairns JR. Expression and purification of rice β -exohydrolase II (OsExoII) in *Escherichia coli*.

10.1.2.3. Pengthaisong S, ... Rovira C, Ketudat Cairns JR. Unusual acid/base geometry of a GH116 beta-glucosidase.

10.1.2.4. Baiya S, Pengthaisong S, Ketudat Cairns JR. Structure of a monolignol β -glucosidase.

Cumulative Impact Factor: 27.736 (published papers only)

10.2. National Journal Publications

10.2.1. Baiya, S., Pengthaisong, S., Ketudat Cairns, J.R. (2017) Crystallization and X-ray diffraction analysis of rice Os4BGlu18. *J. Appl. Sci.* 16(2): 37-46.

10.3. Patent Submissions

10.3.1. Thai patent submission number 1801005294 กรรมวิธีการผลิต 1-อะซิโด-บีตา-ดี-กลูโคส โดยใช้การเร่งปฏิกิริยาค่ายเอนไซม์ English name: Method for production of 1-azido-beta-D-glucose by enzymatic catalysis. Submitted 5 September, 2018.

10.3.2. Thai patent submission number 1901003181

English name: Enzymatic process for the production of α -D-glycosylazide and synthesis of 1,2,3-triazole- α -D-glucoside derivatives. Submitted 28 May, 2019.

10.4. International Meeting Presentations:

10.4.3.1. Ketudat Cairns, J.R. Structural investigation of GH116 β -glucosidase mechanism; Session 3, Invited Lecture 1; Ninth Korea/Japan Seminars on Biomolecular Sciences Experiments and Simulations 13-16 Nov., 2016, Gyeongju, South Korea.

10.4.3.2. Ketudat Cairns, J.R., Svasti, J., Pengthaisong, S., Charoenwattanasatien, R., Kurisu, G., Rimlumduan, T., Hua, Y., Tanaka, T. Invited Lecture. APPA/PST2017, the Protein Society of Thailand, and recent examples of Thai-Japanese collaboration in structural biology. 15th Symposium of the Protein Science Society of Japan. 20-22 June, 2017, Sendai, Japan.

10.4.3.3. Pengthaisong, S., Charoenwattanasatien R, Kurisu, G., Ketudat Cairns, J.R. Poster PP06-3. Structural analysis of *Thermoanaerobacterium xylanolyticum* TxGH116 β -glucosidase mutants corresponding to pathogenic human GBA2 mutations. The 5th Asia Pacific Protein Association Conference/ 12th International Symposium of the Protein Society of Thailand, 11-14 July, 2017, Bangsaen, Thailand. Abstract book Pg. 105.

10.4.3.4. Beagbandee, C., Ketudat Cairns, J.R. Poster PP07-4 Recombinant expression of GBA2 glucosylceramidase in *Escherichia coli*. The 5th Asia Pacific Protein Association Conference/ 12th International Symposium of the Protein Society of Thailand, 11-14 July, 2017, Bangsaen, Thailand. Abstract book Pg. 106.

10.4.3.5. Sala, K., Ketudat Cairns, J.R. Poster PP43-24 Recombinant expression of OsXyl1 β -xylosidase in *Pichia pastoris*. The 5th Asia Pacific Protein Association Conference/ 12th International Symposium of the Protein Society of Thailand, 11-14 July, 2017, Bangsaen, Thailand. Abstract book Pg. 126.

10.4.3.6. Prawisut, A., Ketudat Cairns, J.R. Poster PP47-26 Expression, purification and characterization of rice (*Oryza sativa*) β -D-glucan exoglucanase. The 5th Asia Pacific Protein Association Conference/ 12th International Symposium of the Protein Society of Thailand, 11-14 July, 2017, Bangsaen, Thailand. Abstract book Pg. 128.

10.4.3.7 Ketudat Cairns JR, (2017) Function and application of β -glucosidases and transglucosidases. International Conference of Molecular Bioscience & Biomedical Engineering (ICAMBBE), 5-7 September, 2017, Universitas Brawijaya, Malang, Indonesia, Plenary lecture.

10.4.3.8. Ketudat Cairns JR, Mahong B, Baiya S Komvongsa J, Tran L, Pengthaisong S, Jeon JS. (2017) Insights into rice glycoside hydrolase family 1 function from metabolomic and

structural studies. 15th International Symposium on Rice Functional Genomics. 25-28 September, 2017, Suwon, Korea. Concurrent Program C12, Program pg. 38.

10.4.3.9. Ketudat Cairns JR, Pengthaisong S, Kongdin M, Tran L, Baiya S, Choknud S, Mahong B, Jeon JS. (2017) Structure, function and application of rice β -glucosidases and transglucosidases. 4th International Seminar, Congress and Workshop of the Indonesian Protein Society. 8-9 November, 2017, Jember, Indonesia. Plenary Lecture 2.

10.4.3.10. Ketudat Cairns JR, Ngisara L, Pengthaisong S, Sawangareetrakul P, Champattanachai V, Kantaputra P, Svasti J. (2018) The structural basis for human genetic disease. The 8th Annual Basic Science International Conference (BaSIC), Ijen Resort Hotel, Malang, Indonesia, March, 2018; Keynote Lecture 2.

10.4.3.11. Ketudat Cairns JR (2018) Beta-glucosidase structure and function: from plants to bacteria to human disease. The 6th International Conference on Biochemistry and Molecular Biology, Rayong Resort, Rayong, Thailand, 2018, Plenary Lecture 3. Abstract book page 5.

10.4.3.12. Ketudat Cairns JR, Pengthasong S, (2018) 29th International Carbohydrate Symposium, Universiti Lisboa, Lisboa, Portugal, 15-19 July, 2018. Oral Presentation

10.4.3.13. Ketudat Cairns JR, (2018) Protein science development in Thailand over the past 20 years, one man's perspective 13th International Symposium of the Protein Society of Thailand, Chulabhorn Research Institute, Bangkok, Thailand, 7-9 August, 2018. Jisnuson Svasti PST Award Lecture. Abstract book pg. 4.

10.4.3.14. Pengthaisong S, Charoenwattanasatien R, Kurisu G, Ketudat Cairns JR. (2018) Kinetics and structures of *Thermoanaerobacterium xylanolyticum* TxGH116 β -glucosidase mutants corresponding to pathogenic human GBA2 mutations. The 6th International Conference on Biochemistry and Molecular Biology, Rayong Resort, Rayong, Thailand, 2018, S1-P-34, Abstract book pg 80.

10.4.3.15. Beagbandee C, Pengthaisong S, Ketudat Cairns JR (2018) Purification and characterization of *Thermosynechococcus elongates* TeGH116 β -Glucosidase expressed in *Escherichia coli*. The 6th International Conference on Biochemistry and Molecular Biology, Rayong Resort, Rayong, Thailand, 2018, S1-P-06. Abstract book Pg 52.

10.4.3.16. Pengthaisong S, Ketudat Cairns JR (2018) Nucleophile mutants of *Thermoanaerobacterium xylanolyticum* TxGH116 β -glucosidase for glucoside and oligosaccharide synthesis. 13th International Symposium of the Protein Society of Thailand, Chulabhorn Research Institute, Bangkok, Thailand, 7-9 August, 2018. PP027-7, Abstract book pg. 45.

10.4.3.17. Sala K, Ketudat Cairns JR (2018) Expression of GH3 β -xylosidase from rice (*Oryza sativa* L.) in *Pichia pastoris* and characterization of its enzymatic properties. 13th International Symposium of the Protein Society of Thailand, Chulabhorn Research Institute, Bangkok, Thailand, 7-9 August, 2018. PP028-8, Abstract book pg. 46.

10.4.3.18. Beagbandee C, Gorantla JN, Pengthaisong S, Ketudat Cairns JR (2018) Characterization of *Thermosynechococcus elongatus* TeGH116 β -glucosidase expressed in *Escherichia coli*. 13th International Symposium of the Protein Society of Thailand,

Chulabhorn Research Institute, Bangkok, Thailand, 7-9 August, 2018. PP071-23, Abstract book pg. 61.

10.4.3.19. Prawisut A, Ketudat Cairns JR. (2018) Expression, purification, and characterization of rice (*Oriza sativa*) β -D-glucan exoglucanase. 13th International Symposium of the Protein Society of Thailand, Chulabhorn Research Institute, Bangkok, Thailand, 7-9 August, 2018. PP131-43, Abstract book pg. 81.

10.4.3.20. Ketudat Cairns JR, Ngiwsara L, Pengthaisong S, Jatoorathaweechot P, Beagbandee C, Champattanachai V, Sawangareetrakul P, Svasti J. (2018) Human β -Glucosidases in Health and Disease. The 5th International Seminar and Workshop of Indonesian Protein Society, Hasanuddin University, Makasar, S. Sulawesi, Indonesia, 8-9 November, 2018. Plenary talk 5.

10.4.3.21. Ketudat Cairns JR, Tran LT, Choknud S, Blay Roger V, Robinson RC. (2019) Os9BGlu31 transglucosidase variants with high and promiscuous activity. 33rd Annual Symposium of the Protein Society. 30 June-3 July, 2019. Sheratin Grand Seattle Hotel, Seattle, WA, USA. Poster ABS263

10.4.3.22. Beagbandee C, Charoenwattanasatien R, Pengthaisong S, Kurisu G, Ketudat Cairns JR. (2019) Characterization and crystallization of TeGH116, a cyanobacterial glycoside hydrolase family 116 beta-glucosidase. The 14th International Symposium of the Protein Society of Thailand "Protein Technology for a Better Life" 22-23 July, 2019, Maruay Garden Hotel, Bangkok, Thailand. Poster PO44 Abstracts and Proceedings pg. 84.

10.4.3.23. Huang M, Ketudat Cairns JR. (2019) The roles of active site amino acid residues of TxGH116 in glucose binding and catalysis. The 14th International Symposium of the Protein Society of Thailand "Protein Technology for a Better Life" 22-23 July, 2019, Maruay Garden Hotel, Bangkok, Thailand. Poster PO54 Abstracts and Proceedings pg. 94.

10.5. Graduate Students' Training

10.5.1. Mr. Akkarawit Prawisut, Studying for Ph.D. 2013-present, Suranaree University of Technology, Biochemistry Graduate Program, School of Chemistry.

10.5.2. Ms. Kadsada Sala, Studying for Ph.D. 2015-present, Suranaree University of Technology, Biochemistry Graduate Program, School of Chemistry.

10.5.3. Mr. Meng Huang, Studying for Ph.D. 2016-present, Suranaree University of Technology, Biochemistry Graduate Program, School of Chemistry.

10.6. Collaborations

10.6.1. Prof. Dr. M.R. Jisnuson Svasti, Laboratory of Biochemistry, Chulabhorn Research Institute, Bangkok, Thailand.

10.6.2. Dr. Ratana Charoenwattanasatien, Synchrotron Light Research Institute, Thailand.

- 10.6.3. Dr. Chomphunuch Songsiriritthukul, Synchrotron Light Research Institute, Thailand.
- 10.6.4. Prof. Dr. Maria Hrmova, Australian Centre for Plant Functional Genomics, University of Adelaide, Waite Campus, Glen Osmond, SA, Australia.
- 10.6.5. Prof. Dr. Spencer Williams, University of Melbourne, School of Chemistry and Bio21 Molecular Science and Biotechnology Institute, Parkville, Victoria, Australia.
- 10.6.6. Prof. Dr. Stephen G. Withers, Department of Chemistry, University of British Columbia, Vancouver, British Columbia, Canada.
- 10.6.7. Prof. Dr. Robert C. Robinson, Institute of Molecular and Cell Biology, 61 Biopolis Drive, Proteos, Singapore.
- 10.6.8. Prof. Dr. Atsuo Kimura, Graduate Faculty of Agriculture, Hokkaido University, Sapporo, Hokkaido, Japan.
- 10.6.9. Prof. Dr. Carme Rovira, Computer Simulation and Modeling Laboratory and Institut de Química Teòrica i Computacional (IQTUB), Parc Científic de Barcelona, and Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys, Barcelona, Spain.
- 10.6.10. Prof. Dr. Chun-Jung Chen, National Synchrotron Radiation Research Center, Hsinchu, Taiwan.
- 10.6.11. Prof. Dr. Jong-Seong Jeon, Kyung Hee University, Suwon, S. Korea.
- Prof. Genji Kurisu, Institute for Protein Research, Osaka University, Osaka, Japan
- 10.6.12. Prof. Gideon Davies, Structural Biology Laboratory, Department of Chemistry, University of York, York, UK

11. Copies of Published International Journal Papers.



Research article

Demonstration of monolignol β -glucosidase activity of rice Os4BGlu14, Os4BGlu16 and Os4BGlu18 in *Arabidopsis thaliana* bglu45 mutant

Supaporn Baiya^{a,b}, Bancha Mahong^c, Sang-Kyu Lee^c, Jong-Seong Jeon^{c,**}, James R. Ketudat Cairns^{b,d,e,*}

^a Faculty of Science at Sriracha, Kasetsart University, Sriracha Campus, Chonburi, 20230, Thailand

^b Center for Biomolecular Structure, Function and Application, Suranaree University of Technology, Nakhon Ratchasima, 30000, Thailand

^c Graduate School of Biotechnology, Kyung-Hee University, Yongin, 17104, South Korea

^d School of Chemistry, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima, 30000, Thailand

^e Laboratory of Biochemistry, Chulabhorn Research Institute, Bangkok, 10210, Thailand



ARTICLE INFO

Keywords:

Monolignol β -glucosidases

Monolignol glucosides

Arabidopsis

Rice

Heterologous expression

UPLC-MSMS

ABSTRACT

The glycoside hydrolase family 1 members Os4BGlu14, Os4BGlu16, and Os4BGlu18 were proposed to be rice monolignol β -glucosidases. *In vitro* studies demonstrated that the Os4BGlu16 and Os4BGlu18 hydrolyze the monolignol glucosides coniferin and syringin with high efficiency compared to other substrates. The replacement of the conserved catalytic acid/base glutamate residue by a nonionizable glutamine residue in Os4BGlu14 suggested that it may be inactive as a β -glucosidase. Here, we investigated the activities of Os4BGlu14, Os4BGlu16, and Os4BGlu18 *in planta* by recombinant expression of their genes in the *Arabidopsis bglu45-2* (monolignol β -glucosidase) mutant and analysis of monolignol glucosides by ultra-performance liquid chromatography-tandem mass spectrometry (UPLC-MSMS). The *bglu45-2* line exhibits elevated monolignol glucoside levels, but lower amounts of coniferin, syringin, and *p*-coumaryl alcohol glucoside were seen in *Arabidopsis bglu45-2* rescued lines complemented by the *Os4BGlu14*, *Os4BGlu16*, and *Os4BGlu18* genes. These data suggest that the *bglu45-2* mutant has a broader effect on monolignols than previously reported and that the Os4BGlu14, Os4BGlu16 and Os4BGlu18 proteins act as monolignol β -glucosidases to complement the defect. An Os4BGlu16-GFP fusion protein localized to the cell wall. This apoplastic localization and the effect of these enzymes on monolignol glucoside levels suggest monolignol glucosides from the vacuole may meet the monolignol β -glucosidases, despite their different localization.

1. Introduction

p-Coumaryl alcohol glucoside (*p*CAG), coniferin, and syringin are monolignol glucosides, which have been proposed to be transport and storage forms of monolignols, the precursors of lignin and lignans (Chapelle et al., 2012). Their aglycones, *p*-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol are precursors to the *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units found in lignin (Boerjan et al., 2003), respectively. It is well understood that monolignol glucosides are glucosylated by a uridine diphosphate glucose (UDP-glucose)-dependent glucosyltransferase and deglycosylated by monolignol specific β -D-glucosidases (Dharmawardhana and Ellis, 1998; Steeves et al., 2001; Escamilla-Treviño et al., 2006; Lanot et al., 2006; Baiya et al., 2014; Tsuyama and Takabe, 2015; Lin et al., 2016). Monolignol glucosides accumulate in lignifying xylem close to the cambium of

gymnosperms and also at lower levels in some angiosperms, such as the Magnoliaceae and Linaceae families (Campbell and Sederoff, 1996; Baucher et al., 1998; Huis et al., 2012; Terashima et al., 2016). Recently, localization of coniferin in *Ginkgo biloba* L. by cryo time-of-flight secondary ion mass spectrometry and scanning electron microscopy (cryo-TOF-SIMS/SEM) revealed that coniferin is accumulated in the vacuoles of the tracheid cells of differentiating xylem (Aoki et al., 2016). Although largely found in the stem and leaves, coniferin and syringin were also detected in roots of *Arabidopsis* and they were decreased after infection by *Pythium sylvaticum* (Bednarek et al., 2005).

Dixon et al. (2001) proposed a model of the H, G, and S unit metabolic pathways for synthesis of monolignol related compounds. In this model, the methylation reactions occurred at the level of the free acids cinnamate, 4-coumarate, caffeate, ferulate, 5-hydroxyferulate, and sinapate. The enzymes 4-coumarate CoA ligase (4CL), cinnamoyl CoA

* Corresponding author. School of Chemistry, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima, 30000, Thailand.

** Corresponding author. Graduate School of Biotechnology, Kyung-Hee University, Yongin, 17104, South Korea.

E-mail addresses: jjeon@khu.ac.kr (J.-S. Jeon), cairns@sut.ac.th (J.R. Ketudat Cairns).

<https://doi.org/10.1016/j.plaphy.2018.03.026>

Received 15 July 2017; Received in revised form 22 March 2018; Accepted 23 March 2018

Available online 28 March 2018

0981-9428/ © 2018 Elsevier Masson SAS. All rights reserved.

reductase (CCR), and cinnamyl alcohol dehydrogenase (CAD) then catalyze the conversion of the free acids into the coenzyme A (CoA) ester, aldehyde, and alcohol forms, respectively.

Glycoside hydrolase family 1 (GH1) β -glucosidases are abundant in plants and carry out a wide range of functions, including release of phytohormones, biosynthetic intermediates, aromatic compounds and monolignols from their glucosides (Ketudat Cairns et al., 2015). Analysis of GH1 members encoded in the *Arabidopsis thaliana* genome was first reported by Xu et al. (2004). From phylogenetic analysis, they divided GH1 into 10 subfamilies, which contain 47 genes encoding β -glucosidase-like proteins and the divergent gene *Sensitive to Freezing 2* (*SFR2*), which encodes a galactolipid galactolipid transgalactosidase (Moellering et al., 2010). BGLU45 (*At1g61810*), BGLU46 (*At1g61820*), and BGLU47 (*At4g21760*) are most closely related to *Pinus contorta* coniferin β -glucosidase. Among these, BGLU45 and BGLU46 are most closely related to each other sharing 80%, and their genes are located together on chromosome 1, while BGLU47, which was predicted to be targeted to the peroxisome, is more divergent and its gene is located on chromosome 4 (Xu et al., 2004). Escamilla-Treviño et al. (2006) expressed the glycosylated BGLU45 and BGLU46 enzymes in *Pichia pastoris* and found that BGLU45 could hydrolyze monolignol glucoside substrates (coniferin, syringin, and *p*CAG), while BGLU46 showed a broader substrate specificity, hydrolyzing salicin, *p*-CAG, phenyl β -D-glucoside, 4-methylumbelliferyl β -D-glucoside, coniferin, syringin, arbutin, and amygdalin. In addition, RT-PCR analysis of BGLU45 and BGLU46 genes demonstrated that these genes are expressed in the stem, an expected lignification site. Moreover, immunolocalization demonstrated that the Bglu45 is mainly found in the interfascicular fibers and BGLU46 in the protoxylem of *Arabidopsis* stems. *Arabidopsis* lines with the BGLU45 and BGLU46 genes knocked out displayed a significant increase in the abundance of coniferin, but not for syringin (Chapelle et al., 2012). However, no significant changes in the abundances of coniferyl alcohol and sinapyl alcohol were observed in this experiment. Nonetheless, these data suggest that coniferin is a substrate for BGLU45 and BGLU46 *in planta* and it could serve as an inactive storage form of monolignol precursor, although no significant changes were observed in the lignin of the knockout plants.

When the GH1 β -glucosidase genes in the rice (*Oryza sativa* L.) genome were identified, the Os4BGlu14, Os4BGlu16, and Os4BGlu18 were found to be closely related to *Arabidopsis* BGLU45, BGLU46, and BGLU47 and designated as putative monolignol β -glucosidases (Opassiri et al., 2006). In a recent *in vitro* enzymatic study, Os4BGlu16 and Os4BGlu18 exhibited high catalytic activity for hydrolysis of all three monolignol glucoside substrates when compared with other natural substrates (Baiya et al., 2014). Os4BGlu14 could not be expressed and demonstrated to have β -glucosidase activity, which together with its lack of the conserved ionizable glutamate residue at the catalytic acid/base position, led us to believe it might be an inactive product of a dying gene or have another function. To investigate whether rice Os4BGlu14, Os4BGlu16, and Os4BGlu18 can function as monolignol β -glucosidases *in vivo*, we attempted to rescue an *Arabidopsis* BGLU45 knockout line with all three genes and measure the effects on coniferin

and other monolignol glucoside levels.

2. Materials and methods

2.1. Plant material and growth conditions

Arabidopsis At1g61810 (*bglu45-2*, Col-0 ecotype) mutant seeds corresponding to the Salk_117269 flanking sequence tag were ordered from the Salk Institute Genomic Analysis Laboratory. Homozygous *bglu45-2* mutant plants were isolated by genotype analysis with *Arabidopsis* BGLU45 gene- and T-DNA-specific primers, as described by Chapelle et al. (2012). The *Arabidopsis bglu45-2* mutant lines rescued with Os4BGlu14, Os4BGlu16 and Os4BGlu18, homozygous *bglu45-2* mutant and segregated wild-type plants were grown together in the growth chamber under long-day conditions with some modifications from Chapelle et al. (2012). Briefly, all the plants were grown in 22–25 °C day/18–15 °C night, 16 h of light/8 h dark at 60% relative humidity. Ten replicates of each line were planted, from which four plants with similar growth were sampled for further analysis. Ten-week-old *Arabidopsis* was used for determining monolignol compounds.

2.2. Construction of plant expression vectors and *Arabidopsis bglu45-2* transformation

For expression in *Arabidopsis*, the Os4BGlu14, Os4BGlu16, and Os4BGlu18 cDNA were cloned into pGWB502 expression vector (Nakagawa et al., 2007). The full length cDNA for each gene was amplified with a SolGent™ *Pfu* DNA polymerase (SolGent, Daejeon, Korea) from japonica rice (cv. Nipponbare) cDNA clones [Os4BGlu14 accession number AK067841, Os4BGlu16 accession number AK068772, and Os4BGlu18 accession number AK058333], which were ordered from the Knowledge-based Oryza Molecular biological Encyclopedia (KOME). The primer pairs used for PCR amplification were Bglu14fwd: 5'-CAC CGGAAGACGACGACGAGATGG-3', Bglu16fwd: 5'-CACCTGGGTGG ATCGGTCGAGATGGC-3', and Bglu18fwd: 5'-CACCATGGCAGGAGGC AGT AAGACGCGGA-3' as forward primers and Bglu14rev: 5'-GTAAT ATGGAAGGGAAGAACAG-3', Bglu16rev: 5'-GCTAGACGTATTCTAGC ATCGGA-3', and Bglu18rev: 5'-CTATTGATTTTCATGCAGATTTTG-3' as reverse primers of Os4BGlu14, Os4BGlu16, and Os4BGlu18, respectively. The respective PCR products were cloned into the pENTR/D-TOPO vector (Thermo-Science Invitrogen, Carlsbad, CA, USA), and then the recombinant plasmids were verified by sequencing. The cDNA inserts were cloned via LR clonase recombination (Invitrogen, Gaithersburg, MD, USA) into the Gateway binary vector pGWB502, which contains a 35S promoter and nopaline synthase (NOS) terminator (Fig. 1). The recombinant binary vectors for expression of the three monolignol β -glucosidase-like genes were introduced into *Agrobacterium tumefaciens* GV3101 by electroporation and subsequently transformed into knockout mutant *Arabidopsis* line Salk_117269, in which the T-DNA is inserted into the BGLU45 gene (*bglu45-2*) as described by Chapelle et al. (2012), by the floral dip method (Clough and Bent, 1998).

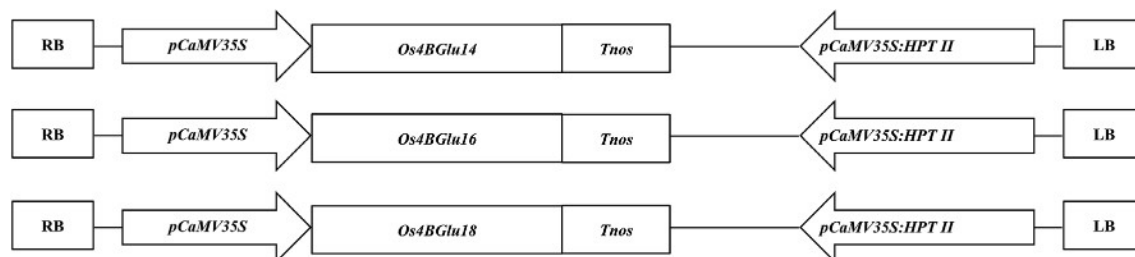


Fig. 1. Schematic diagram of plant expression vectors of the *Os4BGlu14*, *Os4BGlu16*, and *Os4BGlu18* genes.

The three cloned full length cDNAs including partial 5' untranslated region (UTR) and 3'UTR sequences were inserted under the control of the pCaMV 35S promoter. LB and RB, left and right borders of T-DNA; HPT II, hygromycin phosphotransferase II gene; Tnos, nopaline synthase terminator.

2.3. RNA isolation and RT-PCR analysis

Total RNAs from 1 month old leaves of homozygous *Arabidopsis bglu45-2* mutant, pGWB502/Os4BGLu14, pGWB502/Os4BGLu16, and pGWB502/Os4BGLu18 rescued *Arabidopsis bglu45-2*, and segregated wild type were prepared using Trizol reagent (Invitrogen). The isolated RNA extracts were reverse-transcribed with an oligo-dT primer and a First Strand cDNA Synthesis Kit (Roche, Mannheim, Germany). The first-strand cDNAs were used as templates in the RT-PCR with BGLu14qPCRf: 5'-GATGGAAGAAGCAGTCACAT-3', BGLu16qPCRf: 5'-TTCATGGGTACTTTGTCGTC-3', and BGLu18qPCRf: 5'-GATAGTTACACCGACGCAGA-3' as forward primers and BGLu14qPCRr: 5'-TCGAAATCATCAAGGAGAGA-3', BGLu16qPCRr: 5'-GTGTCGAAATCCACTTGTA-3', and BGLu18qPCRr: 5'-AGCTTTGGAGATCTTTCC TG-3' as reverse primers of Os4BGLu14, Os4BGLu16, and Os4BGLu18, respectively. The *Arabidopsis* Tubulin1 (AtG75780) gene-specific primers were used as the internal control (Oppenheimer et al., 1988).

2.4. Subcellular localization analysis of the Os4BGLu16-GFP fusion protein

The Os4BGLu16 entire open reading frame without stop codon was amplified by PCR using a SolGent™ *Pfu* DNA polymerase (SolGent). The primers used were 5'-AAAAAGCAGGCTTCATGGCCGTGGCGGCGGC GAC-3' and 5'-AGAAAGCTGGGTGCGAATCTGCTCTACGTGACC-3' for Os4BGLu16-GFP fusion. The PCR product was cloned into pDONR201 and transferred into between the CaMV35S promoter and Green Fluorescent Protein (GFP) of pH7FWG2 by the Gateway LR reaction (Karimi et al., 2002). The resulting GFP fusion construct was infiltrated into a tobacco leaf by an *Agrobacterium*-mediated infiltration method (Rouyi et al., 2014). For plasmolysis experiments, tobacco leaves expressing Os4BGLu16-GFP were infiltrated with 0.8 M mannitol solution. GFP signal was detected by excitation with the 488 nm line of the argon laser and a capturing emission at 522 nm by laser-scanning confocal microscopy (LSM 510 META, Carl Zeiss, Jena, Germany) at various times after transformation. Chlorophyll autofluorescence was used as markers of chloroplasts.

2.5. Plant extraction and UPLC-MSMS analysis

The leaves and stems of ten week old homozygous *Arabidopsis bglu45-2* mutant, pGWB502/Os4BGLu14, pGWB502/Os4BGLu16, and pGWB502/Os4BGLu18 rescued *Arabidopsis bglu45-2*, and segregated wild type stems were ground in liquid nitrogen. Fifty-milligrams of ground samples were extracted in 500 µl of 80% methanol containing 50 µM of naphthalene acetic acid (NAA) as an internal standard and mixed by vortexing 5 min and sonicating 15 min. Then, the supernatant was collected by centrifugation at $12,470 \times g$ for 15 min. Five microliters of each sample was injected into an Agilent SB-C18 RRHD 1.8 µm, 2.1×150 mm column (Agilent Technologies, CA, USA) on an Agilent 1290 UPLC system inline with an Agilent 6490 triple quadrupole mass spectrometer. A gradient of buffer A (100:1:0.1, water:acetonitrile:2 M ammonium acetate, pH 5.0) and buffer B (100:1:0.1, acetonitrile:water:2 M ammonium acetate, pH 5.0) was used at the flow rate of 0.2 ml/min (Chapelle et al., 2012). The gradient started at 95% A and decreased to 65% in 34 min. The temperature of the column was set at 40 °C. A UV visible spectrum was measured between 190 and 600 nm on the inline diode array detector (DAD). Electrospray ionization (ESI) was used in the negative ion mode with the multiple reactant monitoring (MRM) method. The precursor ion and product ion masses of syringin (m/z 431 > 179, retention time [Rt] 11.6 min, collision energy [CE] 35 V), coniferin (m/z 401 > 179, Rt 10.2 min, CE 5 V), pCAG (m/z 371 > 148.9, Rt 8.2 min, CE 10 V), sinapyl alcohol (m/z 209 > 178.9, Rt 11.6 min, CE 18 V), coniferyl alcohol (m/z 179 > 146, Rt 17.9 min, CE 6 V), p-coumaryl alcohol (m/z 149 > 130.1, Rt 15.6 min, CE 14 V), sinapic acid (m/z 223 > 149.1, Rt 19.5 min, CE 22 V), ferulic acid (m/z 193 > 178, Rt 18.3 min, CE

6 V), p-coumaric acid (m/z 163 > 93, Rt 15.4 min, CE 34 V), vanillic acid (m/z 167 > 108, Rt 8.8 min, CE 10 V), cinnamic acid (m/z 147 > 103, Rt 28.4 min, CE 2 V), caffeic acid (m/z 179 > 135.1, Rt 10.2 min, CE 14 V), and 1-naphthaleneacetic acid [NAA] (m/z 185 > 141, Rt 33.7 min, CE 2 V) were monitored with the cell accelerator voltage at 4 V. Notably, the precursor ions of three monolignol glucosides were detected as their acetate adducts, which have masses 59.0 Da higher than the monolignol molecular weights. Four biological replicates of each sample were injected two times each to carry out statistical analyses. The abundance values were compared to standard curves (Supplementary Figs. S1–S12) of the specific compounds to determine the amount of each monolignol compound in the extract and NAA was used as a reference compound with the abundance shown relative to the measured NAA concentration (NAA standard curve regression: $y = 3.04x + 46.74$, $R^2 = 0.9994$). The compound concentrations in the extracts were calculated by multiplying the values obtained from the standard curve by the concentration of NAA added to the extract and dividing by the concentration of NAA measured in the extract.

2.6. Statistical analysis

To verify the significant differences between the concentration of monolignol compounds in *Arabidopsis bglu45-2* rescued by monolignol β-glucosidases, *Arabidopsis bglu45-2* mutant and segregated wild type lines, all samples were performed in four biological replicates. Statistical significance was evaluated using the IBM SPSS statistics software with one-way ANOVA followed by post hoc Scheffe's test method at a significance level of $P < 0.01$.

3. Results

3.1. Construction of transgenic *Arabidopsis bglu45-2* lines rescued with rice monolignol β-glucosidases

The pGWB502 expression vectors containing the Os4BGLu14, Os4BGLu16, and Os4BGLu18 cDNAs under control of CaMV 35S promoter (Fig. 1) were transformed into homozygous *Arabidopsis thaliana bglu45-2* mutant plants by the floral dip method to obtain putative rice monolignol β-glucosidase-rescued lines. The rescue lines were proven to express the expected putative rice monolignol β-glucosidase gene, but not the BGLU45 gene, by reverse transcription PCR (RT-PCR) (Fig. 2). The positive DNA band was amplified for each specific rice gene in the respective rescue lines, and no band of rice monolignol β-glucosidases genes was detected in segregated wild-type and *bglu45-2* mutant lines.

3.2. UPLC-MSMS analysis of monolignol β-glucosidases rescued *Arabidopsis bglu45-2* mutants

The extracts of the *Arabidopsis bglu45-2* mutant lines rescued with Os4BGLu14, Os4BGLu16 and Os4BGLu18 had levels of monolignol glucosides that were significantly lower than the *bglu45-2* parent line and similar to the Col-0 wild type plant extracts. The relative amounts of syringin, coniferin, and pCAG in *bglu45-2* rescued by all three genes were decreased more than 2-fold, when compared with homozygous *Arabidopsis bglu45-2* mutant (Fig. 3). The levels of the monolignols sinapyl, coniferyl, and p-coumaryl alcohol in the rescued lines were lower than in the *Arabidopsis bglu45-2* mutant and wild type, while the *Arabidopsis bglu45-2* mutant and wild type had similar levels of coniferyl and p-coumaryl alcohols (Fig. 4). The levels of the monolignol biosynthesis intermediate compounds (phenyl propanoic acids) cinnamic acid, p-coumaric acid, and ferulic acid were similar between the extracts of the rescued lines and *bglu45-2* and wild type plants, as was the level of the ferulic acid derivative vanillic acid (Fig. 5). Sinapic acid was significantly lower in the *bglu45-2* line (and higher in the Os4BGLu18

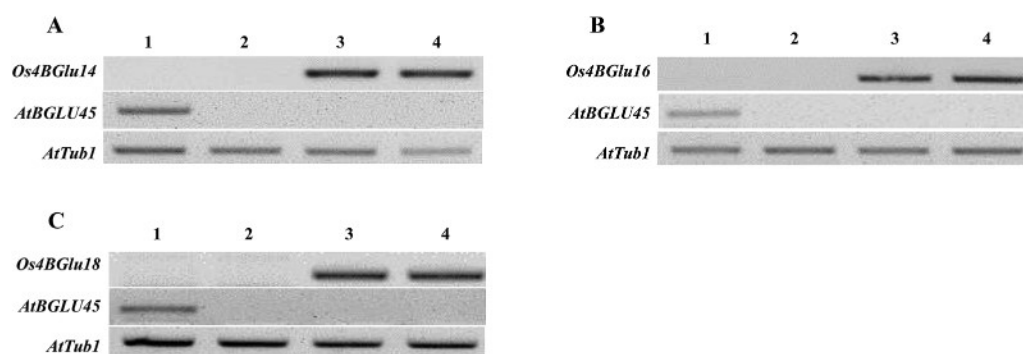


Fig. 2. RT-PCR analysis of β -glucosidase gene expression in *Arabidopsis thaliana* wild type, *bglu45-2* mutant and *bglu45-2* mutant lines rescued by *Os4BGLu14* (A), *Os4BGLu16* (B), and *Os4BGLu18* (C).

Lane 1, segregated wild-type of BGLU45; lane 2, homozygous knock out mutant of *bglu45-2*; lanes 3 and 4, complemented lines of the indicated gene.

rescued line), while caffeic acid was significantly higher in the *bglu45-2* extracts, which demonstrates that there was no systematic increase in the *bglu45-2* mutant extracts compared to the wild type and rescued lines. Nonetheless, the significance of these phenolic metabolite differences is unclear.

3.3. Subcellular localization of *Os4BGLu16*-GFP

Previously, pine tree monolignol β -glucosidase activity was shown to localize to the secondary xylem tissue cell wall (Dharmawardhana et al., 1995) and AtBGLU45 and 46 were also immunolocalized to cell wall (Chapelle et al., 2012), leading to the assumption that monolignol β -glucosidases are found in the cell wall. However, monolignol β -glucosidases are specifically transported to the vacuole, while the free monolignols are transported to the cell wall (Miao and Liu, 2010; Dima et al., 2015). Since a cell wall β -glucosidase might not be expected to affect vacuolar monolignol glucoside levels, *Os4BGLu16* tagged with a C-terminal GFP was transiently expressed in tobacco leaf epithelial cells to check its localization. As can be seen in Fig. 6, the GFP signal was exclusively localized to the plasma membrane and extracellular space in both intact and plasmolysed cells, consistent with an apoplastic or cell wall localization.

4. Discussion

Metabolomics studies have recently revealed several new aspects of lignin biosynthesis and monolignol metabolism (Lim et al., 2003; Chapelle et al., 2012; Vanholme et al., 2012; Morreel et al., 2014; Dima et al., 2015). The identification of secondary metabolite phenolic compounds that are involved in lignin biosynthesis pathway have mostly been investigated by liquid chromatography and mass spectrometry. The inflorescence stem and rosette leaf of *Arabidopsis thaliana* have primarily been studied, although roots were occasionally included. (Lim et al., 2003; Chapelle et al., 2012; Vanholme et al., 2012; Morreel et al., 2014; Dima et al., 2015). It was previously reported that knockout of *Arabidopsis BGLU45*, which is expressed in a stem-specific manner, resulted in an increase in coniferin, vanillin 4-O-hexoside, isodihydrodehydroconiferyl alcohol (IDDDC) hexoside, and isodehydroconiferyl alcohol (IDDC) hexoside in the stem, with smaller effects in the root (Chapelle et al., 2012). However, in that scanning LCMS analysis, the contents of syringin, sinapyl alcohol, and coniferyl alcohol were not significantly different. In contrast, our targeted multiple reactant monitoring (MRM) LCMSMS approach showed significant differences in extracts of homozygous *Arabidopsis bglu45-2* mutant compared with the wild type for syringin, coniferin, *p*CAG, and sinapyl alcohol (Figs. 3 and 4). Since only compounds for which a specific MRM method could be developed were detected, we could not detect all of the compounds described by Chapelle et al. (2012). On the other hand,

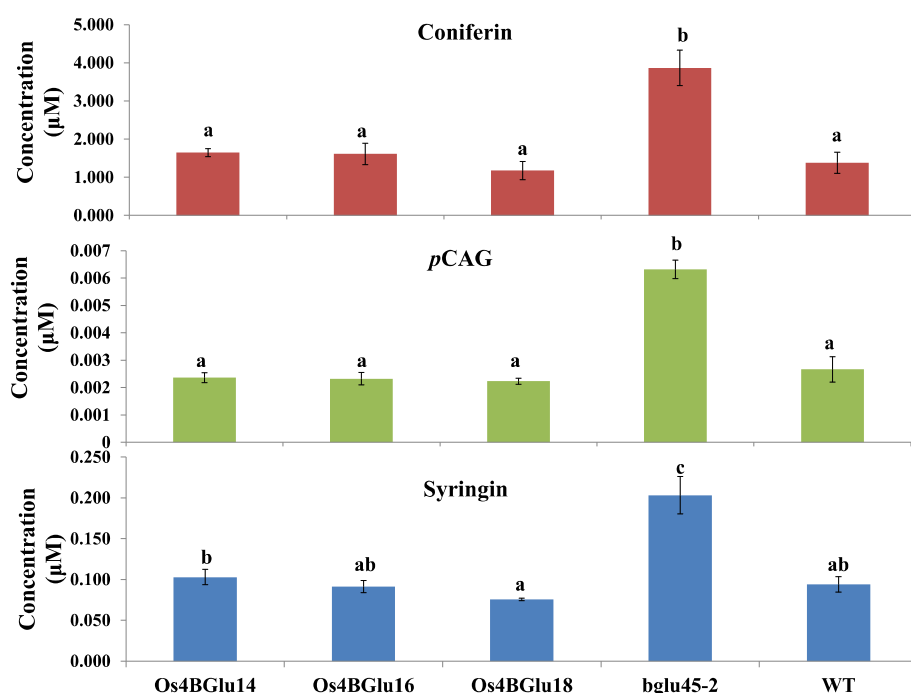


Fig. 3. Concentrations of monolignol glucosides in the extracts of *Arabidopsis bglu45-2* rescued by monolignol β -glucosidases, homozygous *Arabidopsis bglu45-2* mutant and segregated wild type.

The concentrations of each compounds were determined with 2 technical replicates of 4 biological replicates by UPLC-MSMS MRM analysis. Concentrations were determined by comparison to the specific compound standard curve, then divided by the concentration measured for NAA and multiplied by the concentration of NAA expected from the amount added to the sample. Bars marked by the same lower case letter are not significantly different, while bars without the same letter are significantly different ($p < 0.01$).

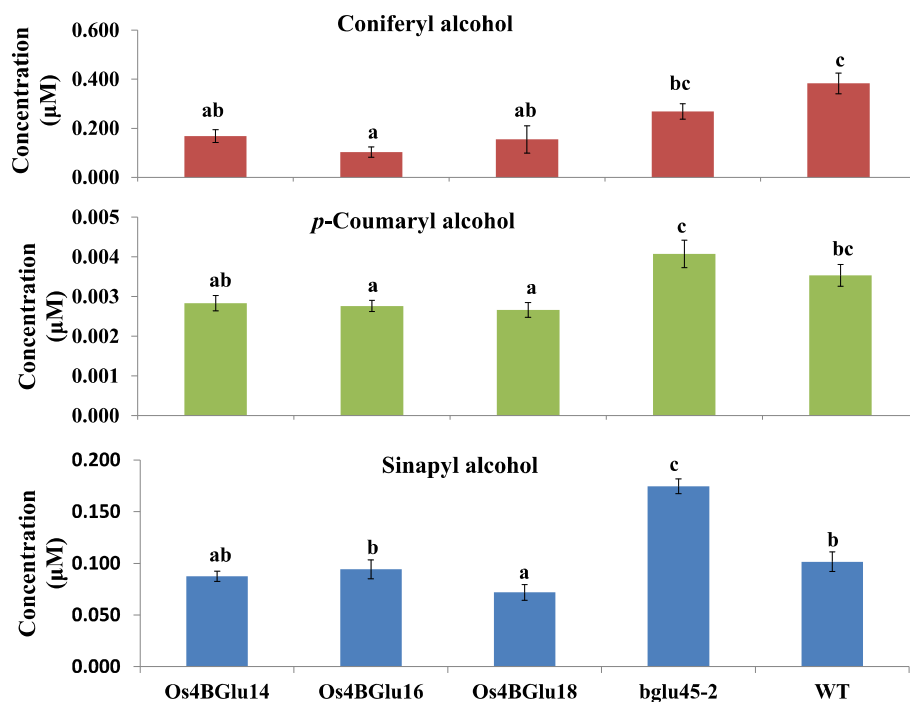


Fig. 4. Concentrations of monolignols in the extracts of *Arabidopsis bglu45-2* rescued by monolignol β -glucosidases, homozygous *Arabidopsis bglu45-2* mutant and segregated wild type. The concentrations of each compounds were determined and relative concentrations calculated as described for Fig. 3. Bars marked by the same lower case letter are not significantly different, while bars without the same letter are significantly different ($p < 0.01$).

since the monitoring of specific fragment ions from specific parent ions in triple quadrupole mass spectrometry is the high precision method for quantifying molecular analytes, it may have provided more sensitivity to the change, compared to the previously utilized ion scanning method. However, we cannot exclude differences in the growth conditions or sample preparation, between our study and that of Chapelle et al. (2012), as contributors to this difference. Nonetheless, the increase in the levels of all three monolignol glucosides in the *bglu45-2* plant extracts is consistent with the activity of recombinant BGLU45 *in vitro* (Escamilla-Treviño et al., 2006).

To rescue coniferin hydrolysis, rice genes encoding monolignol β -glucosidases (Os4BGlu14, Os4BGlu16, and Os4BGlu18) were integrated into the homozygous mutant via *Agrobacterium*-mediated

transformation. Os4BGlu16 and Os4BGlu18 recombinantly expressed in *P. pastoris* and *E. coli* had high catalytic efficiency to hydrolyze syringin and coniferin and lower activity toward pCAG (Baiya et al., 2014). In the current study, extracts of the *Arabidopsis bglu45-2* mutant lines rescued with Os4BGlu16 and Os4BGlu18 had levels of monolignol glucosides that were significantly lower than the *bglu45-2* parent line and similar to the Col-0 wild type plant extracts (Fig. 3).

While Os4BGlu14 lacks the conserved catalytic acid/base residue, it is highly expressed in reproduction stages, such as endosperm, lemma, embryo, pollen, panicle, and flower (Opassiri et al., 2006; Baiya et al., 2014), suggesting that Os4BGlu14 may still function in specific plant cells. The relative amounts of syringin, coniferin, and pCAG in *bglu45-2* rescued by Os4BGlu14 were decreased more than 2-fold, when

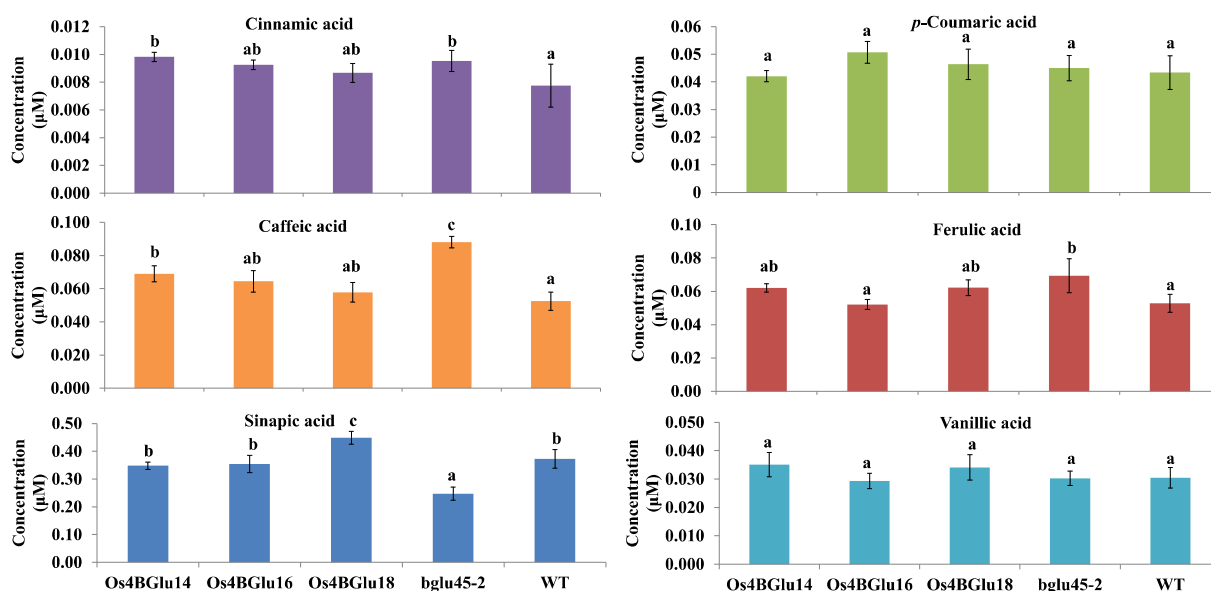


Fig. 5. Concentrations of monolignol intermediates and related compounds in the extracts of *Arabidopsis bglu45-2* rescued by monolignol β -glucosidases, homozygous *Arabidopsis bglu45-2* mutant and segregated wild type.

The concentrations of the compounds were determined and relative concentrations calculated as described in Fig. 3. Bars marked by the same lower case letter are not significantly different, while bars without the same letter are significantly different ($p < 0.01$).

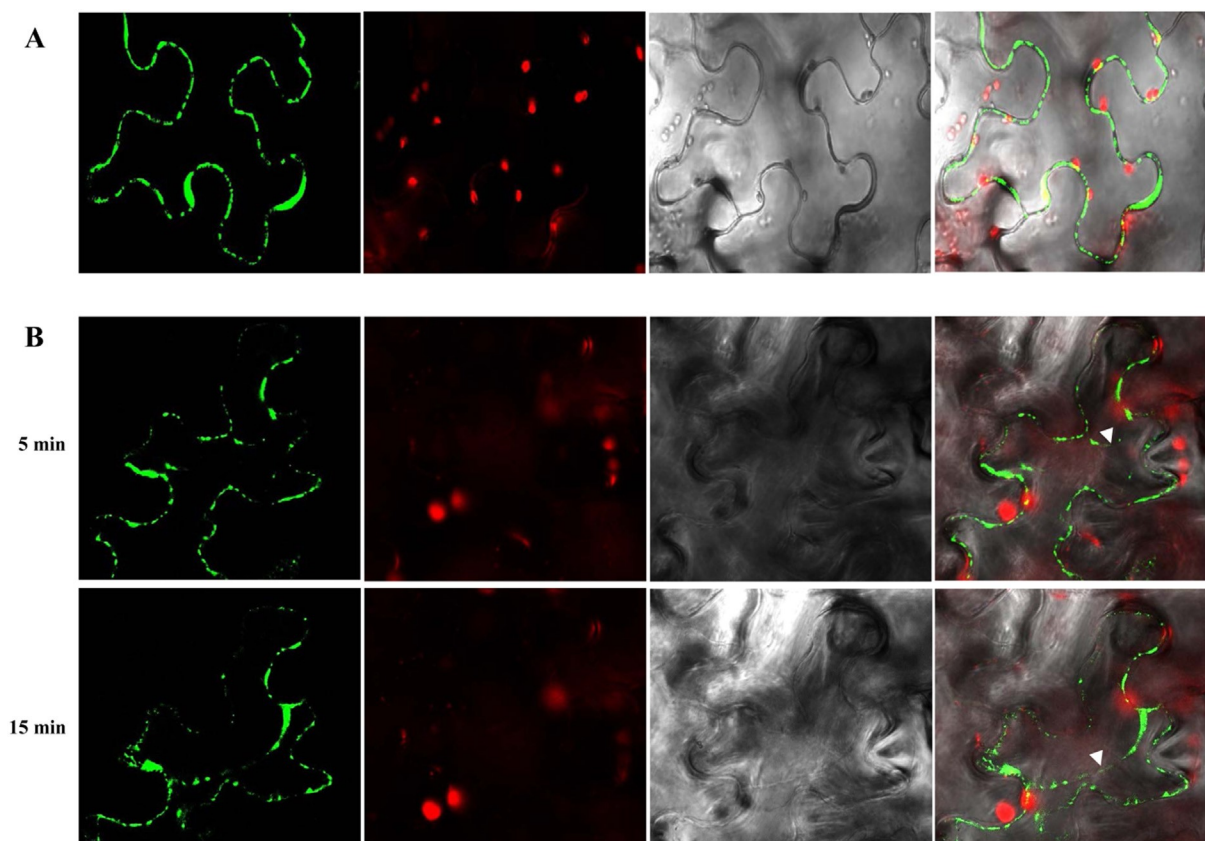


Fig. 6. Subcellular localization of Os4BGlu16-eGFP in tobacco leaf epithelial cells.

(A) Agrobacterium-mediated transient expression of Os4BGlu16-GFP in tobacco leaf epithelial cells. (B) Leaf epithelial cells expressing Os4BGlu16-GFP plasmolyzed by infiltrating 0.8 M mannitol. Microscope images were captured at 5 min (top) and 15 min (bottom) after the infiltration, respectively. Arrowheads indicate the plasma membrane separated from the cell wall. The fluorescent GFP signal, chlorophyll autofluorescence, light microscope view, and merged image are shown from left to right.

compared with homozygous *Arabidopsis bglu45-2* mutant (Fig. 3). This evidence suggests that Os4BGlu14 does serve to release monolignols from their glucosides in the plant cells.

Due to the fact that the amino acid residue in the conserved acid/base position at residue 194 is glutamine rather than glutamate, it is somewhat surprising that Os4BGlu14 expression in the *bglu45-2* mutant *Arabidopsis* yields similar results to expression of the monolignol β -glucosidases Os4BGlu16 and Os4BGlu18. This suggests that the enzyme can somehow function without the catalytic acid/base. Indeed *in vitro* studies of GH1 β -glucosidases in which the acid/base glutamate was replaced with glutamine showed relatively high transferase activity to transfer the glucose to nucleophilic acceptors (Müllegger et al., 2005; Chuenchor et al., 2011). As such, Os4BGlu14 may be acting as a transglucosidase *in vivo*, transferring the glucosyl moiety to organic acids or other suitable acceptors, as has been demonstrated for rice Os9BGlu31 (Luang et al., 2013). However, Os9BGlu31 does not significantly utilize monolignols as substrates, so this would be a unique function for Os4BGlu14.

In contrast to what might have been expected based on their release by the exogenously expressed rice β -glucosidases, sinapyl, coniferyl, and *p*-coumaryl alcohol levels were also low in the rescued lines (Fig. 4). This suggests that upon release from their glucosides by the rice monolignol β -glucosidases, the monolignols may rapidly react to form other compounds. Given that the rice monolignol β -glucosidases are predicted to be extracellular (Opassiri et al., 2006), other monolignol β -glucosidases, including *Arabidopsis* BGLU45 and BGLU46, have been shown to localize to lignifying cell wall (Chapelle et al., 2012), and monolignols from monolignol glucosides have been shown to incorporate in cell walls (Tsuji et al., 2005), lignin is a likely product of

these released monolignols. Surprisingly, the wild type line had the highest amount of coniferyl alcohol, which is the most abundant monolignol precursor for lignin biosynthesis in *Arabidopsis*. This result may be significant, since coniferin was detected at the highest in this work and Chapelle et al. (2012), and coniferyl alcohol-derived G-units are found in high abundance compared to S- and H-units in *Arabidopsis* stem lignin (Dixon et al., 2001). Although syringin and *p*CAG were only detected at low amounts in *Arabidopsis*, they also appeared to be target substrates for the plant expressed rice monolignol β -glucosidases, since the amounts of these substrates in rescued lines was lower than in the homozygous *bglu45-2* mutant line.

The apoplastic localization of OsBGlu16-GFP in tobacco leaf epithelial cells is consistent with the previous localization of pine and *Arabidopsis* monolignol β -glucosidases to the cell wall, and suggests that the other monolignol β -glucosidases are also localized to the apoplast. In contrast, monolignol glucosides have been shown to be specifically transported into the vacuole, while the monolignols themselves are transported through the plasma membrane by ATP-dependent ABC-like transporters (Miao and Liu, 2010). Indeed the storage of glycosylated phenolic compounds in the vacuole has been described in many plants (Leinhos and Savidge, 1993; Santiago et al., 2000; Ferreres et al., 2011; Li et al., 2012; Dima et al., 2015; Le Roy et al., 2016). This "spatial problem" of having the monolignol glucoside and monolignol β -glucosidase in different compartments was noted by Wang et al. (2013), who noted that it suggests a defense role for the monolignol glucosides and β -glucosidases. Defense-related β -glucosidases, such as a specific β -glucosidase for the cyanogenic glycoside dhurrin in sorghum and a specific β -glucosidase for coumarin in *Melilot alba* are also located in different sites from their substrates to prevent mixing before herbivore

attack (Kojima et al., 1979; Oba et al., 1981; Morant et al., 2008). However, the fact that the loss of AtBGLU45 causes a significant increase in monolignol glucosides, while expression of the extracellular rice monolignol β -glucosidases cause a similar decrease in monolignol glucoside levels, suggests that the monolignols and their glucosidases are meeting. Wang et al. (2013) proposed a model where monolignol glucosides may sometimes be transported from the vacuole to the cell wall. On the other hand, plant membrane proteins are sometimes transported through the endosomes and trans-Golgi network to the vacuole (Contento and Bassham, 2012), and a small amount of extracellular β -glucosidase might be able to follow this route to the vacuole for degradation. However, the transport of extracellular β -glucosidases to the vacuole would have to be limited, since rapid release of monolignols and defense compounds from their glucosides in the vacuole would likely be harmful to the plant (Morant et al., 2008; Dima et al., 2015). Whether the monolignol glucosides are cycled through the apoplast at some point, or the monolignol β -glucosidases may be endocytosed into the vacuole and have a significant impact on monolignol glucoside levels is an area for further investigation.

5. Conclusion

In this work, we provided evidence that Os4BGLu14, Os4BGLu16, and Os4BGLu18 break down inactive monolignol glucosides, coniferin, syringin, and pCAG, to their respective active monolignols *in planta*. In our quantitative LCMSMS assay, *Arabidopsis bglu45-2* mutant contained approximately 2-times the levels of coniferin, syringin and pCAG compared to the wild type segregant control. After we introduced rice *Os4BGLu14*, *Os4BGLu16*, and *Os4BGLu18* genes driven by the CaMV 35S promoter to the *bglu45-2* mutant line, the levels of syringin, coniferin, and pCAG decreased to levels similar to the wild type segregant control, indicating that all three genes could compensate for the loss of BGLU45 activity. Moreover, in all three rescued lines, the amounts of sinapyl alcohol and p-coumaric alcohol were also lower than homozygous *bglu45-2* mutant line, suggesting that monolignols released by the exogenously expressed β -glucosidases may rapidly react to form other compounds. The extracellular localization demonstrated for Os4BGLu16 presents an enigma to understand how it lowers the cellular monolignol glucoside stores that are thought to reside primarily in the vacuole, suggesting that our understanding of monolignol glucoside transport and metabolism may be as yet incomplete.

Contributions

SB performed plant extraction, UPLC-MSMS analysis, and wrote the paper. BM contributed plant construction and RT-PCR analysis. SKL conducted the localization experiments. BM, JJ, and JKC designed and advised the experiments. All authors reviewed and proofread the manuscript.

Acknowledgements

This work was supported by grant BRG5980015 from the Thailand Research Fund and Suranaree University of Technology, the National Research University Project of the Commission on Higher Education to Suranaree University of Technology, the Next Generation BioGreen 21 Program of the Rural Development Administration of Korea (PJ013172), and the Basic Science Research Program of National Research Foundation of Korea (NRF-2017R1D1A1B03032724).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.plaphy.2018.03.026>.

References




- Aoki, D., Hanaya, Y., Akita, T., Matsushita, Y., Yoshida, M., Kuroda, K., Yagami, S., Takama, R., Fukushima, K., 2016. Distribution of coniferin in freeze-fixed stem of *Ginkgo biloba* L. by cryo-TOF-SIMS/SEM. *Sci. Rep.* 6, 31525–31533.
- Baiya, S., Hua, Y., Ekkhara, W., Ketudat Cairns, J.R., 2014. Expression and enzymatic properties of rice (*Oryza sativa* L.) monolignol beta-glucosidases. *Plant Sci.* 227, 101–109.
- Baucher, M., Monties, B., Montagu, M.V., Boerjan, W., 1998. Biosynthesis and genetic engineering of lignin. *Crit. Rev. Plant Sci.* 17, 125–197.
- Bednarek, P., Schneider, B., Svatos, A., Oldham, N.J., Hahlbrock, K., 2005. Structural complexity, differential response to infection, and tissue specificity of indolic and phenylpropanoid secondary metabolism in *Arabidopsis* roots. *Plant Physiol.* 138, 1058–1070.
- Boerjan, W., Ralph, J., Baucher, M., 2003. Lignin biosynthesis. *Annu. Rev. Plant Biol.* 54, 519–546.
- Campbell, M.M., Sederoff, R.R., 1996. Variation in lignin content and composition mechanisms of control and implications for the genetic improvement of plants. *Plant Physiol.* 110, 3–13.
- Chapelle, A., Morreel, K., Vanholme, R., Le-Bris, P., Morin, H., Lapierre, C., Boerjan, W., Jouanin, L., Demont-Caulet, N., 2012. Impact of the absence of stem-specific β -glucosidases in lignin and monolignols. *Plant Physiol.* 160, 1204–1217.
- Chuenchor, W., Pengthaisong, S., Robinson, R.C., Yuvaniyama, J., Svasti, J., Ketudat Cairns, J.R., 2011. The structural basis of oligosaccharide binding by rice BGLU1 beta-glucosidase. *J. Struct. Biol.* 173, 169–179.
- Clough, S.J., Bent, A.F., 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735–743.
- Contento, A.L., Bassham, D.C., 2012. Structure and function of endosomes in plant cells. *J. Cell Sci.* 125, 3511–3518.
- Dharmawardhana, D.P., Ellis, B.E., Carlson, J.E., 1995. A β -glucosidase from lodgepole pine xylem specific for the lignin precursor coniferin. *Plant Physiol.* 107, 331–339.
- Dharmawardhana, D.P., Ellis, B.E., 1998. β -Glucosidases and glucosyltransferase in lignifying tissues. *J. Am. Chem. Soc.* 697, 76–83.
- Dima, O., Morreel, K., Vanholme, B., Kim, H., Ralph, J., Boerjan, W., 2015. Small glycosylated lignin oligomers are stored in *Arabidopsis* leaf vacuoles. *Plant Cell* 27, 695–710.
- Dixon, R.A., Chen, F., Guo, D., Parvathi, K., 2001. The biosynthesis of monolignols: a “metabolic grid”, or independent pathways to guaiacyl and syringyl units? *Phytochemistry* 57, 1069–1084.
- Escamilla-Treviño, L.L., Chen, W., Card, M.L., Shih, M.-C., Cheng, C.L., Poulton, J.E., 2006. *Arabidopsis thaliana* β -glucosidases BGLU45 and BGLU46 hydrolyse monolignol glucosides. *Phytochemistry* 67, 1651–1660.
- Ferreres, F., Figueiredo, R., Bettencourt, S., Carqueijeiro, I., Oliveira, J., Gil-Zquierdo, A., Pereira, D.M., Valentão, P., Andrade, P.B., Duarte, P., Barceló, A.R., Sottomayor, M., 2011. Identification of phenolic compounds in isolated vacuoles of the medicinal plant *Catharanthus roseus* and their interaction with vacuolar class III peroxidase: an H₂O₂ affair? *J. Exp. Bot.* 62, 2841–2854.
- Huis, R., Morreel, K., Fliniaux, O., Lucau-Danila, A., Fénart, S., Grec, S., Neutelings, G., Chabbert, B., Mesnard, F., Boerjan, W., Hawkins, S., 2012. Natural hypolignification is associated with extensive oligolignol accumulation in flax stems. *Plant Physiol.* 158, 1893–1915.
- Karimi, M., Inzé, D., Depicker, A., 2002. GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* 7, 193–195.
- Ketudat Cairns, J.R., Mahong, B., Baiya, S., Jeon, J.S., 2015. β -Glucosidases: multitasking, moonlighting or just misunderstood? *Plant Sci.* 241, 246–259.
- Kojima, M., Poulton, J.E., Thayer, S.S., Conn, E.E., 1979. Tissue distributions of ghrurin and of enzymes involved in its metabolism in leaves of *Sorghum bicolor*. *Plant Physiol.* 63, 1022–1028.
- Lanot, A., Hodge, D., Jackson, R.G., George, G.L., Elias, L., Lim, E.-K., Vaistij, F.-E., Bowles, D.J., 2006. The glucosyltransferase UGT72E2 is responsible for monolignol 4-O-glucoside production in *Arabidopsis thaliana*. *Plant J.* 48, 286–295.
- Le Roy, J., Huss, B., Creach, A., Hawkins, S., Neutelings, G., 2016. Glycosylation is a major regulator phenylpropanoid availability and biological activity in plants. *Front. Plant Sci.* 7, 1–19.
- Leinhos, V., Savidge, R.A., 1993. Isolation of protoplasts from developing xylem of *Pinus banksiana* and *Pinus strobus*. *Can. J. For. Res.* 23, 343–348.
- Li, Z., Tang, T., Liang, S., Ning, X., Bai, M., Wu, H., 2012. The synthesis and storage sites of phenolic compounds in the root and rhizome of *Echinacea purpurea*. *AJPS (Asian J. Plant Sci.)* 3, 551–558.
- Lim, E., Higgins, G.S., Li, Y., Bowles, D.J., 2003. Regioselectivity of glucosylation of caffeic acid by a UDP-glucose:glucosyltransferase is maintained *in planta*. *Biochem. J.* 373, 987–992.
- Lin, J.S., Huang, X.X., Li, Q., Cao, Y., Bao, Y., Meng, X.F., Li, Y.J., Fu, C., Hou, B.K., 2016. UDP-glucosyltransferase 72B1 catalyzes the glucose conjugation of monolignols and is essential for the normal cell wall lignification in *Arabidopsis thaliana*. *Plant J.* 88, 26–42.
- Luang, S., Cho, J.-I., Mahong, B., Opasiri, R., Akiyama, T., Phasai, K., Komvongsa, J., Sasaki, N., Hua, Y., Matsuba, Y., Ozeki, Y., Jeon, J.-S., Ketudat Cairns, J.R., 2013. Os9BGLu31 is a transglucosidase with the capacity to equilibrate phenolpropenoid, flavonoid and phytohormone glycoconjugates. *J. Biol. Chem.* 288, 10111–10123.
- Miao, Y., Liu, C., 2010. ATP-binding cassette-like transporters are involved in the transport of lignin precursors across plasma and vacuolar membranes. *Proc. Natl. Acad. Sci. U.S.A.* 107, 22728–22733.
- Moellerling, E.R., Muthan, B., Benning, C., 2010. Freezing tolerance in plants requires lipid remodeling at the outer chloroplast membrane. *Science* 330, 226–228.

- Morant, A.V., Jorgensen, K., Jorgensen, C., Paquette, S.M., Sanchez-Perez, R., Moller, B.L., 2008. β -Glucosidases as detonators of plant chemical defense. *Phytochemistry* 69, 1795–1813.
- Morreel, K., Saeys, Y., Dima, O., Lu, F., Van de Peer, Y., Vanholme, R., Ralph, J., Vanholme, B., Boerjan, W., 2014. Systematic structural characterization of metabolites in *Arabidopsis* via candidate substrate-product pair networks. *Plant Cell* 26, 929–945.
- Müllegger, J., Jahn, M., Chen, H.-M., Warren, R.A.J., Withers, S.G., 2005. Engineering of a thioglycoligase: randomized mutagenesis of the acid-base residue leads to the identification of improved catalysts. *Protein Eng. Des. Sel.* 18, 33–40.
- Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T., Kimura, T., 2007. Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J. Biosci. Bioeng.* 104, 34–41.
- Oba, K., Conn, E.E., Canut, H., Boudet, A.M., 1981. Subcellular localization of 2-(β -D-glucosyloxy)-cinnamic acids and the related β -glucosidase in leaves of *Melilotus alba* Desr. *Plant Physiol.* 68, 1359–1363.
- Opassiri, R., Pomthong, B., Onkokoong, T., Akiyama, T., Esen, A., Ketudat Cairns, J.R., 2006. Analysis of rice glycosyl hydrolase family 1 and expression of Os4bglu12 β -glucosidase. *BMC Plant Biol.* 6, 33–51.
- Oppenheimer, D.G., Haas, N., Silflow, C.D., Snustad, D.P., 1988. The β -tubulin gene family of *Arabidopsis thaliana*: Preferential accumulation of the β 1 transcript in roots. *Gene* 63, 87–102.
- Rouyi, C., Baiya, S., Lee, S.-K., Mahong, B., Jeon, J.-S., Ketudat-Cairns, J.R., Ketudat-Cairns, M., 2014. Recombinant expression and characterization of the cytoplasmic rice β -glucosidase Os1BGlu4. *PLoS One* 9, e96712.
- Santiago, L.J.M., Louro, R.P., De Oliveira, D.E., 2000. Compartmentation of phenolic compounds and phenylalanine ammonia-lyase in leaves of *Phyllanthus tenellus* Roxb. And their induction by copper sulphate. *Ann. Bot.* 86, 1023–1032.
- Steeves, V., Förster, H., Pommer, U., Savidge, R., 2001. Coniferyl alcohol metabolism in conifers-I. Glucosidic turnover of cinnamyl aldehydes by UDPG: coniferyl alcohol glucosyltransferase from pine cambium. *Phytochemistry* 57, 1085–1093.
- Terashima, N., Ko, C., Matsushita, Y., Westermark, U., 2016. Monolignol glucosides as intermediate compounds in lignin biosynthesis. Revisiting the cell wall lignification and new ^{13}C -tracer experiments with *Ginkgo biloba* and *Magnolia liliiflora*. *Holzforschung* 70, 801–810.
- Tsuji, Y., Chen, F., Yasuda, S., Fukushima, K., 2005. Unexpected behavior of coniferin in lignin biosynthesis of *Ginkgo biloba* L. *Planta* 222, 58–69.
- Tsuyama, T., Takabe, K., 2015. Coniferin β -glucosidase is ionically bound to cell wall in differentiating xylem of poplar. *J. Wood Sci.* 61, 438–444.
- Vanholme, R., Storme, V., Vanholme, B., Sundin, L., Christensen, J.H., Goeminne, G., Halpin, C., Rohde, A., Morreel, K., Boerjan, W., 2012. A systems biology view of responses to lignin biosynthesis perturbations in *Arabidopsis*. *Plant Cell* 24, 3506–3529.
- Wang, Y., Chantreau, M., Sibout, R., Hawkins, S., 2013. Plant cell wall lignification and monolignol metabolism. *Front. Plant Sci.* 4, 220.
- Xu, Z., Escamilla-Treviño, L., Zeng, L., Lalgondar, M., Bevan, D., Winkel, B., Mohamed, A., Cheng, C.L., Shih, M.C., Poulton, J., Esen, A., 2004. Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. *Plant Mol. Biol.* 55, 343–367.



Cite this: *Green Chem.*, 2019, **21**, 2823

Engineering faster transglycosidases and their acceptor specificity†

Linh T. Tran,  ‡^{a,b} Vincent Blay,  *^{c,d} Sukanya Luang,^e Chatchakorn Eurtivong,^f Sunaree Choknud,^{a,b} Humbert González-Díaz^{i,h} and James R. Ketudat Cairns  *^{a,b,g}

Transglycosidases are enzymes that have the potential to catalyze the synthesis of a wide range of high-value compounds starting from biomass-derived feedstocks. Improving their activity and broadening the substrate range are important goals to enable the widespread application of this family of biocatalysts. In this work, we engineered 20 mutants of the rice transglycosidase Os9BGlu31 and evaluated their catalysis in 462 reactions over 18 different substrates. This allowed us to identify mutants that expanded their substrate range and showed high activity, including W243L and W243N. We also developed double mutants that show very high activity on certain substrates and exceptional specificity towards hydrolysis, such as L241D/W243N. In order to guide a more general use of Os9BGlu31 variants as transglycosylation catalysts, we built cheminformatics models based on topological descriptors of the substrates. These models showed useful predictive potential on the external validation set and are allowing the identification of efficient catalytic routes to novel phytohormone and antibiotic glucoconjugates of interest.

Received 19th February 2019,
Accepted 24th April 2019

DOI: 10.1039/c9gc00621d

rs.c.li/greenchem

^aSchool of Chemistry, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand. E-mail: cairns@sut.ac.th

^bCenter for Biomolecular Structure, Function and Application, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand

^cFisher College of Business, The Ohio State University, Gerlach Hall, 2108 Neil Ave. Columbus, OH 43210, USA. E-mail: blayroger.1@osu.edu

^dMolecular Topology & Drug Design Research Unit, Departamento de Química Física, Universitat de València, Av. V. A. Estellés, s/n, 46100 Burjassot, Spain

^eDepartment of Biochemistry, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand

^fProgram of Chemical Biology, Chulabhorn Graduate Institute, Chulabhorn Royal Academy of Science, Bangkok 10210, Thailand

^gDepartment of Organic Chemistry II, University of the Basque Country UPV/EHU, 48940 Leioa, Spain

^hIKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain

ⁱLaboratory of Biochemistry, Chulabhorn Research Institute, Bangkok 10210, Thailand

†Electronic supplementary information (ESI) available: Fig. S1 illustrates the data analysis pipeline used. Chromatograms in Fig. S2 to S7 show the transfer of glucose to different acceptors by wild type Os9BGlu31 and its mutants. Fig. S8 demonstrates the synthesis of the phytohormone glucoconjugate abscisic acid glucose ester enabled by this study. Table S1 presents the sequences of the oligonucleotides used as primers in the site-directed mutagenesis of Os9BGlu31. Table S2 indicates the docking scores on the wild type and W243L mutant. Table S3 indicates the names of the descriptors shortlisted. Table S4 is the correlation matrix between selected descriptors. Table S5 presents ensemble activity predictions for potential new transglucosylation acceptors. File SI01.xlsx contains the activity results from the screening of all mutants and the epistasis analysis. File SI02.xlsx presents the data from the high activity-mutants used to train the ML models. File SI03.xlsx contains cheminformatics data (descriptors and results of the ANNs). See DOI: 10.1039/c9gc00621d

‡Current Address: Research Institute for Interdisciplinary Science, Okayama University, Okayama, Japan.

*These authors contributed equally.

1. Introduction

Glycosylation is an essential mechanism for building structural and bioactive components of cells and their matrices and for providing blocking groups that restrain reactive and bioactive molecules from disrupting the metabolism of the cells in which they are synthesized.¹ Understanding the roles of the various glycosides in nature requires their synthesis or purification from natural sources, where they are often found in minute quantities. Moreover, the glycosylation of some bioactive compounds may modulate their uptake and bioactivity, which holds interest for the fine chemical and pharmaceutical industries. The large-scale synthesis of these glucoconjugates may require complicated strategies or expensive starting materials, while their extraction and purification may require large amounts of solvents. Therefore a simple and efficient system to produce such compounds in aqueous solvents is highly desirable.

Generally, glycosylation begins with the activation of a sugar by its attachment to a nucleotide, yielding nucleotide sugars, such as UDP-glucose (uridine diphosphate- α -D-glucopyranoside). The sugar can then be transferred by a Leloir-type glycosyltransferase (GT) to a carbohydrate or noncarbohydrate molecule to form a glycoside.² Glycosides and carbohydrates are broken down by cleavage of the glycosidic bond created by the GTs by glycoside hydrolases (GHs), usually by hydrolysis. However, many GHs are able to catalyze transglycosylation, that is, the transfer of the glycosyl moiety to another carbo-

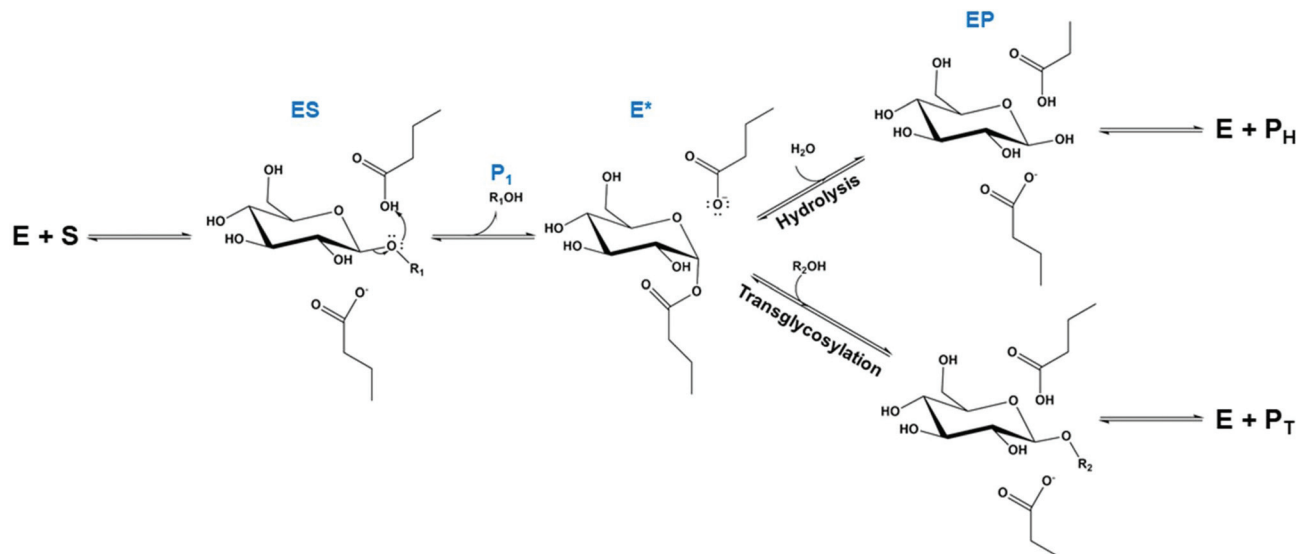


Fig. 1 Mechanism of transglycosylation and hydrolysis reactions in retaining a GH and GT. The enzyme complexes are labeled with the symbols used in the general nomenclature for a ping-pong bi-bi mechanism. The glycosyl intermediate is labeled E*. The second substrate is water in the case of hydrolysis or a hydroxyl group (R_2OH) in the case of transglycosylation. The glycon moiety shown is glucose, in line with the transglucosidase activity described in the current paper.

hydrate or aglycon moiety (Fig. 1). Some enzymes related to retaining GHs act as transglycosidases (TGs, also referred to as non-Leloir-type GTs), which preferentially catalyze transglycosylation rather than hydrolysis.

The transglycosidases that have been studied to date belong to families of retaining GHs and are believed to catalyze transglycosylation by the same mechanism used by the related GHs, shown in Fig. 1. The retaining mechanism of GHs (and GTs) starts with a nucleophilic residue in the active site attacking the anomeric carbon. At the same time, the departure of the aglycon moiety is facilitated by a general acid in the active site. Since the glucosyl moiety becomes covalently bound to the nucleophilic residue, this first step is called glycosylation.^{3,4} In the second step (deglycosylation), a nucleophilic acceptor molecule can attack from the same side from where the aglycon departed. This is a ping-pong bi-bi mechanism, in which two S_N2 reactions, with an inversion of stereochemistry at each step, return the anomeric carbon to its original stereochemical configuration. In the case of hydrolysis, water acts as the nucleophile in the deglycosylation step, whereas in transglycosylation another molecule acts as the nucleophile.

As noted above, TGs are related to GHs and have been included in the GH families grouped by their amino acid sequence similarity at the carbohydrate active enzyme database (<http://www.cazy.org>). Most of the transglycosidases that have been described catalyze synthesis and remodeling of oligo- and polysaccharides. For instance, xyloglucan endotransferases (XET) belong to the family GH16, which also contains xyloglucan endohydrolases. Differences in a surface loop in the XET and the xyloglucan endohydrolases have been shown to contribute to the transferase *versus* hydrolase specificity,⁵ but this principle cannot be readily transferred to other

families. Several attempts have been made to engineer GHs to make TGs, including the production of a β -transglucosidase from a bacterial GH1 β -glycosidase,⁶ a *trans*- α -L-arabinofuranosidase from GH51,⁷ and α -glucosyl transferases for the production of oligosaccharide antigens.^{8,9} In each of these cases, the target products were oligosaccharides, whereas relatively little work has been done on the use of TGs to synthesize glycosides.

The use of directed evolution approaches allowed the identification of changes that led to higher transglycosylation to hydrolysis ratios in the GHs that were converted to TGs.^{6,7} While it was anticipated that decreasing the binding of water and increasing sugar binding could increase transglycosylation, it was surmised that several of the mutations actually worked by decreasing the function of the catalytic nucleophile or acid/base. This was hypothesized to prolong the lifetime of the intermediate by destabilizing the transition states of the two reaction steps (Fig. 1) and by decreasing the ability of water, a weaker nucleophile that requires deprotonation, to compete with sugars as the acceptor in the deglycosylation step. This strategy resembles the action of glycosynthase and thioglycosylase mutations of retaining GHs, in which the nucleophile and acid/base, respectively, are mutated to non-nucleophilic, non-ionizable residues, thereby allowing the transfer of the glycon from activated donor glycosides to suitable sugar acceptors without hydrolysis.^{10,11}

In the last several years, some GH1 enzymes that catalyze the transglycosylation of lipids, anthocyanins, and other non-carbohydrate small molecules have been described.^{12–16} The *sensitive to freezing 2* (*SFR2*) gene in *Arabidopsis thaliana* was found to encode a galactolipid/galactolipid galactosyltransferase. Several anthocyanin glucosyltransferases were found to

actually be GH1 TG,^{13,15,16} whereas rice (*Oryza sativa*) Os9BGlu31 was found to be a general TG that can transfer glucose among phenolic acids, phytohormones, and flavonoids and it also appears to deglycosylate fatty acids and other substrates in the plant.^{14,17} Little engineering of these types of enzymes has been reported, although we recently demonstrated that mutation of the active site cleft residue Trp243 (W243) to Asn (W243N) increased the activity and broadened the specificity of Os9BGlu31.¹⁸

Promiscuous GTs have been used to glycosylate several medicinal compounds to modulate their solubility and bioactivity.¹⁹ For instance, several microbial antibiotics are glycosides and these promiscuous GTs have allowed the transfer of unusual sugars in glycodiversification to develop new or more robust activities.^{20,21} From an industrial point of view, the development of transglycosidase (bio)catalysts would be desirable because they do not require a nucleotide sugar substrate or intermediate in the transfer of sugars to compounds of interest. Ideally, the sugars could be obtained from biomass.^{22,23} Recently, cyclodextrin glucosyltransferases from *Thermoanaerobacterium* sp. have been reported to transfer glucose onto aryl glucopyranosides and furanosides, achieving an unusual substrate specificity toward alkyl furanosides.²⁴

Although most of the transglycosidases studied to date show a high specificity for the transfer of the glucosyl moiety to certain sugars or aglycon groups, the promiscuity of Os9BGlu31 makes it especially promising to aid in the synthesis of a wide range of compounds.^{14,18} Nonetheless, wild-type Os9BGlu31 cannot efficiently transfer glucosyl groups to all acceptors; therefore new variants with broader or different specificities are of interest to increase its potential as a biosynthetic tool. The development of catalysts that work under mild conditions with environmentally friendly reagents would greatly facilitate the production of compounds for studies of phytohormone metabolism¹ and antibiotic glycosides with modified pharmacokinetic properties.¹⁹

In this work, we have explored the potential of rice Os9BGlu31 active site cleft mutations to broaden its acceptor potential by engineering and evaluating 14 different amino acids at residue position 243. Moreover, we combined the mutant W243N with hydrophilic mutations of residues in positions that could interact with water during hydrolysis and evaluated their effect on the hydrolysis and transglycosylation specificity. Since the structure of Os9BGlu31 has not been solved yet, homology models of the covalent glycosyl-enzyme intermediate were made to evaluate the interaction of acceptor substrates with the active site. Although docking into such models provides insight into possible modes of binding, it could not quantitatively predict activity differences, due to the low resolution of the models. Therefore, we have applied cheminformatics tools to obtain further insights and anticipate the behavior of the different transglycosidases when presenting new substrates.

There are two main possible approaches to build a cheminformatics model on experimental results. In the absence of a mechanistic hypothesis, it is possible to compute many mole-

cular descriptors and then select statistically a phenomenological route that may be possible to interpret *a posteriori*.^{25,26} Alternatively, one can hypothesize a mechanistic route that selects a reduced number of physicochemical descriptors beforehand to build a model with them.^{27,28} In this work, we explored both approaches and built artificial neural network (ANN) models on a subset of physicochemical descriptors that can predict the activity of multiple enzymes against new substrates of interest.

2. Materials and methods

2.1 Construction of pET32a/DEST/TEV/Os9BGlu31

The wild-type pET32a/DEST/TEV/Os9BGlu31 expression vector was constructed as follows. The cDNA fragment encoding the mature Os9BGlu31 gene was amplified with the AK121679F and AK121679R primers indicated in Table S1† by *Pfu* DNA polymerase with the AK121679 clone plasmid as the template. The PCR product (~1.5 kb) was purified from the agarose gel, cloned into the pENTR/TEV/D-TOPO Gateway entry vector (Invitrogen) and incubated at 22 °C for 18 h. The entry clone size was checked by *SacI* restriction endonuclease digestion. Then, pENTR/TEV/D-TOPO/Os9BGlu31 was recombined into the pET32a/DEST destination vector²⁹ by a Gateway LR Clonase (Invitrogen) reaction and cloned in *Escherichia coli* strain DH5α, selected on 50 µg ml⁻¹ ampicillin LB agar. The recombinant expression vector clone was verified to contain the inserted gene by digestion with *EcoRI* restriction endonuclease and DNA sequencing (Macrogen Corp.).

2.2 Site-directed mutagenesis of Os9BGlu31

The tryptophan residue W243 in wild-type Os9BGlu31 was changed to 14 different residues by QuikChange site-directed mutagenesis (Agilent) with the pET32a/DEST/TEV/Os9BGlu31 vector as the template and the primers shown in Table S1.† These mutants comprise cysteine (C), glutamic acid (E), glycine (G), histidine (H), isoleucine (I), lysine (K), leucine (L), asparagine (N), proline (P), glutamine (Q), arginine (R), serine (S), threonine (T) and valine (V). Furthermore, double mutants I172T/W243N, L183Q/W243N and L241D/W243N were prepared by introducing one of the mutations I172T, L183Q or L241D in the vector, in addition to W243N. The sequences of all mutant clones were confirmed by DNA sequencing (Macrogen Corp.).

2.3 Expression and purification of Os9BGlu31 and its mutants

The pET32a/DEST/TEV/Os9BGlu31 wild-type plasmid and all the mutants were transformed into Origami B(DE3) and the proteins expressed, extracted, and purified by an initial immobilized metal ion affinity chromatography (IMAC) step as previously described for the pET32a/DEST/Os9BGlu31 expression vector.^{14,18} All of the purified fractions were assayed in 150 µl total volume with 2 mM 4-nitrophenyl β-D-glucopyranoside (4NPGlc) in 50 mM acetate buffer, pH 4.5, at 30 °C for 30 min. The reactions were stopped by adding 75 µl of 2 M

Na_2CO_3 and the absorbance at 405 nm was then measured. The fractions showing activity with 4NPGlc were pooled and the imidazole was removed by buffer exchange with equilibration buffer (150 mM NaCl in 20 mM Tris HCl, pH 8.0) in 30 kDa molecular weight cutoff (MWCO) centrifugal filters. The N-terminal fusion tag was removed by cleavage with 1 mg TEV protease per 50 mg of the fusion protein at 4 °C for 16 h. Then, the digested proteins were loaded onto a second IMAC column in equilibration buffer, and the flow-through fractions showing activity to cleave 4NPGlc were pooled and concentrated with 30 kDa MWCO centrifugal filters.

2.4 Relative activities of wild type Os9BGlu31 and mutants

The activities of Os9BGlu31 and its mutants were compared upon a range of acceptor substrates for transglycosylation and water was used as the acceptor for hydrolysis in 50 mM citrate buffer (pH 4.5). Unless stated otherwise, the enzymatic assays were set up with 0.25 mM acceptor, 2.5 vol% dimethyl sulfoxide (DMSO), 5 mM 4NPGlc as the donor, and various amounts of wild-type Os9BGlu31 and its mutants so that conversion of the limiting substrate was below 10% (ESI files SI01.xlsx and SI02.xlsx†). The reactions were conducted at 30 °C for 15 min, during which the concentrations of products evolved linearly with time, and they were stopped by the addition of 1% formic acid. The reaction mixtures were centrifuged at 10 000g for 10 min to remove the enzymes, and the supernatants were evaluated by reverse-phase UPLC, as previously described.¹⁸ Briefly, 2 μL of the reaction mixtures were injected into a ZORBAX SB-C18 (1.8 μm , 2.1 \times 150 mm) column equilibrated in 95% solvent A (0.2% formic acid in water) and 5% solvent B (0.2% formic acid in acetonitrile) in an Agilent 1290 UPLC system with a diode-array detector (DAD). The compounds were eluted by a linear gradient from 5% to 50% B (v/v) for 13 min, 50% to 70% B (v/v) for 1 min, and 70% to 5% B (v/v) for 2 min, at a flow rate of 0.3 ml min⁻¹. Relative activities were evaluated from the absorbance at 360 nm of the 4-nitrophenol (4NP) released, which eluted at 10.5 min.

2.5 Transglycosylation of substrates to multiple glucoconjugates by Os9BGlu31 variants

To identify differences in glycosylation on one acceptor that has multiple nucleophilic groups, the products of transglycosylation by wild type Os9BGlu31 and W243 mutants (C, I, L, N, T, and V) were determined in the reaction with 5 mM 4NPGlc as the donor, 0.5 mM acceptor, and 5 μg of proteins in 50 mM citrate buffer (pH 4.5). The reactions were conducted at 30 °C for 1 hour and then stopped with 1% formic acid. The mixtures were injected into the UPLC-DAD as described in section 2.4. Absorbance was measured at wavelengths between 190 and 500 nm to detect the glucoconjugate compounds.

2.6 Linear cheminformatics model

The glucosyl acceptors were characterized by 1D, 2D, and 3D molecular descriptors of different classes (constitutional, molecular format, autocorrelation, Basak, burden, connectivity, topological, charge descriptors, etc.). 1132 descriptors

were computed using the software ChemDes,³⁰ which integrates multiple state-of-the-art packages (Pybel, CDK, RDKit, Chemopy, etc.). 1824 additional molecular descriptors were computed with the newly released software Mordred,³¹ which computes 3D conformers by MM optimization and implements fully revised algorithms for popular indices found in Dragon and PaDEL.

To select the descriptors that will participate in the model, we first removed any index with zero variance across the range of substrates studied. We then reduced the set of descriptors under consideration by looking at the pairwise correlation between all of them. The correlation coefficient matrix was processed by the *findCorrelation* function in the *caret* R package.³² We established the cutoff so that no pair of descriptors had a correlation coefficient >0.8. When a pair of descriptors has a correlation above the cutoff, *findCorrelation* computes the mean absolute correlation of each descriptor in the pair with all others and removes the one with the largest mean absolute correlation. This step reduces the redundancy and the number of descriptors under consideration. In a second step, we ranked the descriptors using as the criterion the information gain as defined in the *FSelector* package³³ and the Pearson correlation of each individual descriptor with the average enzymatic activity measured in triplicate experiments. The top descriptors were then considered by a forward stepwise selection algorithm (*StepAIC*, *MASS* package) to build the multiple linear regression model.³⁴ A scheme of the pipeline devised is shown in Fig. S1.†

2.7 Nonlinear ANN models

The molecular descriptors used in the nonlinear models are D_1 = ALOGP (an estimate of log P), D_2 = MR (molar refractivity), and D_3 = TPSA (topological polar surface area), which can be found in file SI03.xlsx.† STATISTICA 10 was used to implement Artificial Neural Networks (ANN) as nonlinear ML models. The ANNs tested are Multi-Layer Perceptrons (MLP) with one dense hidden layer. The MLP models have up to 7 hidden neurons in the hidden layer. Different activation functions were evaluated for the hidden and output layers. The models were trained within 200 cycles of the Broyden-Fletcher-Goldfarb-Shanno (BFGS) backpropagation algorithm.

2.8 Docking study

The Os9BGlu31 model was calculated in MODELLER 9.19³⁵ based on the structure of Os3BGlu6 in a covalent complex with 2-fluoroglucoside (PDB ID: 3GNR).³⁶ To understand the structure of Os9BGlu31 in a complex with glucose, the structure of Os3BGlu6 in the complex with a 2-fluoroglucoside structure was modified by replacing the fluorine at C'2 with oxygen. Five models each of Os9BGlu31 and Os9BGlu31/glucose complexes were generated from the Os3BGlu6 covalent complex with 2-fluoroglucoside. The models with the lowest scores of the MODELLER Objective Function (MOF)³⁷ were selected. Os9BGlu31 mutants with and without glucose were generated from the Os9BGlu31 wild type models using FoldX³⁸ after insertion of the covalently linked α -D-glucoside coordinates

into the models with glucose. The final models were evaluated for quality using the ProSa2003³⁹ and PROCHECK software.⁴⁰

The acceptor substrate 3D structures were created with the drawing and manipulation tools in the Scigress 3.3.2 software suite. The homology models were protonated. The acceptor substrate structures and the newly added protons, glucosides, and mutated residues were then subjected to energy minimization using the MM2 force-field. The Genetic Optimization for Ligand Docking (GOLD) software version 5.6.2 was used for the docking calculations. The center of binding was defined at coordinates ($x = -0.581$, $y = -6.381$, $z = -21.102$) with a 10 Å radius. Thirty docking runs were allowed for each ligand with default search efficiency (100%). Lysine and arginine were defined as protonated; aspartic and glutamic acid were considered deprotonated. The GoldScore scoring function was used to evaluate the predicted binding modes.⁴¹

3. Results and discussion

3.1 Activity of wild-type Os9BGlu31 and mutants

Previously, we showed that wild-type Os9BGlu31 and several mutants are able to catalyze transglycosylation, transferring a

glucosyl moiety from a donor substrate such as 4NPGlc to an acceptor.^{14,18} In this study, we studied the effect of amino acid substitutions in the position W243 which complete the set of 20 natural amino acids at this position, when combined with our previous work.¹⁸ These included the W243N mutant that showed the highest activity in our previous study and 13 new mutations (W243C, E, G, H, I, K, L, P, Q, R, S, T, and V). We also introduced the second point mutations I172T, L183Q, and L241D in the W243N mutant with the goal to test whether multiple hydrophilic substitutions increase the hydrolysis to transglycosylation ratio and broaden the substrate/product specificity to offer novel biosynthetic possibilities.

Wild type Os9BGlu31 and its mutants were evaluated in the glycosylation of 22 acceptors, including water (for hydrolysis). The relative activities of wild type Os9BGlu31 and its mutants are measured by quantifying the release of 4NP from the reaction of 4NPGlc with the different acceptors. The activity results are presented in Fig. 2. Details on the activities are provided in the ESI file SI01.xlsx.[†]

Among the compounds tested, the preferred acceptor for the wild-type Os9BGlu31 is ferulic acid, while caffeic acid and 1-naphthaleneacetic acid are the substrates preferred by other Os9BGlu31 variants (Fig. 2). Most of the Os9BGlu31 variants

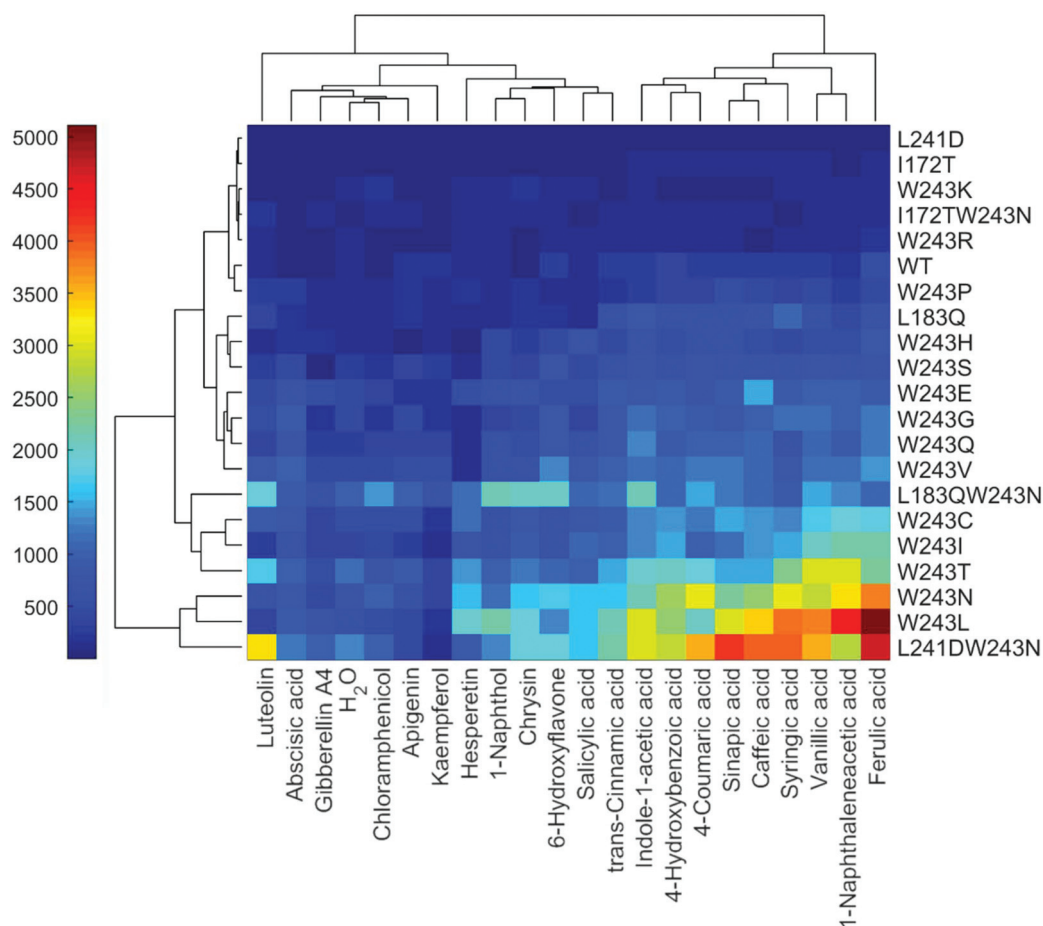


Fig. 2 Heatmap of the 4NP release (in $\text{nmol min}^{-1} \text{mg}^{-1}$) catalyzed by 21 Os9BGlu31 variants when glycosylating 22 different acceptors (462 reactions). The dendrograms were constructed using the average Euclidian distance between clusters as the linkage method. For reaction conditions see SI01.xlsx.[†]

Table 1 Reaction rates (in nmol min⁻¹ mg⁻¹) and types of epistatic interactions observed for the double mutants in this work. For reaction conditions see SI01.xlsx

Acceptor	Wild type	L183Q	L241D	I172T	W243N	L183Q/W243N/ epistasis	L241D/W243N/ epistasis	I172T/W243N/ epistasis
1-Naphthaleneacetic acid	185	40	16	68	3300	1344/–	2750	108
1-Naphthol	84	12	4	33	1186	2104/(sign)	1316	88
4-Coumaric acid	275	26	15	82	3108	1482/–	3515	85
4-Hydroxybenzoic acid	324	29	15	82	2592	1066/–	2869	85
6-Hydroxyflavone	286	14	14	43	1703	1996	1877	134
Absciscic acid	60	17	15	45	861	911	1208/(sign)	78
Apigenin	165	12	0	33	603	1004/(sign)	407	83
Caffeic acid	252	29	15	114	2544	1081/–	3957/(sign)	94
Chloramphenicol	55	10	0	61	1005	1401/(sign)	993	156
Chrysin	67	18	14	43	1655	1996	1877	134
Ferulic acid	550	37	24	99	3759	1088/–	4674/+	115
Gibberellin A4	77	8	15	34	479	591/+	1029/(sign)	110
Hesperetin	92	24	0	34	1521	1167	957/–	84
Indole-1-acetic acid	295	27	0	89	2152	2092	2977/(sign)	135
Kaempferol	210	12	0	25	321	653/(sign)	105	64
Luteolin	141	19	0	53	708	1913/(sign)	3348/(sign)	225
Salicylic acid	138	17	15	35	1674	1050/–	1638	78
Sinapic acid	251	32	0	82	2276	1198/–	4220/(sign)	82
Syringic acid	312	28	6	81	3051	897/–	3933/(sign)	62
<i>trans</i> -Cinnamic acid	275	20	6	72	1617	1131	2188/(sign)	92
Vanillic acid	318	36	15	92	2812	1482/–	3515	85
Water	103	11	6	41	727	968/(sign)	1316/(sign)	75

generated yield higher rates of 4NP release with phenolic acceptors than with buffer alone, which results in hydrolysis by water. Among the single mutants, W243N is arguably the most active overall. W243N also displays an increased activity with most of the substrates, although none is particularly favored, which agrees with previous results.¹⁸ These mutations increase the activity of the enzyme and its potential use for the glycosylation of substrates of interest.

The mutants I172T, L183Q, and L241D show lower activity and stability than the wild type or W243N. The less soluble L241D mutant, in particular, shows the lowest performance over multiple acceptors (Fig. 2). Remarkably, the introduction of L241D in the Os9BGlu31 W243N variant caused a significant activity increase for most substrates but the flavonoids (Table 1). Epistasis is the non-linear interaction between mutants such that their effects over function, in this case transglycosylation activity, are significantly higher or lower than those expected by simple addition of their individual effects. In Table 1, we observe several examples of positive epistasis^{42–44} (higher improvements in function than expected) against several substrates. Note that the double mutant L241D/W243N also exhibits the highest transglycosylation rate observed in this work using syringic acid as the acceptor. This is an example of sign epistasis^{42–44} (the point mutant L241D has an adverse effect on the activity of the single mutant, but it has a positive effect on the double mutant L241D/W243N).

Another second mutation was introduced to generate the mutant L183Q/W243N. Although the second mutation barely affected the transglycosylation rate of phenolic compounds over W243N, it increased the activity with other acceptors and positive epistasis was observed in several cases (Table 1). Very prominent sign epistasis was also observed in this double

mutant against the substrate 1-naphthol. Lastly, the introduction of the second mutation W243N in the mutant I172T could barely restore any of the activity lost upon the first mutation, suggesting that conservation of the amino acid I172 is important to catalyze the transglycosylation reaction. The entries in Table 1 with no annotation on epistasis indicate that these mutations may be purely additive or that the epistasis was not strong given the expected uncertainty in the data. Details on how the thresholds for the epistatic interactions were defined are provided in the ESI file SI01.xlsx.†

In addition to the mutants I172T, L183Q, and L241D, five W243 mutants (H, K, P, Q, and R) also showed lower transglycosylation and hydrolysis activities than the wild type. The rest of the mutants showed intermediate activities on the different substrates (Fig. 2). Notably, the function of Os9BGlu31 could be affected by both single and double mutants, but none of the mutants studied showed a higher activity on all the substrates. Based on the results obtained, we selected the following 8 Os9BGlu31 variants with high activity for further study, along with the wild type as the reference: W243C, W243I, W243L, W243N, W243T, W243V, L183Q/W243N, and L241D/W243N.

The preceding results were obtained with different concentrations of DMSO, which were adjusted based on the substrate solubility (0.25 vol% for phenolic compounds, 1.25 vol% for phytohormones and flavonoids, and 0 vol% for water). We suspected that the DMSO used as a co-solvent might have a slight effect on the catalytic activity of the enzymes. Therefore, we performed the activity study of the 9 high activity enzymes keeping the concentration of DMSO fixed at 2.5% by volume. These reactions were repeated in triplicate (file SI02.xlsx†). The results are presented in Fig. 3 and the activity trends are in agreement with the previous discussion.

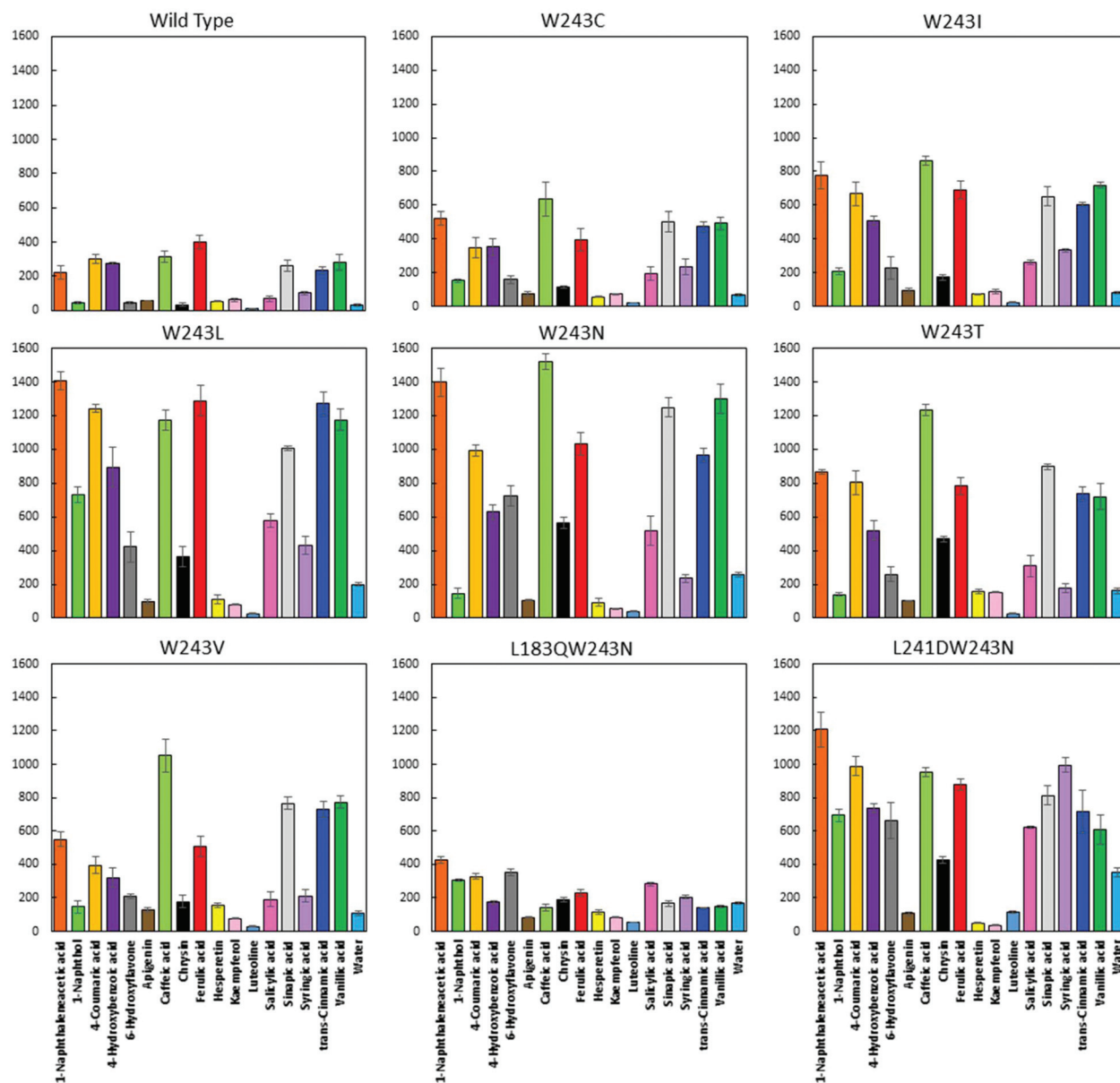


Fig. 3 Relative rates of 4NP release by the Os9BGluc1 wild-type enzyme, W243 mutants, and L183Q/W243N and L241D/W243N double mutants against 22 different acceptors, including water. For reaction conditions, see SI02.xlsx.†

It was found that the conversion of several hydrophilic residues did not improve substantially the hydrolysis activity of Os9BGluc1.¹⁸ Remarkably, in the present study we have identified several high activity mutants that can hydrolyze 4NPGlc faster than the wild-type (Fig. 4). In particular, Os9BGluc1 W243N shows the highest hydrolysis activity overall among the single mutant enzymes. This suggested that the mutation could be used in combination with additional hydrophilic mutations to further increase the hydrolysis rate of the enzyme. In fact, the double mutants engineered catalyze both reactions, hydrolysis and transglycosylation, with the L183Q/W243N variant showing the highest overall ratio of hydrolysis to transglycosylation and L241D/W243N showing the second highest. These findings may validate the idea that

hydrophobic residues near the putative water binding site contribute to transglycosylation activity, which was not apparent in previous studies.^{6,7,18} The selectivity to each of the two reactions can be determined by monitoring the products being synthesized (section 3.2).

3.2 Effect of the substrate on product distribution

To further understand the function of Os9BGluc1, we analyzed the different transglycosylation products during longer reaction times with a diverse group of substrates and enzyme mutants. Interestingly, several acceptors containing more than one hydroxyl group can accept glucose from the W243 mutants in different positions to produce multiple glucoconjugate compounds. These acceptors are the phenolic acids 4-hydroxy-

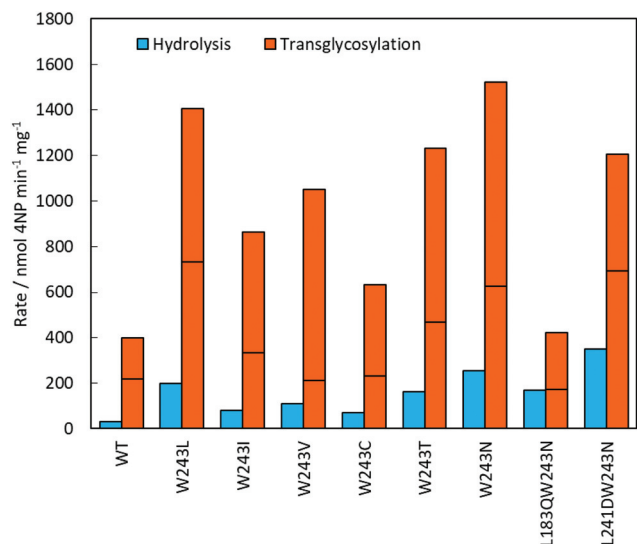


Fig. 4 Rate of release of 4NP with the best substrate and with water for wild type Os9BGlu31 and high-activity mutants. The line inside the bar indicates the median activity over 18 different substrates. For reaction conditions see SI02.xlsx.†

benzoic acid, ferulic acid, and vanillic acid and the flavonoids apigenin, luteolin, and kaempferol (Fig. 5).

Notably, the wild-type enzyme transferred the sugar exclusively to one hydroxyl group position of the phenolic acids (Fig. S2 to S4†) or the flavonoids, which is a good example of regioselectivity. Mutants W243 C, I, L, and N produced more than two products when provided with ferulic acid, including the bis-glucoside at 5.3 min, the glucoside at 6.4 min, and the glucose ester at 7.1 min (based on our previous characterization¹⁸), while only one more product (ferulic acid glucose ester, FAG) was formed by W243T and W243V (Fig. S2†). The W243 mutants produced different amounts of glucoconjugates, with the glucosyl moiety being transferred to different positions to a variable extent. For instance, while the wild type and mutants yielded significant amounts of the 4HB-glucose ester, these yields were even higher for the mutants W243N and W243L. Since both glucosides and glucose esters of these phenolic acids have been reported in nature,⁴⁵ the ability to synthesize them will be useful to study their biological roles.

We measured the product range formed from flavonoid substrates, some of which contain several hydroxyl groups (apigenin, luteolin, and kaempferol). Kaempferol and luteolin have similar structures with four hydroxyl groups (positions 3, 5, 7, and 4' in kaempferol and 5, 7, 3', and 4' in luteolin). W243N was shown to produce multiple kaempferol glucosides and bis-glucosides by transglycosylation.¹⁸ In this study, we also observed various glycosides with the newly generated mutants (Fig. S5†) and were able to extend this to other flavonoids with different hydroxyl group positions.

The chromatograms in Fig. S6 and S7† show the presence of multiple products in reactions catalyzed by Os9BGlu31 mutants. Four additional products were eluted in most of the reactions with apigenin catalyzed by

Os9BGlu31 mutants. The product eluting at 9.45 min (apigenin 7-O-glucoside)¹⁸ was synthesized by both wild-type and mutant enzymes but no other products were produced by the wild type, while additional peaks at 7.5, 9.55 and 11.3 min were seen for the mutants. The peak at 7.5 min which was particularly prominent in the reactions with the W243L and W243N variants is likely to be a bis-glucoside, based on the more hydrophilic elution position and previous observations with kaempferol glucosides.¹⁸ This suggests that Os9BGlu31 W243L and W254N are efficient at transglycosylating two positions on apigenin, resulting in high bis-glucoside production. Besides, the presence of three products eluting in the range of monoglucosides (9–12 min) suggests that the selected mutants can glycosylate all three hydroxyl groups on apigenin.

Luteolin could also accept the transfer of glucose by Os9BGlu31 transglucosidase variants. W243L appears more active than the other mutants, based on the abundance of new compounds, especially the one eluting at 7.4 min. As noted for the corresponding apigenin glucoconjugate, this compound is likely to be a bis-glucoside, based on its relatively early elution, which might result from the high activity of W243L towards two hydroxyl positions. W243L also showed a high yield of the product eluting at 9.4 min, in addition to the peak at 8.7 min, which was high in most variants. W243T showed a lower yield of the peak at 8.7 min, as it is more selective to the compound eluting at 9.4 min. The relatively high peak at 10.6 min for the W243T reaction and the relatively low hydrolysis rate shown in Fig. 3 and 4 suggest that it may produce another glucoside hidden in the *p*-nitrophenol peak, but we were unable to resolve this putative glucoside. These data suggest that most variants could be used to produce the glycoside at 8.7 min, while W243L and W243T could be used to produce the bis-glucoside and other luteolin glucosides.

3.3 Structural model of the ternary complex including the acceptor substrate

Previously, we produced a homology model of the wild type Os9BGlu31 structure,¹⁸ but the apo structure is not appropriate to evaluate the reactivity toward acceptor substrates, since they react with the glucosyl-enzyme intermediate (Fig. 1). To obtain a better understanding of the glucosyl transfer mechanism and visualize the binding orientations of the acceptor substrates within the Os9BGlu31 active site, we modeled the wild type and mutant proteins as their covalent glucosyl intermediates and docked the potential acceptor substrates on them. In these models, the glucoside is deeply buried in the active site and the access path is narrow, but small molecules can access it. This could be related to the lower activities of larger flavonoid substrates which are evident in Fig. 3, as the narrow binding site would cause difficulty in the access of very large substrates. Molecular docking revealed that the substrates are oriented on the opposite face of the glucose from the catalytic nucleophile, E387. The binding mode of the substrates is exemplified by ferulic acid and 1-naphthaleneacetic acid in Fig. 6. Although ferulic acid was positioned somewhat distant from the anomeric carbon atoms in the wild type active

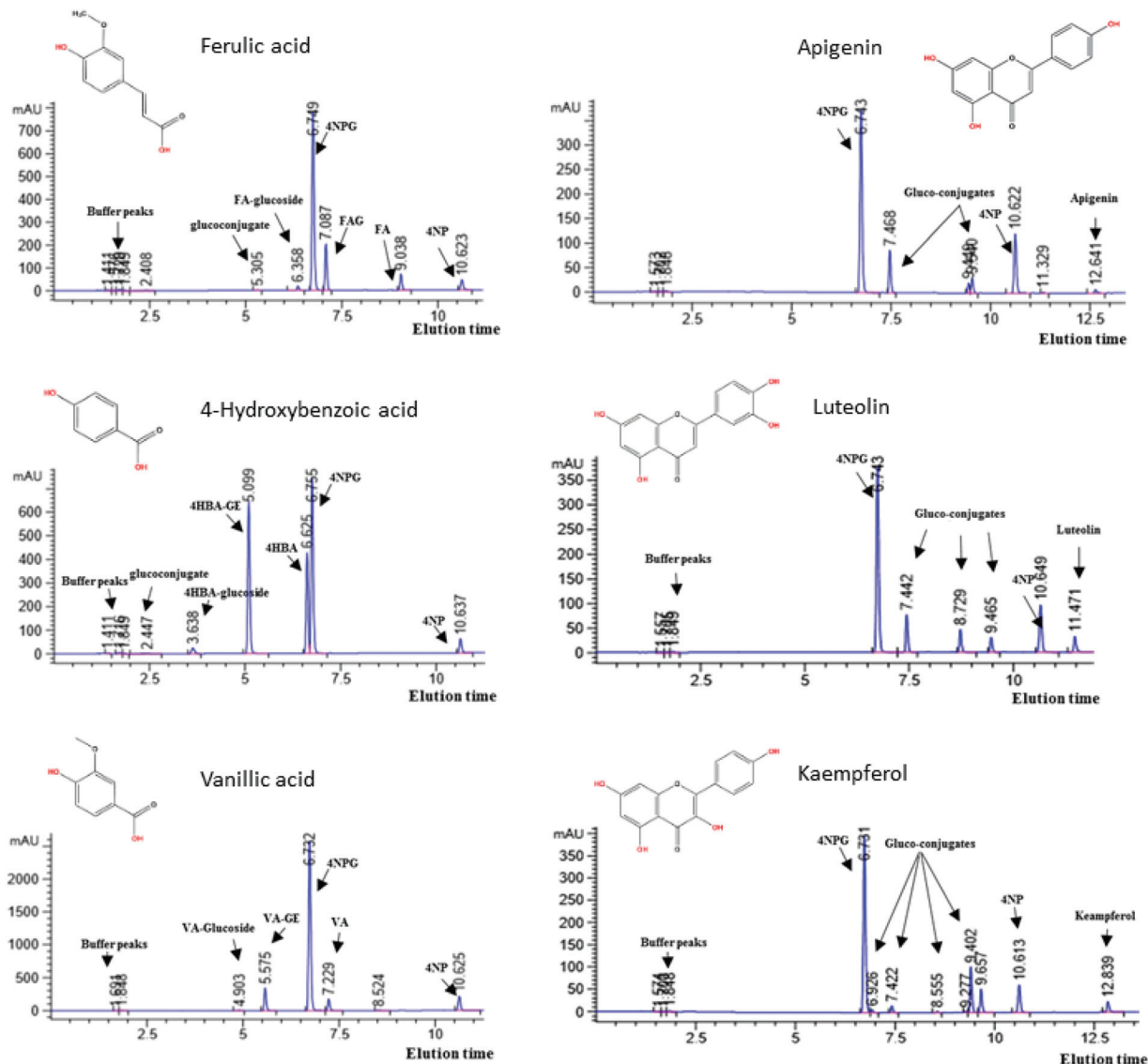


Fig. 5 UPLC chromatograms of products of reactions catalyzed by the Os9BGlu31 W243L variant using acceptors with multiple -OH groups and 4NPGlc as the glucose donor (absorbance at $\lambda = 360$ nm). The structures of the acceptors are shown at the top of the corresponding chromatogram. The reaction mixtures contained 5 mM 4NPGlc as the donor, 0.5 mM acceptor, and 5 μ g of Os9BGlu31 W243L in 50 mM citrate buffer (pH 4.5), and were incubated at 30 °C for 1 hour and then stopped with 1% formic acid. The components were separated over a C18 UPLC column with the solvent and gradient conditions described in the Methods. The large peaks of 4NPGlc at 6.7 min and 4NP at 10.6 min were used to assess the total enzymatic activity. Other peaks are marked according to their previous identification.¹⁸ FA is ferulic acid, FAG is ferulic acid 1-O-glucose ester, 4NP is 4-nitrophenol, 4NPG is 4NPGlc (4NP- β -D-glucoside), 4HBA is 4-hydroxybenzoic acid, 4HBA-GE is 4-HBA 1-O-glucose ester, VA is vanillic acid, and VA-GE indicates the VA glucose ester.

site, the models shown are in line with the predicted S_N2 mechanism proposed, as this requires the nucleophilic substrates to be positioned on the opposite face of the leaving group, E387, which ultimately results in the re-inversion of stereochemistry to form the β -anomeric products. Moreover, as shown in Table S2,[†] the binding scores for the acceptor substrates on the glycosylated enzyme intermediate are generally higher for the W243L mutant compared to the wild type, suggesting that a higher free energy of binding of the acceptor may contribute towards lowering the activation energy of the deglycosylation step and increasing the kinetic rate.

3.4 Cheminformatics

3.4.1 Linear, phenomenological model. In order to generalize the findings of this work to other possible substrates, we constructed different cheminformatics models. To build an appropriate model we require a set of descriptors that is relevant (*i.e.*, is highly related to the enzymatic activity that we want to predict) and at the same time minimizes the redundancy of information captured by the different indices under consideration (*i.e.*, the descriptors are not totally correlated with each other).

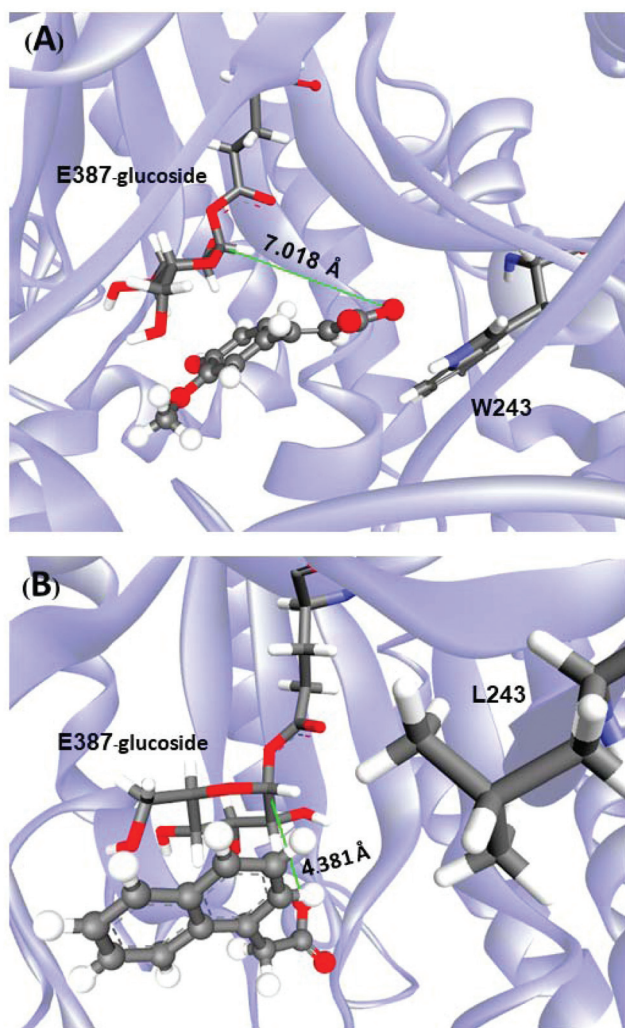


Fig. 6 Graphical representations of the binding modes of (A) ferulic acid to wild-type Os9BGlu31 and (B) 1-naphthaleneacetic acid to W243L mutant covalent glucosyl intermediates. The blue ribbons depict the secondary structure of the proteins. The acceptor substrates are shown in a ball-and-stick representation. E387-glucoside, W243, and L243 are shown as sticks only. The solid green lines indicate the distances measured between the nucleophilic carboxylate and the electrophilic C1 on the glucoside.

We considered the results with all 9 high-activity enzymes altogether on all substrates except water. Catalysis on water is different from the rest of substrates and its small size prevents the calculation of several molecular descriptors. For instance, *RotRatio* is undefined for molecules that do not have a rotatable bond; *HybRatio* cannot be computed for molecules with no carbon atoms; *AATSp*, *AATSCp*, *MATSp*, and *GATSp* are undefined when $p > \text{number of atoms}$, etc.

Firstly, we computed a large number of molecular features describing each of the substrates (2956 descriptors). We then applied a redundancy-reduction filter so that no pair of molecular descriptors had a correlation above 0.8. This step reduced the number of descriptors under consideration to 100, which suggests that some molecular descriptors that have

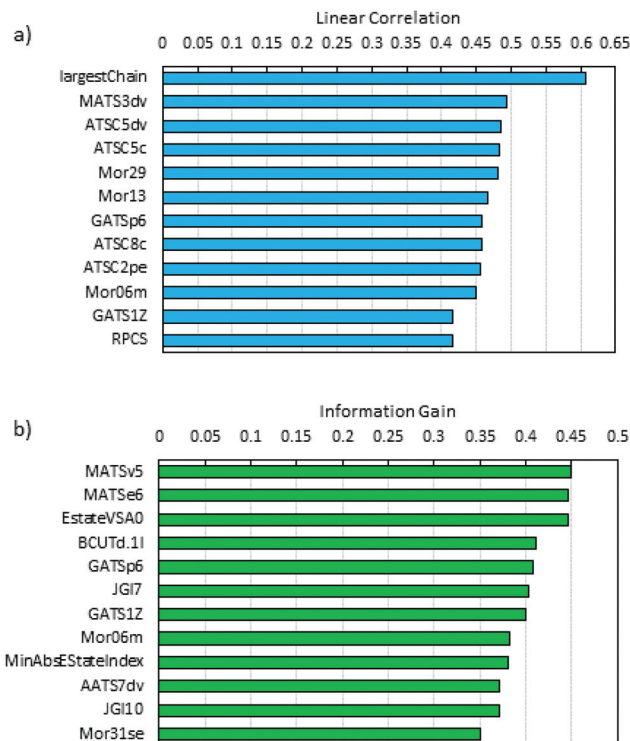


Fig. 7 Top feature selection results using as criteria the linear correlation (a) and the information gain (b) between the descriptors and the average enzymatic activity measured across all enzymes.

been proposed in the literature along the years may be more correlated with others than one might desire.⁴⁶ In a second step, we ranked the descriptors using as the criterion the Pearson correlation of each individual descriptor with the average enzymatic activity. The 12 most relevant descriptors according to this criterion are presented in Fig. 7a. Their names are indicated in Table S3.† We also considered another criterion for the ranking: in Fig. 7b, the descriptor importance is measured in terms of information gain, an entropy-based metric.

In Fig. 7, we can observe some interesting results. The descriptor most correlated with the enzymatic activity on its own is *largestChain*. This is a 1D constitutional descriptor from the RDKit package and, as its name suggests, it quantifies the number of atoms in the largest chain on the molecular graph. *MATS3dv* is the Moran coefficient of lag 3 weighted by valence electrons. This is a 2D centered autocorrelation descriptor computed by the Mordred software. This descriptor being selected as important suggests that relatively symmetrical acceptors, such as aromatic group-containing substrates,⁴⁷ may favor faster transglucosylation rates. *ATSC5dv* and *ATSC5c* are the centered Moreau–Broto autocorrelation descriptors of lag 5 weighted by valence electrons and by charges, respectively.

If we look at the maximum information gain as the criterion for selecting descriptors, we also find multiple autocorrelation descriptors: *MATSV5* and *MATSp6* are the Moran auto-

correlation coefficients of lag 5 weighted by van der Waals atomic volumes, and of lag 6 weighted by atomic polarizability, respectively. *GATSp6* and *GATS1Z* are the Geary autocorrelation coefficients of lag 6 weighted by atomic polarizability and of lag 1 weighted by atomic number, respectively. These two descriptors are considered relevant by both filtering algorithms. The fact that autocorrelation descriptors with different lags and different weighting are considered relevant by the algorithm suggests that the symmetry of the substrate, both in terms of connectivity and atomic makeup, would favor higher transglucosylation rates.

It is also interesting to see that, thanks to the pairwise-correlation reduction step, descriptors of diverse nature are considered in the ranking. For instance, *Mor29* and *Mor13* are the unweighted 3D-Morse descriptors of distances $d = 29$ and 13 , as computed by Mordred. The large value of the scattering parameter d should make these descriptors insensitive to atom pairs situated at large distances but highly sensitive to short distances ($<3 \text{ \AA}$).⁴⁸ This index may thus capture the effect of heteroatoms and multiple bonds (and thus bond distances) on the ability of the substrates to accept the glucosyl moiety. *Mor06m* is another 3D-Morse descriptor that is considered relevant. Notably, the topological charge indices *JGI7* and *JGI10* are also considered important by the information-based criterion. These indices have been shown to capture the charge distribution (e.g., the dipolar moment) within a molecule based solely on its topology, thus bypassing the complexity of 3D-optimizing the molecular structure.⁴⁹

A major limitation of most filter methods is the fact that they barely consider the correlation between features.⁵⁰ Thanks to our filter strategy, however, this limitation could be greatly reduced. In Table S4,† we can see that in a few cases the correlation between some of the indices shortlisted using the correlation with the enzymatic rate is close to the cutoff of 0.8. On average, however, the correlation between features is much lower, around 0.5, thanks to the design of the *findCorrelation* algorithm. Interestingly, the correlation among features when using the information gain as the criterion is even lower, but the correlation with the enzymatic rate is more diverse, in the range $R = 0.05\text{--}0.39$ vs. $R = 0.49\text{--}0.61$ when using the correlation criterion (Table S4†).

With the subset of 21 different molecular descriptors shortlisted in Fig. 7, it becomes possible to train a variety of models to seek a relationship among the molecular descriptors, the type of enzyme, and the enzymatic activity over a given acceptor substrate. One possibility is to build a linear model. The filtering algorithm has reduced the dimensionality of the problem by rejecting variables that are not very informative. Next, we conduct a *variable selection* to build a model (note that we still have 22 variables in our consideration set: 21 continuous molecular descriptors and 1 categorical variable or factor representing the enzyme type). One way to select variables is to use a stepwise selection method. To this end, we used the function *stepAIC* in the *MASS* R package. This algorithm adds or removes one variable at a time so that the new model leads to the highest reduction in the Akaike

Table 2 Coefficients for the independent term as a function of the enzyme in the linear cheminformatics model

Level	Intercept (nmol min ⁻¹ mg ⁻¹)
WT	8.69
L183Q/W243N	79.74
W243C	121.46
W243V	216.75
W243I	248.40
W243T	330.22
L241D/W243N	462.87
W243N	519.52
W243L	562.16

Information Criterion (AIC) relative to the previous model considered. The AIC is a statistic that allows evaluating the quality of a model: it is closely related to the correlation coefficient R , but it introduces a penalty for every degree of freedom consumed upon introduction of a new variable.³⁴ This way, we selected the following model:

$$\text{Rate (nmol min}^{-1} \text{ mg}^{-1}) =$$

$$\text{Intercept(Enzyme)} + 89.44\text{largest chain} + 2.31\text{ATSC5dv}$$

$$n = 153 \quad R^2 = 0.68 \quad R_{\text{adj}}^2 = 0.65 \quad F(10 \ 142) = 29.5$$

$$\text{AIC} = 1665.6 \quad p < 0.01$$

in which the independent term takes the values indicated in Table 2 depending on the query enzyme.

Not surprisingly, the algorithm has identified the type of enzyme and the *largestChain* descriptor as the most relevant variables, and it has also selected *ATSC5dv*. One advantage of linear models is their ease of interpretability. On the one hand, the molecular descriptors *largestChain* and *ATSC5dv* have a positive effect on the turnover of a substrate, with *largestChain* having a predominant effect. On the other hand, differences between values in Table 2 correspond to the expected differences in activity between enzymes when presenting a given substrate. It thus becomes clear that mutants W243L and W234N are much more active on average than the wild-type enzyme.

The linear model does not overfit the data, as indicated by the close values of R^2 and adjusted R^2 and the relatively low value of the AIC, and the model and every variable are considered highly significant by the corresponding F-tests. Remarkably, this simple linear model yields a high determination coefficient, $R^2 = 0.68$. One factor that may contribute to this is that, in our study, most of the enzyme variants considered are single mutants at the same position 243. It thus seems conceivable that these mutations have a localized steric effect within the active site cleft of the enzyme. This effect can thus be represented reasonably well by a constant factor depending on the specific residue introduced and a variable related to the substrate chain size, which is in agreement with the docking study in the previous section. On the other hand, this model has been built and trained considering all the data (except for the hydrolysis reactions). To increase the predictive

Table 3 Performance of 10 different artificial neural networks predicting transglycosylation activity with multiple substrates and enzymes (for details, see SI03.xlsx)

Network			Training		Validation	
Topology	Hidden activation	Output activation	RSE	R	RSE	R
MLP 12-3-1	Logistic	Identity	33 438	0.7264	33 494	0.7397
MLP 12-5-1	Logistic	Identity	24 856	0.8063	23 629	0.8241
MLP 12-7-1	Logistic	Identity	30 873	0.7554	27 619	0.7786
MLP 12-3-1	Exponential	Identity	36 041	0.7016	31 832	0.7381
MLP 12-5-1	Exponential	Identity	20 831	0.8403	27 899	0.7996
MLP 12-7-1	Exponential	Identity	18 715	0.8574	30 130	0.7978
MLP 12-3-1	Exponential	Identity	24 504	0.8170	23 676	0.8017
MLP 12-5-1	Exponential	Identity	17 788	0.8665	23 239	0.8120
MLP 12-7-1	Exponential	Identity	18 145	0.8639	25 942	0.7997
MLP 12-7-1	Identity	Identity	46 663	0.5822	32 346	0.7488

power of the model, in the next section we explore more flexible models and assess their performance against validation data.

3.4.2. Nonlinear, mechanistic models. In this section, we report a model based only on the molecular descriptors classically used in Hansch's method. The method has the advantage of using a few pre-selected descriptors to seek the models, making unnecessary the exploration of large spaces of chemical descriptors.⁵¹ Hansch's model is based on an extra-thermodynamic linear free energy relationship (LFER). The LFER model decomposes a biological mechanism into a series of several steps. Then, it seeks a linear combination or weighted summation (additive model) of the free energies of the different steps. It considers as free energy contributions parameters such as the negative logarithm of the ionization constant ($\text{p}K_i = -\log K_i$) and the logarithm of the partition coefficient ($\log P$). These parameters are related to the free energy of the ionization process and to the posterior membrane transport (partition) steps of the biological compound, respectively. Specifically, P is the partition coefficient of the biological compound in the system n -octanol/water and it measures the compound's lipophilicity. Hansch's model also considers other steric and electronic properties of the compounds, such as their polar surface area or molar refractivity, which increases the flexibility of the method to describe enzymatic processes such as the one investigated. The molecular descriptors used in this work are $D_1 = \text{ALOGP}$ (an estimate of $\log P$), $D_2 = \text{MR}$ (molar refractivity), and $D_3 = \text{TPSA}$ (topological polar surface area). Moreover, as shown in Table S4,† these different parameters show a limited correlation between them.

We evaluated multiple non-linear models using Artificial Neural Network (ANN) algorithms. The models use as input variables the type of enzyme and the 3 molecular descriptors calculated for the query compound. The networks built are Multi-Layer Perceptrons (MLP) with one hidden layer comprising 3, 6 or 9 neurons. Different combinations of identity, logistic, and exponential activation functions were considered. Table 3 summarizes the results obtained. On the one hand, we can see that the performance of the ANN is satisfactory (the coefficient of variation of experimental repeats is around

5–15%), with values of the correlation coefficient around $R = 0.75$ – 0.86 for the training and validation sets. In this case, 4/17 of the compounds were assigned randomly to the validation set across all the enzymes (see SI03.xlsx†). Thus, the activity results for 1-naphtol, 4-coumaric acid, kaempferol, and *trans*-cinnamic acid were not used in training the ANNs. Notably, the predictions for these compounds achieved a similar accuracy to those used to train the networks (Table 3, validation).

On the other hand, we observe that there is no great advantage in using a very flexible model over using a linear model with appropriate descriptors in our problem. In fact, the ANN that only uses identity functions shows a limited performance, which stems from the descriptors specified. Abruzzo *et al.* also reported that variable selection had a greater impact than the selection of the model in a comparative study of transcriptomics data classifiers.⁵² Furthermore, we observed that the training of larger dense hidden layers is more prone to encounter local optima and it may also increase the chances of undesired overfitting. Taken together, these results also suggest that the true relationship between inputs and output is not highly nonlinear.

Importantly, the models built can be used in practice to estimate the enzymatic activity of new acceptors over the different enzymes and thus used to screen libraries of potential substrates computationally. Given the reasonable performance of the different networks in Table 3, we averaged the individual predictions by all of them to reach an ensemble or consensus prediction. In Table S5,† we report the results of exploring *in silico* the transglycosylation rates of a new set of putative acceptors by the wild-type Os9BGlu31 and different mutants. Although several of these compounds are expected to exhibit low activity, the results suggest that 3,4-dihydroxybenzoic acid, 6-benzylaminopurine, indole-3-acetic acid, and serotonin may be particularly reactive transglycosylation substrates over the enzyme variants W243L, W243N, and L241D/W243N. Aided by these predictions, we are already advancing the production of novel antibiotic and phytohormone glucoconjugates of interest, such as abscisic acid glucoside (preliminary results shown in Fig. S8†), on which we recently submitted a patent application.

In these results, glucose was transferred from 4NPGlc to a variety of compounds and acceptor positions, allowing us to prepare a broad array of valuable glucoconjugates. Other glucose donors have also been proposed.⁵³ For example, β -glucosyl fluoride has recently been shown to be an effective donor for certain Leloir-type glycosyltransferases in the presence of catalytic amounts of UDP.⁵⁴ With regard to the large-scale production of glucoconjugates, more atom-efficient donors could be pursued, which opens up another area of potential research. The cheminformatics pipeline proposed in this work could also be extended to new combinations of substrates and catalysts, thus enabling increasing possibilities for transglycosidases and other enzymes.

4. Conclusion

Glucoconjugate molecules are of great interest to the fine chemistry and phytochemistry industries, and many of these compounds could be synthesized from feedstocks derived from biorefineries. However, the production of complex glucoconjugates using conventional synthetic approaches may not be sustainable or economical, as multiple synthetic steps with limited yield would be involved. In this work, we studied 16 new Os9BGlu31 variants and compared them to the wild-type Os9BGlu31 to identify new high-activity catalysts, such as W243L, and to correlate their activities with properties of the substrates being converted. Furthermore, the combination of the mutations designed to increase the hydrophilicity of the active site, L183Q and L241D, with the high-activity W243N mutation led to catalysts with high hydrolytic activity compared to their transglycosylation activity. These mutations exhibited positive epistasis on other substrates as well, which may be related to their proximity in the protein fold. We went a step further and used computational methods to facilitate a greener catalyst selection depending on the target substrate with a pipeline that can be used or adapted by other researchers. Cheminformatics models can help minimize the resources consumed to investigate new substrates and maximize the throughput of new catalytic processes. Our models suggest that steric constraints in the active site have a crucial effect on the activity of a substrate, which agrees with the docking simulations. Careful variable selection can afford linear models with reasonable performance, although higher accuracy could be attained with nonlinear artificial neural networks. These models are proving useful to identify substrates and efficiently produce new glucoconjugates of interest, such as phytohormone glucosides for enzymatic and plant physiology studies or antibiotic glycosides with improved solubility and pharmacological properties.

Conflicts of interest

The authors declare no competing financial interests.

Acknowledgements

This work was supported by grants from Suranaree University of Technology (SUT), the Thailand Research Fund (BRG5980015), and the National Research University Project from the Commission on Higher Education (Thailand) to SUT. Ministerio de Economía y Competitividad (FEDER CTQ2016-74881-P and CTQ2013-41229-P) and Basque Government (IT1045-16) are acknowledged for financial support. V. B. thanks the support from the Center for Biomolecular Structure, Function and Application (SUT). V. B. thanks Dr. Jie Dong (Central South University, China) for useful comments and Prof. Jisnusun Svasti (Chulabhorn Research Institute, Thailand) for his very kind support.

References


- 1 J. R. Ketudat Cairns, B. Mahong, S. Baiya and J. S. Jeon, *Plant Sci.*, 2015, **241**, 246–259.
- 2 L. L. Lairson and S. J. Withers, *Chem. Commun.*, 2004, **20**, 2243–2248.
- 3 D. E. Koshland, *Biol. Rev.*, 1953, **28**, 416–436.
- 4 S. G. Withers, R. A. J. Warren, I. P. Street, K. Rupitz, J. B. Kempton and R. Aebersold, *J. Am. Chem. Soc.*, 1990, **112**, 5887–5889.
- 5 M. J. Baumann, J. M. Eklöf, G. Michel, A. M. Kallas, T. T. Teeri, M. Czjzek and H. Brumer, *Plant Cell*, 2007, **19**, 1947–1963.
- 6 D. Teze, J. Hendrickx, M. Czjzek, D. Ropartz, Y. H. Sanejouand, V. Tran, C. Tellier and M. Dion, *Protein Eng., Des. Sel.*, 2014, **27**, 13–19.
- 7 B. Bissaro, J. Durand, X. Biarnés, A. Planas, P. Monsan, M. J. O'Donohue and R. Faure, *ACS Catal.*, 2015, **5**, 4598–4611.
- 8 E. Champion, I. André, C. Moulis, J. Boutet, K. Descroix, S. Morel, P. Monsan, L. A. Mulard and M. Remaud-Siméon, *J. Am. Chem. Soc.*, 2009, **131**, 7379–7389.
- 9 E. Champion, F. Guérin, C. Moulis, S. Barbe, T. H. Tran, S. Morel, K. Descroix, P. Monsan, L. Mourey, L. A. Mulard, S. Tranier, M. Remaud-Siméon and I. André, *J. Am. Chem. Soc.*, 2012, **134**, 18677–18688.
- 10 L. F. Mackenzie, Q. P. Wang, R. A. J. Warren and S. G. Withers, *J. Am. Chem. Soc.*, 1998, **120**, 5583–5584.
- 11 M. Jahn, J. Marles, R. A. J. Warren and S. G. Withers, *Angew. Chem., Int. Ed.*, 2003, **42**, 532–534.
- 12 E. R. Moellering, B. Muthan and C. Benning, *Science*, 2010, **330**, 226–228.
- 13 Y. Matsuba, N. Sasaki, M. Tera, M. Okamura, Y. Abe, E. Okamoto, H. Funabashi, M. Takatsu, M. Saito, H. Matsuoka, K. Nagasawa and Y. Ozeki, *Plant Cell*, 2010, **22**, 3374–3389.
- 14 S. Luang, J. L. Cho, B. Mahong, R. Opasiri, T. Akiyama, K. Phasai, J. Komvongsa, N. Sasaki, Y. Hua, Y. Matsuba, Y. Ozeki, J. S. Jeon and J. R. Ketudat Cairns, *J. Biol. Chem.*, 2013, **288**, 10111–10123.
- 15 T. Miyahara, R. Sakiyama, Y. Ozeki and N. Sasaki, *J. Plant Physiol.*, 2013, **170**, 619–624.

- 16 Y. Nishizaki, M. Yasunaga, E. Okamoto, M. Okamoto, Y. Hirose, M. Yamaguchi, Y. Ozeki and N. Sasaki, *Plant Cell*, 2013, **25**, 4150–4165.
- 17 J. Komvongsa, B. Mahong, K. Phasai, Y. Hua, J. S. Jeon and J. R. Ketudat Cairns, *J. Agric. Food Chem.*, 2015, **63**, 9764–9769.
- 18 J. Komvongsa, S. Luang, J. V. Marques, K. Phasai, L. B. Davin, N. G. Lewis and J. R. Ketudat Cairns, *Biochim. Biophys. Acta*, 2015, **1850**, 1405–1414.
- 19 R. W. Gantt, P. Peltier-Pain and J. S. Thorson, *Nat. Prod. Rep.*, 2011, **28**, 1811–1853.
- 20 C. Zhang, B. R. Griffith, Q. Fu, C. Albermann, X. Fu, I.-K. Lee and L. Li, *Science*, 2006, **313**, 1291–1294.
- 21 V. Křen and T. Řezanka, *FEMS Microbiol. Rev.*, 2008, **32**, 858–889.
- 22 I. Finore, A. Poli, P. Di Donato, L. Lama, A. Trincone, M. Fagnano, M. Mori, B. Nicolaus and A. Tramice, *Green Chem.*, 2016, **18**, 2460–2472.
- 23 R. A. Sheldon, *ACS Sustainable Chem. Eng.*, 2018, **6**, 4464–4480.
- 24 A. Pennec, L. Legentil, L. Herrera-Estrella, V. Ferrières, A.-L. Chauvin and C. Nugier-Chauvin, *Green Chem.*, 2014, **16**, 3803–3809.
- 25 V. Blay, R. Garcia-Domenech and J. Galvez, *ChemTexts*, 2017, **3**, 2.
- 26 V. Blay, J. Gullon-Soleto, M. Galvez-Llompart, J. Galvez and R. Garcia-Domenech, *ACS Sustainable Chem. Eng.*, 2016, **4**, 4224–4231.
- 27 B. Bhattacharai, R. Garg and P. Gramatica, *Mol. Inf.*, 2010, **29**, 511–522.
- 28 C. B. Santiago, J.-Y. Guo and M. S. Sigman, *Chem. Sci.*, 2018, **9**, 2398–2412.
- 29 R. Opassiri, B. Pomthong, T. Onkoksoong, T. Akiyama, A. Esen and J. R. Ketudat Cairns, *BMC Plant Biol.*, 2006, **6**, 33.
- 30 J. Dong, D. S. Cao, H. Y. Miao, S. Liu, B. C. Deng, Y. H. Yun, N. N. Wang, A. P. Lu, W. B. Zeng and A. F. Chen, *J. Cheminf.*, 2015, **7**, 60.
- 31 H. Moriwaki, Y. S. Tian, N. Kawashita and T. Takagi, *J. Cheminf.*, 2018, **10**, 4.
- 32 M. Kuhn, *J. Stat. Softw.*, 2008, **28**, 1–26.
- 33 T. Cheng, Y. Wang and S. H. Bryant, *Bioinformatics*, 2012, **28**, 2851–2852.
- 34 T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2nd edn, 2009.
- 35 A. Sali and T. L. Blundell, *J. Mol. Biol.*, 1993, **234**, 779–815.
- 36 S. Seshadri, T. Akiyama, R. Opassiri, B. Kuaprasert and J. K. Cairns, *Plant Physiol.*, 2009, **151**, 47–58.
- 37 M. Y. Shen and A. Sali, *Protein Sci.*, 2006, **15**, 2507–2524.
- 38 J. W. Schymkowitz, F. Rousseau, I. C. Martins, J. Ferkinghoff-Borg, F. Stricher and L. Serrano, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 10147–10152.
- 39 M. J. Sippl, *Proteins*, 1993, **17**, 355–362.
- 40 R. A. Laskowski, M. W. Macarthur, D. S. Moss and J. M. Thornton, *J. Appl. Crystallogr.*, 1993, **26**, 283–291.
- 41 G. Jones, P. Willett, R. C. Glen, A. R. Leach and R. Taylor, *J. Mol. Biol.*, 1997, **267**, 727–748.
- 42 J. F. Storz, *Curr. Opin. Struct. Biol.*, 2018, **50**, 18–25.
- 43 T. N. Starr and J. W. Thornton, *Protein Sci.*, 2016, **25**, 1204–1218.
- 44 M. Camps, A. Herman, E. Loh and L. A. Loeb, *Crit. Rev. Biochem. Mol. Biol.*, 2007, **42**, 313–326.
- 45 S. Klick and K. Herrmann, *Phytochemistry*, 1988, **27**, 2177–2180.
- 46 R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors*, Wiley, 2008, p. 692, ISBN: 9783527613106.
- 47 B. Hollas, *J. Math. Chem.*, 2003, **33**, 91–101.
- 48 O. Devinyak, D. Havrylyuk and R. Lesyk, *J. Mol. Graphics Modell.*, 2014, **54**, 194–203.
- 49 J. Galvez, R. Garcia, M. T. Salabert and R. Soler, *J. Chem. Inf. Comput. Sci.*, 1994, **34**, 520–525.
- 50 G. Chandrashekar and F. A. Sahin, *Comput. Electr. Eng.*, 2014, **40**, 16–28.
- 51 P. Abeijon, X. Garcia-Mera, O. Caamano, M. Yanez, E. Lopez-Castro, F. J. Romero-Duran and H. Gonzalez-Diaz, *Curr. Drug Targets*, 2017, **18**, 511–521.
- 52 L. V. Abruzzo, L. L. Barron, K. Anderson, R. J. Newman, W. G. Wierda, S. O'Brien, A. Ferrajoli, M. Luthra, S. Talwalkar, R. Luthra, D. Jones, M. J. Keating and K. R. Coombes, *J. Mol. Diagn.*, 2007, **9**, 546–555.
- 53 S. Shoda, *Proc. Jpn. Acad., Ser. B*, 2017, **93**, 125–145.
- 54 A. Lepak, A. Gutmann and B. Nidetzky, *ACS Catal.*, 2018, **8**, 9148–9153.

PAPER

[View Article Online](#)
[View Journal](#) | [View Issue](#)Cite this: *RSC Adv.*, 2019, 9, 6211

Gram scale production of 1-azido- β -D-glucose via enzyme catalysis for the synthesis of 1,2,3-triazole-glucosides†

Jaggaiah N. Gorantla,^a Salila Pengthaisong,^a Sunaree Choknud,^a
Teadkai Kaewpuang,^b Tanaporn Manyum,^a Vinich Promarak^b and James R. Ketudat
Cairns *^a

The production of analytical amounts of azido sugars is used as a means of verifying catalytic acid/base mutations of retaining glycosidase, but application of this process to preparative synthesis has not been reported. The catalytic acid/base mutant of *Thermoanaerobacterium xylanolyticus* GH116 β -glucosidase, TxGH116D593A, catalyzed the gram scale production of 1-azido- β -D-glucose (**1**) from *p*-nitrophenyl- β -D-glucopyranoside (pNPGlc) and azide via a transglucosylation reaction. Overnight reaction of the enzyme with pNPGlc and NaN₃ in aqueous MES buffer (pH 5.5) at 55 °C produced **1** (3.27 g), which was isolated as a white foamy solid in 96% yield. This **1** was successfully utilized for the synthesis of fifteen 1,2,3-triazole- β -D-glucosyl derivatives (**2–16**) containing a variety of functional groups, via click chemistry.

Received 28th January 2019
Accepted 14th February 2019

DOI: 10.1039/c9ra00736a

rsc.li/rsc-advances

Introduction

Enzymes are well known for their unique ability to catalyze stereospecific reactions and produce stereoselective products.^{1–3} Retaining β -glycosidases are one group of enzymes displaying this property via generation of products with the same β -anomeric configuration as their substrates. The retaining β -glycosidase mechanism involves the displacement of the aglycone leaving group from the anomeric carbon of the sugar by the catalytic nucleophile with acid-assistance by the catalytic acid/base, which protonates the leaving group.^{4,5} The α -glycosyl enzyme intermediate, thus formed, is resolved by attack of a water molecule or other nucleophile on the anomeric carbon to displace the catalytic nucleophile residue and release a β -sugar or β -glycoside, respectively. This second step is generally facilitated by the deprotonation of the incoming nucleophile (water) molecule by the catalytic acid/base residue.

Acid/base mutants of β -glycosidase enzymes, in which the catalytic carboxylate is replaced with a nonionizable amino acid sidechain, have been frequently reported to catalyze the formation of β -glucosides without significant hydrolysis.^{4,5} In an

early study with the acid/base mutant of *Agrobacterium* β -glucosidase, the substrates 2,4-dinitrophenyl- β -D-glucopyranoside (DNPGlc) and 4-nitrophenyl- β -D-glucopyranoside (pNPGlc) served as donor substrates for the transglucosylation of various nucleophiles, including formate, azide, acetate, cyanide and benzoate.⁵ Similarly, the *Thermoanaerobacterium xylanolyticum* TxGH116 β -glucosidase acid/base mutant, TxGH116D593A, was previously produced and rescued by azide to verify that the D593 acts as a catalytic acid/base.⁶ Kinetic characterization of transglucosylation reactions catalyzed by β -glucosidase catalytic acid/base mutants to produce 1-azido- β -D-glucose (**1**) has also been reported,^{5–7} but exploration of this reaction as a preparative method has not been reported.

Synthesis of 1-azido- β -D-glucose (**1**) has also been reported from unprotected glucose using 1,3-dimethylimidazolinium salts, base and NaN₃ in deuterium oxide,^{8–10} and other reagents.^{11,12} 1-Azido- β -D-glucose (**1**) has been used in the copper-catalyzed azide-alkyne cycloaddition (CuAAC) reaction for the construction of triazole-glucosides.^{13–16} Triazole-glucosides exhibit a wide range of glucosidase,^{17–19} glycogen phosphorylase,^{20–22} and carbonic anhydrase^{23,24} inhibitory activities (Fig. 1). The triazole-glucosides also exhibit binding affinity towards various proteins,^{25–28} and were found to be potential radiopharmaceuticals,^{29,30} pesticides,³¹ water soluble prodrugs, and inhibitors against cancer associated carbonic anhydrases,^{32,33} and to have antinociception³⁴ activities (Fig. 1). Triazole-glucosides and their various activities have been reviewed.^{35,36}

In the present study, a catalytic process for the gram scale production of the 1-azido- β -D-glucose (**1**) by an enzyme (TxGH116D593A) in water was developed. The purified enzyme

^aSchool of Chemistry, Institute of Science, & Center for Biomolecular Structure, Function and Application, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand. E-mail: cairns@sut.ac.th; Fax: +66 44 224185; Tel: +66 44 224193

^bDepartment of Materials Science and Engineering, School of Molecular Science and Engineering, Vidyasirimedhi Institute of Science and Technology, Rayong 21210, Thailand

† Electronic supplementary information (ESI) available: Supplementary material contains Fig. S1–S6, including purification, ¹H NMR spectra of **1–16**, ¹³C NMR spectra of **1–16**, and HRMS spectrum of compound **16**. See DOI: 10.1039/c9ra00736a

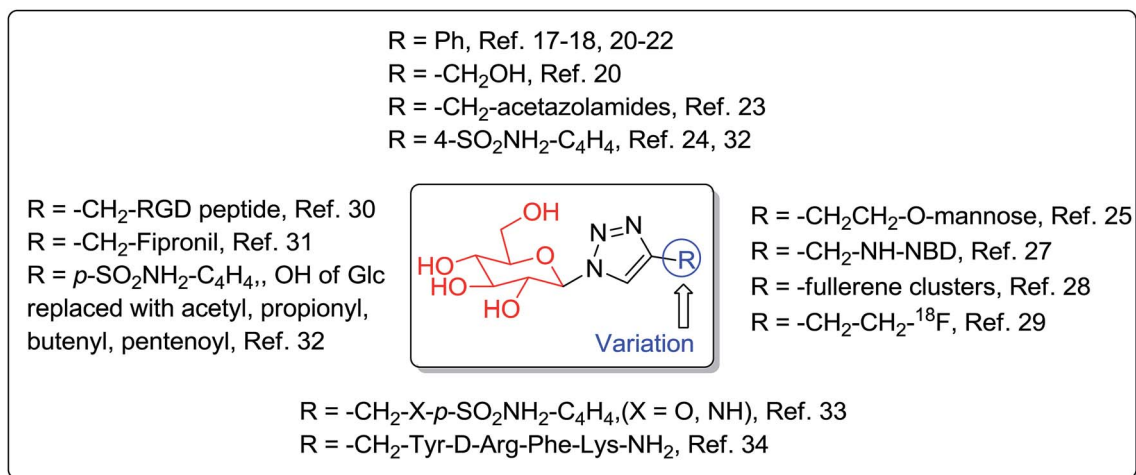


Fig. 1 Reported bioactive 1,2,3-triazole- β -D-glucosides.

catalyzed the production of gram scale amounts of **1** from *p*NPGlc and sodium azide in aqueous MES buffer (pH = 5.5), in an overnight reaction at 55 °C. After the overnight incubation, thin layer chromatography (TLC) indicated the nearly complete conversion of *p*NPGlc substrate into **1**, by the enzyme *via* transglucosylation with azide as the acceptor nucleophile. The product (**1**) could be conveniently purified and reacted with various alkynes to form a variety of 1,2,3-triazole- β -D-glucosyl derivatives (**2–16**), without the need for further deprotection of the sugar.

Results and discussion

Initially, we optimized the catalytic transglucosylation reaction conditions by varying the concentration of enzyme, substrate,

and sodium azide, while the MES buffer was maintained at pH 5.5, based on previous pH optimization. Following these initial experiments, the production was screened with 1 : 3, 1 : 6, 1 : 9, 1 : 12 and 1 : 15 (w/w) ratios of enzyme : *p*NPGlc substrate, 100 mM sodium azide and 50 mM MES buffer, pH 5.5, in 1 mL water, and yellow color was observed after the 24 h incubation at 55 °C. TLC (Fig. 2) showed the formation of a new polar compound, with the *R_f* of 1-azido- β -D-glucose and near complete conversion of substrate, except at the 1 : 15 ratio, where the substrate conversion was clearly not completed.

Again, we ran the same experiment as explained above with 1 : 13 and 1 : 14 (w/w) ratios of enzyme : substrate, the substrate was almost completely converted into a product, as indicated by TLC (Fig. 2). Taking advantage of this optimization of reaction conditions, we used the 1 : 14 (w/w) ratio of enzyme : substrate

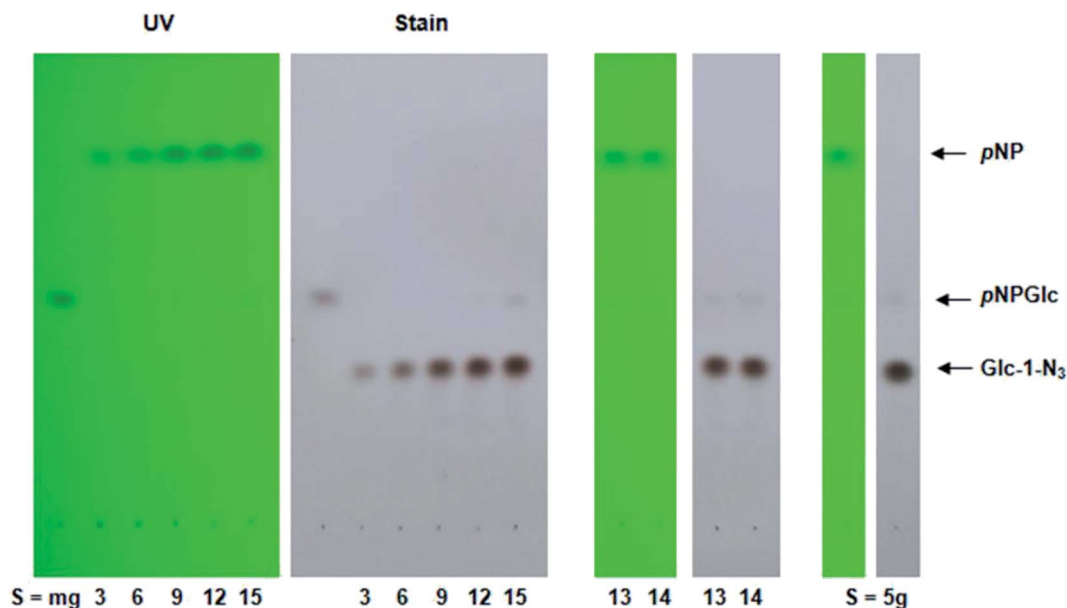


Fig. 2 TLC profile of 1-azido- β -D-glucose (**1**) formation after 24 h from 3 mg/1 mg to 15 mg/1 mg and 5 g/370 mg of *p*NPGlc/enzyme (7xGH116D593A). In each case, detection under UV light is shown on the left and carbohydrate staining with 10% sulphuric acid in ethanol and charring is shown on the right. The numbers below the TLC lanes indicate ratios of *p*NPGlc to enzyme.



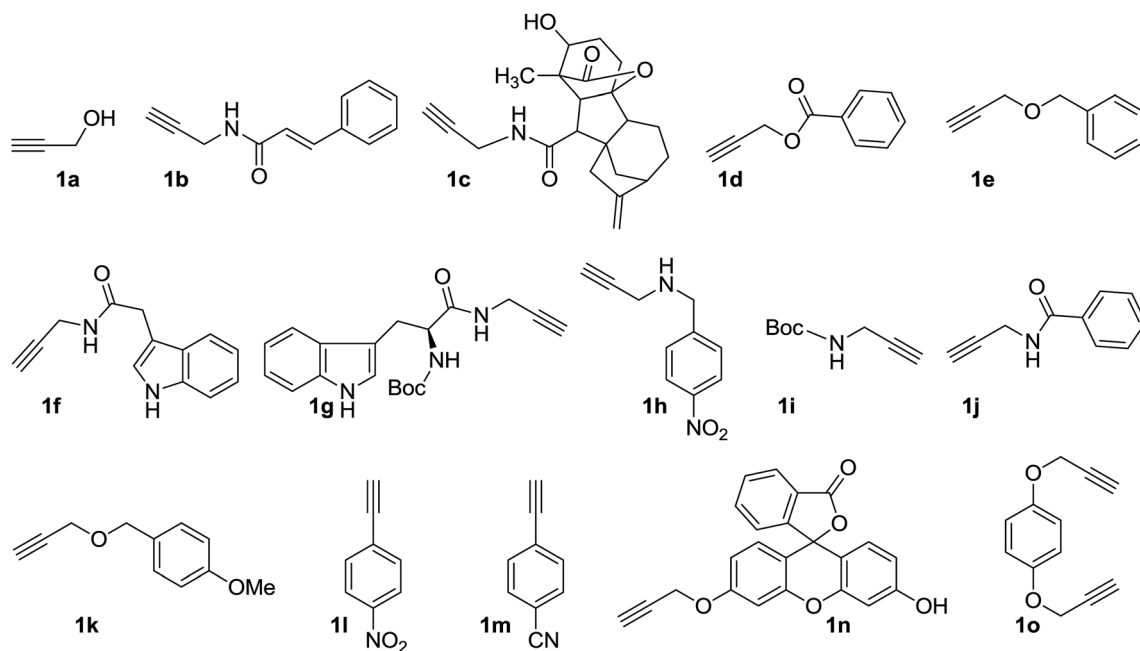


Fig. 3 Structures of alkynes (1a–o) used in click chemistry.

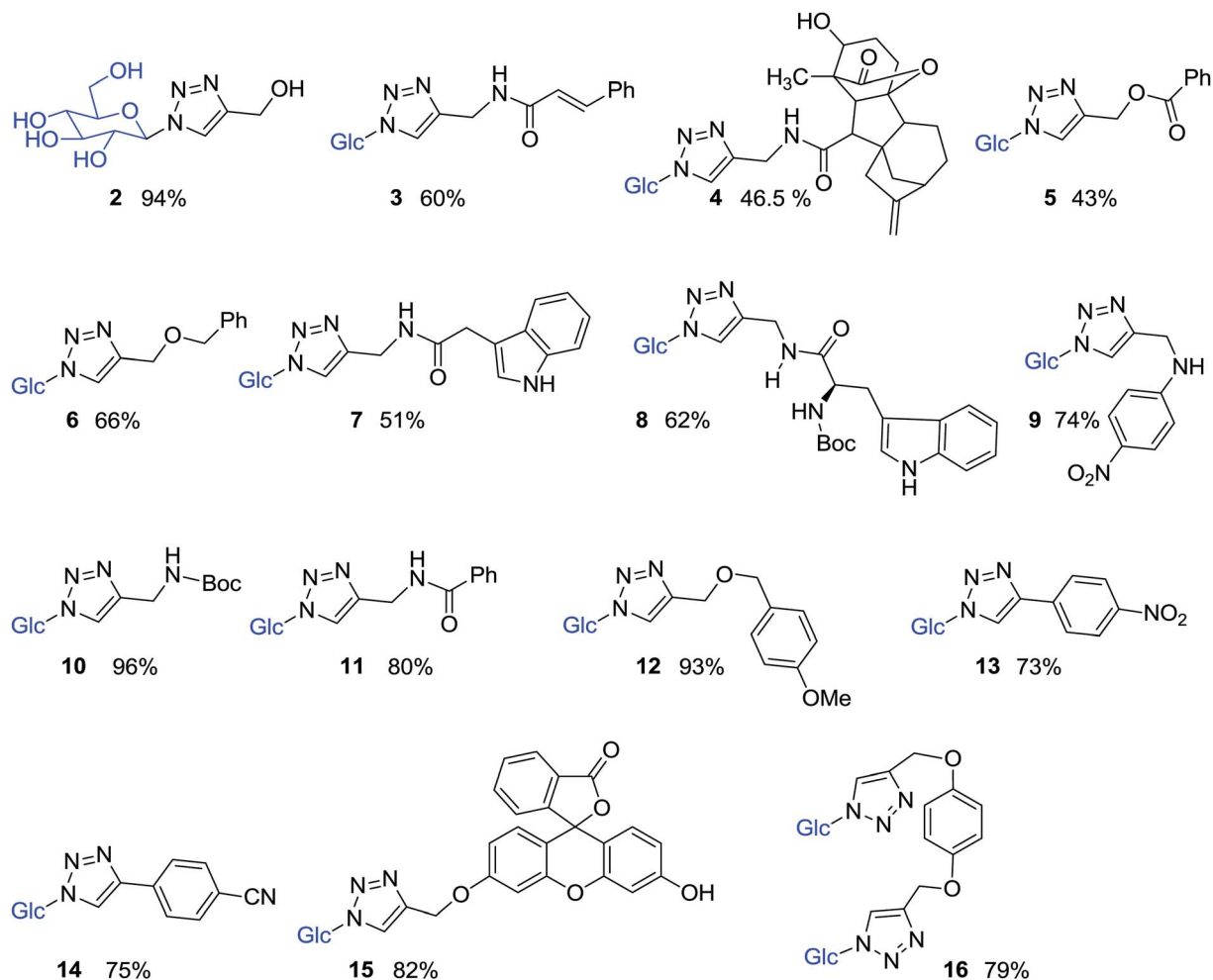


Fig. 4 Structures of 1,2,3-triazole-β-D-glucosyl derivatives 2–16.



for the large scale synthesis of 1-azido- β -D-glucose (**1**), from *p*NPGlc and azide catalyzed by TxGH116D593A in a 370 mL reaction volume (see ESI, Fig. S1†). After 24 hours of reaction at 55 °C, the reaction mixture was lyophilized to dryness, redissolved in methanol (MeOH) and purified by silica gel column chromatography. The *p*NP product eluted in 100% ethyl acetate (EtOAc, 300 mL), and then the unreacted *p*NPGlc was isolated as a mixture with 1-azido- β -D-glucose (**1**) (2.645 g) (see ESI Fig. S1†), with EtOAc/MeOH (95 : 5, v/v). The product **1** (800 mg) (see ESI Fig. S1†), was isolated in pure form by elution with EtOAc/MeOH (94 : 6, v/v). Finally, a trace of glucose byproduct (see ESI Fig. S1†) was eluted with EtOAc/MeOH (1 : 1, v/v). The pure 1-azido- β -D-glucose (**1**) was dried under reduced pressure to give a white foamy solid, R_f = 0.36 (CHCl₃/MeOH/NH₃, 7 : 2.8 : 0.2, v/v/v). Compound **1** was characterized with ¹H, ¹³C NMR and HRMS analysis and the ¹H NMR spectrum resembles that reported in the literature (see ESI†).⁶ In compound **1**, the anomeric proton (H-1) shifted to downfield and merged with water solvent residual peak at δ 4.8 ppm in 600 MHz (see ESI page S4†), but the H-1 proton was observed at δ 4.55 ppm (H-1, d, J = 9 Hz, 1H) in the 400 MHz spectrum (see ESI page S3†).

To increase the yield, the obtained mixture (2.645 g) was redissolved in 6 mL of water/methanol (95 : 5, v/v) and separated over a Sephadex LH-20 column in the same solvent (ESI Fig. S2†). The product (**1**) eluted first in pure form, followed by a minor amount of *p*NPGlc. Drying this portion of product yielded an additional 2.470 g of compound **1** as a foamy white solid. Added with the 0.800 g from the first step, the 3.270 g of

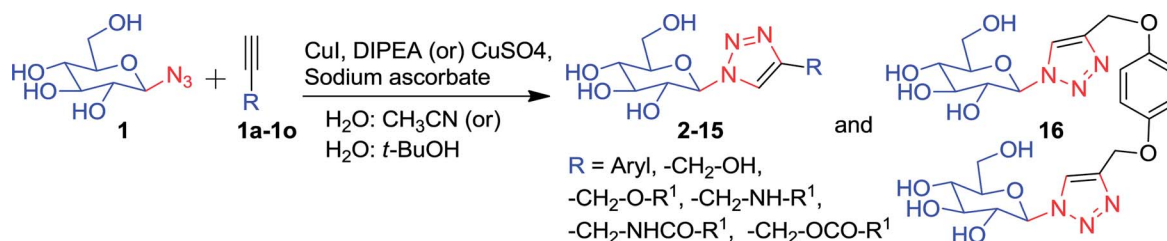
compound **1** represented 96% of the theoretical yield from the 5 g of *p*NPGlc used in the reaction. This was similar to the yield of 230 mg obtained from 330 mg of *p*NPGlc (98% yield) that we obtained in a preliminary small-scale reaction, suggesting reproducibly efficient conversion at milligram to gram scale. The purified compound (**1**) was thus ready for the synthesis of β -D-glucosyl-1,2,3-triazole derivatives *via* a click chemistry.

Compound **1** was subsequently joined with different compounds in which a propargyl group was attached to an alcohol, ethers, esters, amines, and amides (Fig. 3), as well as phenyl acetylenes in a copper-catalyzed azide-alkyne cycloaddition (CuAAC) reactions, giving rise to the 1,2,3-triazole- β -D-glucosyl derivatives **2–16** (Fig. 4, Table 1).

Initially, except for **1a**, all the alkynes **1b–1k** and **1n–o** (Fig. 3) were freshly prepared from their respective starting materials by EDC coupling, reductive amination and alkylation reactions. The phenyl acetylenes (**1l–m**), were obtained from their respective aldehydes by the Seyferth–Gilbert homologation reaction.³⁷

Pure 1-azido- β -D-glucose (**1**) was successfully coupled with propargyl groups attached to hydroxyl, ether, amine, amide and ester functional groups by copper-catalyzed azide-alkyne cycloaddition (CuAAC) reactions at room temperature. In all the reactions, two equivalents each of CuI and DIPEA were used for the coupling of **1** and alkynes **1a–k** (Fig. 3) giving rise to most of the water soluble 1,2,3-triazole- β -D-glucosyl derivatives **2**,²⁰ **3**,³⁸ **4**, **5**, **6**,³⁹ **7**, **8**, **9**, **10**,³⁸ **11**,³⁸ **12**, in good to excellent yields (Fig. 4, Table 1). When the phenyl acetylenes were reacted with

Table 1 General reaction of the substrates used in click reactions



Entry	Alkyne	Cu salt (eq.)	Base (eq.)	Solvent system	Triazole-glucoside	Isolated yield
1	1a	CuI (2)	DIPEA (2)	H ₂ O : CH ₃ CN (1 : 2)	2	94%
2	1b	CuI (2)	DIPEA (2)	H ₂ O : CH ₃ CN (1 : 2)	3	60%
3	1c	CuI (2)	DIPEA (2)	H ₂ O : CH ₃ CN (1 : 2)	4	46.5%
4	1d	CuI (2)	DIPEA (2)	H ₂ O : CH ₃ CN (1 : 2)	5	43%
5	1e	CuI (2)	DIPEA (2)	H ₂ O : CH ₃ CN (1 : 2)	6	66%
6	1f	CuI (2)	DIPEA (2)	H ₂ O : CH ₃ CN (1 : 2)	7	51%
7	1g	CuI (2)	DIPEA (2)	H ₂ O : CH ₃ CN (1 : 2)	8	62%
8	1h	CuI (2)	DIPEA (2)	H ₂ O : CH ₃ CN (1 : 2)	9	78%
9	1i	CuI (2)	DIPEA (2)	H ₂ O : CH ₃ CN (1 : 2)	10	96%
10	1j	CuI (2)	DIPEA (2)	H ₂ O : CH ₃ CN (1 : 2)	11	80%
11	1k	CuI (2)	DIPEA (2)	H ₂ O : CH ₃ CN (1 : 2)	12	93%
12	1l	CuI (2)	DIPEA (2)	H ₂ O : CH ₃ CN (1 : 2)	No reaction	—
13	1m	CuI (2)	DIPEA (2)	H ₂ O : CH ₃ CN (1 : 2)	No reaction	—
14	1l	CuSO ₄ · 5H ₂ O (0.1)	NaAsc (0.2)	H ₂ O : <i>t</i> -BuOH (1 : 5)	13	73%
15	1m	CuSO ₄ · 5H ₂ O (0.1)	NaAsc (0.2)	H ₂ O : <i>t</i> -BuOH (1 : 5)	14	75%
16	1n	CuSO ₄ · 5H ₂ O (0.2)	NaAsc (0.4)	H ₂ O : <i>t</i> -BuOH (1 : 1)	15	82%
17	1o	CuSO ₄ · 5H ₂ O (0.2)	NaAsc (0.4)	H ₂ O : <i>t</i> -BuOH (1 : 2)	16	79%



compound **1** in the same solvent system, no triazole product was observed, prompting us to explore different click reaction conditions. Later, we found the phenyl acetylenes (**1l–m**) successfully reacted with compound **1** when catalytic amounts of $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ (10%) and sodium ascorbate (20%) were used in water and *tert*-butanol (1 : 5, v/v), leading to the formation of triazole-glucoside derivatives **13** (ref. 40) and **14** (Fig. 4, Table 1). In the above reaction conditions in water : *tert*-butanol (1 : 1 and 1 : 2, v/v), the compounds **15** and **16** (ref. 41) (Fig. 4, Table 1), were obtained by joining compound **1** with mono-propargylated fluorescein and di-propargylated hydroquinone (**1n** and **1o**), respectively.

It should be noted that compounds such as **4**, **5** and **15** (Fig. 4), could be difficult to produce from click reactions with tetra-acetyl-1-azido- β -D-glucose, since the internal ester and lactone groups may be liable during deprotection reactions. For instance, the deprotection of the glucose in generation of the glucose ester of gibberellin **4** (GA_4 , the starting material for compound **4**) led to relatively low yield.⁷ In fact, for click chemistry, the tetra-acetyl-1-azido- β -D-glucose could be deprotected before coupling and there are synthetic routes to 1-azido- β -D-glucose directly.^{8–12} However, the enzymatic generation of compound **1** reported here represents a convenient method for its generation. Compound **1** can then be coupled to such labile groups without the need for deprotection.

Conclusions

The enzyme TxGH116D593A was successfully produced in a multi-milligram scale and subsequently utilized for the gram scale synthesis of 1-azido- β -D-glucose (**1**) in aqueous medium. The compound **1** was further conjugated with different alkynes *via* a 1,2,3-triazole as a linker. The complete process proceeds in a safe and convenient way for the generation of novel 1,2,3-triazole- β -D-glucosyl derivatives (**2–16**) in a short route by avoiding deprotection steps. In addition, compound **1** was produced from pNPGlc in an inexpensive way, when compared with the commercial price of the product. This is the foremost method developed for the gram scale synthesis of compound **1** using an enzyme (TxGH116D593A), and its utilization for the synthesis of glucoconjugates. Compound **1** is widely used in different fields of synthesis and **2–16** are under screening for their inhibitory activities.

Methods and experimental section

Silica gel 60 F₂₅₄ aluminum TLC plates were used to monitor the reactions with short-wavelength ultraviolet light and by charring the TLC plate after spraying with 10% sulfuric acid in ethanol to visualize the spots. Column chromatography was performed on silica gel 60–200 mesh and Sephadex LH-20 resin. ¹H NMR and spectra were recorded at 400 MHz and 600 MHz and ¹³C NMR spectra at 100 MHz, 125 MHz and 150 MHz on a Bruker Avance III 600 MHz instrument with a Cryo probe Prodigy. Either deuterated water (D_2O), or deuterated DMSO (CD_3SO) was used as the solvent. Chemical shifts are given in parts per million and coupling constants in hertz. HR-ESI-MS

analysis was performed on a Thermo Scientific Exactive Mass Spectrometer with ions denoted in *m/z*.

Process for the gram scale production of 1-azido- β -D-glucose (**1**) from pNP-glucose using enzyme (TxGH116D593A) and purification of **1**

The TxGH116D593A enzyme was produced and purified as previously described,⁶ but without cleavage and removal of the fusion tag. Compound **1** was produced in reactions with 100 mM NaN_3 (since higher concentrations did not produce significantly more product, data not shown), in 50 mM 4-morpholineethanesulphonic acid (MES buffer), pH 5.5, and varying concentrations of pNPGlc (Henan Tianfu Chemical Co. Ltd., Zhengzhou, China) substrate and TxGH116D593A enzyme. For the large-scale reaction, 5 g pNPGlc (45 mM) was reacted with 370 mg TxGH116D593A (0.01 mM) as catalyst in a total volume of 370 mL in a 1 L Duran bottle (see ESI Fig. S1†). After 24 hours of incubation in a water bath at 55 °C, the yellow-colored reaction mixture was frozen and subjected to lyophilization (ESI Fig. S1†) for 48 h. The resulting yellow-colored dried mixture was redissolved in methanol, the insoluble material was removed by filtration through cotton and the soluble portion was adsorbed on silica gel and loaded onto a silica gel column.

Purification of 1-azido- β -D-glucose (**1**) using silica gel column chromatography

The silica gel with the adsorbed yellow-colored crude mixture was loaded onto a silica gel (60–200 mesh) column (18.4 cm \times 4 cm), in ethyl acetate (EtOAc, 100%). The column was washed with 100% EtOAc (300 mL) to elute the *p*-nitrophenol, and then with EtOAc/MeOH (95 : 5, v/v, 5 \times 150 mL) to elute the unreacted pNPGlc in a mixture with 1-azido- β -D-glucose (**1**). Elution with EtOAc/MeOH (94 : 6, v/v, 5 \times 150 mL) gave pure product (see ESI Fig. S1†). Finally, the column was eluted with EtOAc/MeOH (1 : 1, v/v) to elute the trace of glucose (see ESI Fig. S1†). The compound was monitored by charring the TLC plate after spraying with 10% (v/v) sulfuric acid in ethanol. The eluent fractions containing pure compound **1** and compound **1** mixed with pNPGlc were separately concentrated under reduced pressure. The pure 1-azido- β -D-glucose (**1**) dried to a white foamy solid (800 mg), $R_f = 0.36$ ($\text{CHCl}_3/\text{MeOH}/\text{NH}_3$, 7 : 2.8 : 0.2, v/v/v). The mixture of pNPGlc and compound **1** dried to yield white powder (2.645 g) (ESI Fig. S1†).

Purification of mixture of 1-azido- β -D-glucose (**1**) and pNP- β -D-glucoside by Sephadex LH-20 column chromatography

The obtained mixture (2.645 g, ESI Fig. S1†) was redissolved in 6 mL of water : methanol (95 : 5, v/v) and loaded onto an 18 cm \times 3.5 cm Sephadex LH-20 column (see ESI Fig. S2†). The column was eluted with 2 \times 100 mL of water/methanol (95 : 5, v/v). The product (**1**) eluted in pure form, after which the minor amount of pNPGlc eluted. The product purity in the selected fractions was shown by TLC (ESI Fig. S2†), and these fractions were pooled into a round bottom flask, and freeze-dried, resulting in 2.470 g of compound **1** as a white foamy solid (ESI Fig. S2†).



1-Azido- β -D-glucose (1)

^1H NMR (D_2O , 400 MHz): δ 4.55 (H-1, d, J = 9 Hz, 1H), 3.73 (H-6a, dd, J = 2, 12.4 Hz, 1H), 3.55 (H-6b, dd, J = 5.6, 12.8, Hz, 1H), 3.37–3.30 (H-4, H-5, m, 2H), 3.25 (H-2, appt, J = 9.4 Hz, 1H), and 3.072 (H-3, appt, J = 8.5 Hz, 1H); ^{13}C NMR (D_2O , 600 MHz). In the 600 MHz NMR analysis, the anomeric proton peak merged with the solvent peak. The chemical shifts of the observed peaks were δ 3.89 (H-6a, d, J = 12 Hz, 1H), 3.72 (H-6b, d, J = 12 Hz, 1H), 3.51–3.48 (H-4, H-5, m, 2H), 3.40 (H-2, appt, J = 9 Hz, 1H), and 3.24 (H-3, appt, J = 9 Hz, 1H); ^{13}C NMR (D_2O , 150 MHz) δ 90.1, 77.9, 75.7, 72.8, 69.2 and 60.6; HR-ESI-MS m/z [$\text{M} + \text{Na}$] $^+$ $\text{C}_6\text{H}_{11}\text{N}_3\text{O}_5\text{Na}$ calc'd for 228.05964, found 228.05975.

Compound (2)

To a solution of **1** [26 mg, 0.126 mmol, 1 equivalent (eq.)] in water/acetonitrile (1 : 2, 2 mL) was added propargyl alcohol **1a** (22 μL , 0.38 mmol, 3 eq.), followed by copper iodide (48 mg, 0.252 mmol, 2 eq.) and DIPEA (44 μL , 0.252 mmol, 2 eq.), and the resulting yellow color reaction mixture was stirred at room temperature overnight. TLC indicated the complete conversion of the starting material into a polar compound. The reaction mixture was diluted with 5 mL methanol and then passed through a celite pad by washing with 10 mL methanol; the filtrate was concentrated under reduced pressure. The resulting crude product was redissolved in 4 mL methanol and adsorbed on silica gel and subjected to purification by silica gel chromatography in solvent ethyl acetate/methanol (80 : 20, v/v), which gave rise to pure compound **2** (31 mg, 94% yield) as a white solid. R_f = 0.13 (EtOAc/MeOH, 9 : 1, v/v); ^1H NMR (D_2O , 600 MHz) δ 8.39 (s, 1H), 5.94 (d, J = 8 Hz, 1H), 4.94 (s, 2H), 4.24–4.20 (m, 1H), 4.09 (d, J = 11.7 Hz, 1H), 3.97–3.90 (m, 3H), and 3.82–3.81 (m, 1H); ^{13}C NMR (D_2O , 150 MHz) δ 147.2, 123.4, 87.5, 78.9, 76.0, 72.3, 69.0, 60.5 and 54.7; HR-ESI-MS m/z [$\text{M} + \text{Na}$] $^+$ calc'd for $\text{C}_9\text{H}_{15}\text{N}_3\text{O}_6\text{Na}$ 284.08585, found 284.08640.

Compound 3

To a solution of **1** (25 mg, 0.121 mmol, 1 eq.) and propargylcinnamide (**1b**) (45 mg, 0.243 mmol, 2 eq.) in water/acetonitrile (1 : 2, v/v, 2 mL), were added copper iodide (46 mg, 0.243 mmol, 2 eq.) and DIPEA (42 μL , 0.243 mmol, 2 eq.), and the resulting yellow colored reaction mixture was stirred at room temperature for 2.5 h. TLC indicated the formation of a polar compound. The reaction mixture was filtered through a celite pad, washed with methanol (15 mL), adsorbed on silica gel and purified by column chromatography using ethyl acetate/methanol (90 : 10, v/v), which afforded compound **3** (28 mg, 60%) as a white solid. R_f = 0.2 (EtOAc/MeOH, 8.5 : 1.5, v/v); ^1H NMR (D_2O , 600 MHz) δ 8.14 (s, 1H), 7.60 (bs, 2H), 7.51 (d, J = 15.7 Hz, 1H), 7.51 (bs, 3H), 6.61 (d, J = 15.7 Hz, 1H), 5.72 (d, J = 8.7 Hz, 1H), 4.59 (s, 2H), 3.97 (appt, J = 8.5 Hz, 1H), 3.88 (d, J = 12 Hz, 1H), 3.76–3.68 (m, 3H), and 3.60 (appt, J = 9 Hz, 1H); ^{13}C NMR (D_2O , 150 MHz) δ 168.8, 141.6, 134.3, 130.3, 129.0, 128.0, 123.1, 119.8, 87.4, 78.9, 75.9, 72.3, 68.9, 60.4, and 34.6; HR-ESI-MS m/z [$\text{M} + \text{Na}$] $^+$ calc'd for $\text{C}_{18}\text{H}_{22}\text{N}_4\text{NaO}_6$ 413.14370, found 413.14374.

Compound 4

To a solution of **1** (20 mg, 0.097 mmol, 1.2 eq.) and compound **1c** (30 mg, 0.081 mmol, 1 eq.) in water/acetonitrile (1 : 2, v/v, 1.5 mL), were added copper iodide (31 mg, 0.162 mmol, 2 eq.) and DIPEA (28 μL , 0.162 mmol, 2 eq.), and the resulting yellow colored reaction mixture was stirred at room temperature 1 h. TLC indicated the almost complete conversion of the starting material into a polar compound. The reaction mixture was filtered through a celite pad, washed with methanol (15 mL), adsorbed on silica gel, and purified by chromatography using ethyl acetate/methanol (90 : 10, v/v) as solvent. Compound **4** (26 mg, 46.5% yield) was obtained as a pale brown solid. It had a silica gel TLC R_f = 0.30 (EtOAc/MeOH 9 : 1, v/v, as solvent); ^1H NMR (D_2O , 600 MHz): δ 8.12 (s, 1H), 5.71 (d, J = 8.7 Hz, 1H), 4.99 (s, 1H), 4.85 (s, 1H), 4.52–4.44 (m, 3H), 3.96 (appt, J = 8.5 Hz, 1H), 3.90 (d, J = 12 Hz, 1H), 3.80–3.75 (m, 2H), 3.71–3.68 (m, 2H), 3.62–3.59 (m, 1H), 3.07 (d, J = 10 Hz, 1H), 2.62–2.60 (m, 2H), 2.06–2.04 (m, 2H), 1.91–1.85 (m, 2H), 1.80–1.56 (m, 6H), 1.35–1.29 (m, 2H), and 1.03 (s, 3H); ^{13}C NMR (D_2O , 150 MHz): δ 181.9, 174.1, 158.3, 123.4, 106.5, 97.1, 87.4, 78.9, 75.9, 72.2, 69.6, 69.0, 60.5, 55.0, 53.0, 52.5, 51.3, 51.1, 43.6, 38.8, 36.6, 34.3, 30.8, 27.2, 26.5, 15.6, and 13.7; HR-ESI-MS m/z [$\text{M} + \text{Na}$] $^+$ calc'd for $\text{C}_{28}\text{H}_{38}\text{N}_4\text{NaO}_9$ 597.25365, found 597.25415.

Compound 5

To a solution of **1** (25 mg, 0.121 mmol, 1 eq.) and propargylbenzoate **1d** (35 mg, 0.304 mmol, 1.8 eq.) in water/acetonitrile (1 : 2, v/v, 2 mL), were added copper iodide (46 mg, 0.243 mmol, 2 eq.) and DIPEA (42 μL , 0.243 mmol, 2 eq.), and the resulting yellow color reaction mixture was stirred at room temperature overnight. TLC indicated the formation of a polar compound. The reaction mixture was filtered through celite pad, washed with methanol (20 mL), and adsorbed on silica gel. Purification by silica gel chromatography using ethyl acetate/methanol (90 : 10, v/v) as solvent gave rise to compound **5** (19 mg, 43% yield) as a white solid. Silica gel TLC gave an R_f = 0.26 (EtOAc/MeOH, 9 : 1, v/v as solvent); ^1H NMR: (D_2O , 600 MHz) δ 8.37 (s, 1H), 8.01 (d, J = 6.6 Hz, 2H), 7.67 (s, 1H), 7.51 (d, J = 6.1 Hz, 2H), 5.78 (d, J = 8.8 Hz, 1H), 5.51 (s, 2H), 4.01 (appt, J = 8.6 Hz, 1H), 3.91 (d, J = 12 Hz, 1H), 3.79–3.71 (m, 3H), 3.63 (appt, J = 8.9 Hz, 1H); ^{13}C NMR (D_2O , 150 MHz) δ 168.2, 134.0, 129.5, 129.0, 128.8, 125.0, 87.6, 78.9, 75.9, 72.3, 69.0, 60.5, and 57.9; HR-ESI-MS m/z [$\text{M} + \text{Na}$] $^+$ calc'd for $\text{C}_{16}\text{H}_{19}\text{N}_3\text{NaO}_7$ 388.11207 amu, found 388.11189.

Compound 6

To a solution of **1** (18 mg, 0.087 mmol, 1 eq.) and propargylbenzyl ether **1e** (63 mg, 0.435 mmol, 5 eq.) in water/acetonitrile (1 : 2, v/v, 2 mL), were added copper iodide (33 mg, 0.174 mmol, 2 eq.) and DIPEA (30 μL , 0.174 mmol, 2 eq.), and the resulting yellow colored reaction mixture was stirred at room temperature overnight. TLC indicated formation of a polar compound, just below to the starting material. The reaction mixture was filtered through a celite pad, washed with methanol (15 mL), adsorbed on silica gel and purified by silica



gel chromatography using ethyl acetate/methanol (90 : 10, v/v) as solvent to afford compound **6** (20 mg, 66% yield) as a pale yellow viscous solid. Silica gel TLC showed an $R_f = 0.25$ (EtOAc/MeOH, 9 : 1, v/v, as solvent); ^1H NMR (D_2O , 600 MHz): δ 8.24 (s, 1H), 7.44–7.42 (m, 5H), 5.76 (d, $J = 8.9$ Hz, 1H), 4.75 (s, 2H, merged with solvent peak), 4.65 (s, 2H), 4.01 (appt, $J = 8.9$ Hz, 1H), 3.92 (d, $J = 12$ Hz, 1H), 3.80–3.71 (m, 3H), and 3.64 (appt, $J = 9$ Hz, 1H); ^{13}C NMR (D_2O , 150 MHz): δ 136.9, 128.8, 128.7, 128.4, 124.5, 87.4, 78.9, 75.9, 72.3, 72.3, 69.0, 62.2, and 60.4; HR-ESI-MS m/z $[\text{M} + \text{Na}]^+$ calc'd for $\text{C}_{16}\text{H}_{21}\text{N}_3\text{NaO}_6$ 374.13281, found 374.13373.

Compound 7

To a solution of **1** (27 mg, 0.131 mmol, 1 eq.) and propargyl-indole-3-acetamide **1f** (45 mg, 0.210 mmol, 1.6 eq.) in water/acetonitrile (1 : 2, v/v, 2 mL), were added copper iodide (50 mg, 0.263 mmol, 2 eq.) and DIPEA (46 μL , 0.263 mmol, 2 eq.), and the resulting yellow colored mixture was stirred at room temperature for 4 h. Silica gel TLC indicated the almost complete conversion of the starting material into a polar compound. The reaction mixture was filtered through celite pad, washed with methanol (20 mL), and adsorbed on silica gel. Purification by silica gel chromatography using ethyl acetate/methanol (90 : 10, v/v) as solvent afforded compound **7** (27.3 mg, 51% yield) as a pale brown solid. Silica gel TLC showed an $R_f = 0.30$ (with EtOAc/MeOH, 9 : 1, v/v, as solvent); ^1H NMR (D_2O , 600 MHz): δ 7.68 (s, 1H), 7.50 (appt, $J = 8.4$ Hz, 2H), 7.28 (s, 1H), 7.23 (appt, $J = 7.2$ Hz, 1H), 7.12 (appt, $J = 7.2$ Hz, 1H), 5.58 (d, $J = 9$ Hz, 1H), 4.42 (s, 2H), 3.86 (d, $J = 11.4$ Hz, 1H), 3.82–3.79 (m, 4H), 3.66–3.62 (m, 2H), and 3.55 (appt, $J = 9.6$ Hz, 1H); ^{13}C NMR (D_2O , 150 MHz): δ 175.4, 145.3, 136.3, 126.5, 125.1, 122.4, 122.2, 119.5, 118.3, 112.0, 107.5, 87.3, 78.8, 75.8, 72.2, 68.9, 60.4, 34.6, and 32.5; HR-ESI-MS m/z $[\text{M} + \text{Na}]^+$ calc'd for $\text{C}_{19}\text{H}_{23}\text{N}_5\text{NaO}_6$ 440.15460, found 440.15543.

Compound 8

To a solution of **1** (22 mg, 0.107 mmol, 1 eq.) and propargyl-tryptophanamide **1g** (44 mg, 0.128 mmol, 1.2 eq.) in water/acetonitrile (1 : 2, v/v, 2 mL), were added copper iodide (41 mg, 0.214 mmol, 2 eq.) and DIPEA (37 μL , 0.214 mmol, 2 eq.), and the resulting reaction mixture was stirred at room temperature for 30 min. TLC indicated the complete conversion of starting material into a polar compound. The reaction mixture was filtered through celite pad, washed with methanol (25 mL), and adsorbed on silica gel. Purification by chromatography using ethyl acetate/methanol (92 : 8, v/v) as solvent afforded compound **8** (35.5 mg, 62% yield) as a white solid. Silica gel TLC gave an $R_f = 0.2$ (EtOAc/MeOH, 8.5 : 1.5, v/v, as solvent); ^1H NMR (D_2O , 600 MHz): δ 7.60 (d, $J = 7.8$ Hz, 1H), 7.50 (s, 1H), 7.46 (d, $J = 8.4$ Hz, 1H), 7.22 (appt, $J = 7.2$ Hz, 1H), 7.13 (appt, $J = 7.2$ Hz, 1H), 7.05 (s, 1H), 5.64 (d, $J = 9$ Hz, 1H), 4.40–4.18 (m, 3H), 3.97–3.93 (m, 2H), 3.78 (dd, $J = 5.4, 12.6$ Hz, 1H), 3.74–3.73 (m, 2H), 3.69 (appt, $J = 9$ Hz, 1H), 3.61 (appt, $J = 9$ Hz, 1H), 3.14 (s, 2H), 1.34 (s, 6H) and 1.22 (s, 3H); ^{13}C NMR (D_2O , 150 MHz): δ 174.4, 157.2, 136.0, 126.7, 124.4, 121.8, 119.2, 118.3, 111.8, 108.7, 87.3, 78.8, 75.9, 72.1, 68.9, 60.4, 55.6, 34.1, 27.5,

and 27.1; HR-ESI-MS m/z $[\text{M} + \text{Na}]^+$ calc'd for $\text{C}_{25}\text{H}_{34}\text{N}_6\text{NaO}_8$ 569.23358, found 569.23474.

Compound 9

To a solution of **1** (25 mg, 0.121 mmol, 1 eq.) and *p*-nitro-benzyl-propargylamine **1h** (32 mg, 0.18 mmol, 1.5 eq.) in water/acetonitrile (1 : 2, v/v, 2 mL), were added copper iodide (46 mg, 0.242 mmol, 2 eq.) and DIPEA (42 μL , 0.242 mmol, 2 eq.), and the resulting yellow colored reaction mixture was stirred at room temperature for 6 h. TLC indicated the completion of starting material into a polar compound. The reaction mixture was filtered through celite pad, washed with methanol (20 mL), and adsorbed on silica gel. Purification by silica gel chromatography using ethyl acetate/methanol (70 : 30, v/v) as solvent afforded compound **9** (35 mg, 74% yield) as white solid. Silica gel TLC gave an $R_f = 0.29$ (EtOAc/MeOH, 7 : 3, v/v, as solvent); ^1H NMR (D_2O , 600 MHz): δ 8.23 (d, $J = 6.9$ Hz, 1H), 8.1 (s, 1H), 7.56 (d, $J = 6.7$ Hz, 1H), 5.75 (d, $J = 8.5$ Hz, 1H), 4.94 (s, 2H), 4.06–4.02 (m, 3H), 3.93 (d, $J = 12$ Hz, 1H), 3.81–3.71 (m, 3H), and 3.64 (appt, $J = 9$ Hz, 1H); ^{13}C NMR (D_2O , 150 MHz): δ 147.2, 129.7, 123.9, 87.5, 79.0, 76.0, 72.3, 69.0, 60.5, 54.5, and 50.9; HR-ESI-MS m/z $[\text{M} + \text{Na}]^+$ calc'd for $\text{C}_{16}\text{H}_{21}\text{N}_5\text{NaO}_7$ 418.13387, found 418.13445.

Compound 10

To a solution of **1** (26 mg, 0.126 mmol, 1 eq.) and Boc-propargylamine **1i** (59 mg, 0.378 mmol, 3 eq.) in water/acetonitrile (1 : 2, v/v, 2 mL) were added copper iodide (48 mg, 0.253 mmol, 2 eq.) and DIPEA (44 μL , 0.253 mmol, 2 eq.), and the resulting yellow colored reaction mixture was stirred at room temperature for 3 h. TLC indicated the complete conversion of the starting material into a polar compound. The reaction mixture was filtered through celite pad, washed with methanol (15 mL), and adsorbed on silica gel. Purification by silica gel chromatography using ethyl acetate/methanol (96 : 4, v/v) as solvent afforded compound **10** (44 mg, 96% yield) as a white solid. Silica gel TLC demonstrated an $R_f = 0.35$ (with EtOAc/MeOH, 8.5 : 1.5, v/v, as solvent); ^1H NMR (D_2O , 600 MHz): δ 8.11 (s, 1H), 5.72 (d, $J = 8.9$ Hz, 1H), 4.36 (s, 2H), 3.99 (appt, $J = 8.5$ Hz, 1H), 3.91 (d, $J = 12$ Hz, 1H), 3.79–3.70 (m, 3H), 3.62 (appt, $J = 9$ Hz, 1H), and 1.47 (s, 9H); ^{13}C NMR (D_2O , 150 MHz): δ 158.0, 122.9, 87.4, 78.9, 75.9, 72.3, 69.0, 60.4, and 27.6; HR-ESI-MS m/z $[\text{M} + \text{Na}]^+$ calc'd for $\text{C}_{14}\text{H}_{24}\text{N}_4\text{NaO}_7$ 383.15427, found 383.15500.

Compound 11

To a solution of **1** (20 mg, 0.097 mmol, 1 eq.) and propargyl-benzamide **1j** (31 mg, 0.195 mmol, 2 eq.) in water/acetonitrile (1 : 2, v/v, 2 mL), were added copper iodide (37 mg, 0.195 mmol, 2 eq.) and DIPEA (34 μL , 0.195 mmol, 2 eq.), and the resulting yellow colored reaction mixture was stirred at room temperature for 40 minutes. TLC indicated the complete conversion of the starting material into a polar compound. The reaction mixture was filtered through celite pad, washed with methanol (15 mL), and adsorbed on silica gel. Purification by silica gel chromatography using ethyl acetate/methanol (90 : 10, v/v) as solvent afforded compound **11** (28 mg, 80% yield) as



a white solid. Silica gel TLC demonstrated an $R_f = 0.27$ (with EtOAc/MeOH, 8.5 : 1.5, v/v, as solvent); ^1H NMR (D_2O , 600 MHz): δ 8.15 (s, 1H), 7.76 (d, $J = 7.2$ Hz, 2H), 7.59 (appt, $J = 7.6$ Hz, 1H), 7.50 (appt, $J = 8.4$ Hz, 2H), 5.71 (d, $J = 9.6$ Hz, 1H), 4.68 (s, 2H), 3.96 (appt, $J = 9$ Hz, 1H), 3.87 (d, $J = 11.4$ Hz, 1H), 3.76–3.66 (m, 3H), and 3.59 (appt, $J = 9.6$ Hz, 1H); ^{13}C NMR (D_2O , 150 MHz): δ 170.9, 145.1, 133.2, 132.3, 128.8, 127.1, 123.1, 87.4, 78.8, 75.9, 72.2, 68.9, 60.4, and 34.9; HR-ESI-MS m/z [$\text{M} + \text{Na}$] $^+$ calc'd for $\text{C}_{16}\text{H}_{20}\text{N}_4\text{NaO}_6$ 387.12805, found 387.12903.

Compound 12

To a solution of **1** (50 mg, 0.243 mmol, 1 eq.) and PMB-propargylether **1k** (129 mg, 0.731 mmol, 3 eq.) in water/acetonitrile (1 : 2, v/v, 4 mL) were added copper iodide (93 mg, 0.486 mmol, 2 eq.) and DIPEA (85 μL , 0.486 mmol, 2 eq.), and the resulting yellow colored reaction mixture was stirred at room temperature overnight. TLC indicated the complete conversion of the starting material into a polar spot. The reaction mixture was filtered through a celite pad, washed with methanol (25 mL), and adsorbed on silica gel. Purification by silica gel chromatography using ethyl acetate/methanol (90 : 10, v/v) as solvent afforded compound **12** (87 mg, 93% yield) as a pale yellow solid. Silica gel TLC demonstrated an $R_f = 0.27$ (with EtOAc/MeOH, 8.5 : 10, v/v, as solvent); ^1H NMR (D_2O , 600 MHz): δ 8.17 (s, 1H), 7.32 (d, $J = 8.4$ Hz, 2H), 6.98 (d, $J = 7.8$ Hz, 2H), 5.71 (d, $J = 9.6$ Hz, 1H), 4.68 (s, 2H), 4.54 (s, 2H), 3.96 (appt, $J = 9$ Hz, 1H), 3.88 (d, $J = 12.4$ Hz, 1H), 3.80 (s, 3H), 3.76–3.66 (m, 3H), and 3.59 (appt, $J = 9$ Hz, 1H); ^{13}C NMR (D_2O , 150 MHz): δ 158.9, 144.4, 130.4, 129.5, 124.4, 87.4, 78.9, 75.9, 72.2, 71.9, 68.9, 61.9, 60.4, and 55.4; HR-ESI-MS m/z [$\text{M} + \text{Na}$] $^+$ calc'd for $\text{C}_{17}\text{H}_{23}\text{N}_3\text{NaO}_7$ is 404.14337, found 404.14513.

Compound 13

A solution of **1** (25 mg, 0.0121 mmol, 1 eq.) and 4-nitrophenylacetylene **1l** (27 mg, 0.182 mmol, 1.5 eq.) in water/*tert*-butanol (1 : 5, v/v, 3.5 mL) was heated in a 50 °C water bath until the reaction mixture became homogeneous and then $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ (3 mg, 0.0121 mmol, 0.1 eq.) and sodium ascorbate (5 mg, 0.0242 mmol, 0.2 eq.), were added and the resulting colorless reaction mixture was stirred at room temperature overnight. TLC indicated the complete conversion of the starting material into a polar compound. The reaction mixture was filtered through celite pad, washed with methanol (15 mL), and adsorbed on silica gel. Purification by silica gel chromatography using ethyl acetate/methanol (94 : 6, v/v) as solvent afforded compound **13** (31 mg, 73% yield) as a white solid. Silica gel TLC (ran twice), demonstrated an $R_f = 0.20$ (with EtOAc/MeOH, 9 : 1, v/v, as solvent); ^1H NMR (DMSO- d_6 , 600 MHz): δ 9.13 (s, 1H), 8.34 (d, $J = 8.4$ Hz, 2H), 8.17 (d, $J = 7.8$ Hz, 2H), 5.62 (d, $J = 9.6$ Hz, 1H), 5.46 (d, $J = 6$ Hz, 1H), 5.33 (d, $J = 4.8$ Hz, 1H), 5.18 (d, $J = 6$ Hz, 1H), 4.62 (appt, $J = 6$ Hz, 1H), 3.82–3.78 (m, 1H), 3.73 (dd, $J = 5.4$, 11.4 Hz, 1H), 3.53–3.43 (m, 3H), and 3.28–3.25 (m, 1H); ^{13}C NMR (DMSO- d_6 , 150 MHz): δ 147.2, 144.9, 137.4, 126.4, 124.9, 123.1, 88.3, 80.4, 77.2, 72.8, 72.0, and 61.2; HR-ESI-MS m/z [$\text{M} + \text{Na}$] $^+$ calc'd for $\text{C}_{14}\text{H}_{16}\text{N}_4\text{NaO}_7^+$ 375.0916, found 375.0919.

Compound 14

A solution of **1** (25 mg, 0.0121 mmol, 1 eq.) and 4-cyanophenylacetylene **1m** (23 mg, 0.182 mmol, 1.5 eq.) in water/*tert*-butanol (1 : 5, v/v, 3 mL) was heated in a water bath until the reaction mixture became homogeneous. Then, $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ (3 mg, 0.0121 mmol, 0.1 eq.) and sodium ascorbate (5 mg, 0.0242 mmol, 0.2 eq.) were added and the resulting colorless reaction mixture was stirred at room temperature overnight. TLC indicated the complete conversion of the starting material into a polar compound. The reaction mixture was filtered through celite pad, washed with methanol (20 mL), and adsorbed on silica gel. Purification by silica gel chromatography using ethyl acetate/methanol (94 : 6, v/v) as solvent afforded compound **14** (30 mg, 75% yield) as a white solid. Silica gel TLC (ran twice), demonstrated an $R_f = 0.20$ (with EtOAc/MeOH, 9 : 1, v/v, as solvent); ^1H NMR (DMSO- d_6 , 600 MHz): δ 9.05 (s, 1H), 8.09 (d, $J = 7.8$ Hz, 2H), 7.94 (d, $J = 7.8$ Hz, 2H), 5.61 (d, $J = 9$ Hz, 1H), 5.45 (d, $J = 5.4$ Hz, 1H), 5.33 (d, $J = 2.4$ Hz, 1H), 5.17 (d, $J = 4.8$ Hz, 1H), 4.63 (d, $J = 5.4$ Hz, 1H), 3.81–3.77 (m, 1H), 3.72 (dd, $J = 5.4$, 10.8 Hz, 1H), 3.52–3.42 (m, 3H), and 3.29–3.25 (m, 1H); ^{13}C NMR (DMSO- d_6 , 150 MHz): δ 145.3, 135.5, 133.5, 126.2, 122.7, 119.2, 110.7, 88.2, 80.4, 77.2, 72.7, 70.0, and 61.2; HR-ESI-MS m/z [$\text{M} + \text{Na}$] $^+$ calc'd for $\text{C}_{15}\text{H}_{16}\text{N}_4\text{NaO}_5$ 355.10184, found 355.10254.

Compound 15

To a solution of **1** (20 mg, 0.097 mmol, 1 eq.) and mono-propargylated fluorescein **1n** (20 mg, 0.048 mmol, 0.5 eq.) in water/*tert*-butanol (1 : 1, v/v, 2 mL) were added $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ (3 mg, 0.0097 mmol, 0.2 eq.) and sodium ascorbate (4 mg, 0.019 mmol, 0.4 eq.), and the resulting yellow colored reaction mixture was stirred at room temperature overnight. TLC indicated the complete conversion of the starting material into a fluorescent polar compound. The reaction mixture was filtered through celite pad, washed with methanol (20 mL) and adsorbed on silica gel. Purification by silica gel chromatography using ethyl acetate/methanol (94 : 6, v/v) as solvent afforded compound **15** (23 mg, 82% yield) as a yellow solid. Silica gel TLC demonstrated an $R_f = 0.22$ (with EtOAc/MeOH, 8.5 : 1.5, v/v, as solvent); ^1H NMR (DMSO- d_6 , 600 MHz): δ 10.15 (s, 1H), 8.47 (s, 1H), 8.01 (d, $J = 7.8$ Hz, 1H), 7.80 (appt, $J = 7.2$ Hz, 1H), 7.73 (appt, $J = 7.2$ Hz, 1H), 7.28 (d, $J = 7.8$ Hz, 1H), 7.12 (s, 1H), 6.79 (d, $J = 8.4$ Hz, 1H), 6.72 (s, 1H), 6.66 (d, $J = 8.4$ Hz, 1H), 6.58 (s, 2H), 5.56 (d, $J = 9.6$ Hz, 1H), 5.39 (d, $J = 6$ Hz, 1H), 5.26 (d, $J = 4.2$ Hz, 1H), 5.23 (s, 2H), 4.62–4.61 (m, 1H), 3.80–3.76 (m, 1H), 3.69 (dd, $J = 5.4$, 9 Hz, 1H), 3.46–3.37 (m, 4H), and 3.26–3.23 (m, 1H); ^{13}C NMR (DMSO- d_6 , 150 MHz): δ 169.1, 160.2, 160.0, 152.9, 152.3, 152.2, 142.6, 136.1, 130.6, 129.5, 129.4, 126.5, 125.1, 124.5, 124.4, 113.3, 112.8, 111.8, 109.9, 102.7, 102.0, 88.0, 83.1, 80.4, 77.4, 72.5, 70.0, 61.8, and 61.2; HR-ESI-MS m/z [$\text{M} + \text{H}$] $^+$ calc'd for $\text{C}_{29}\text{H}_{26}\text{N}_3\text{O}_{10}$ 576.16182, found 576.16227.

Compound 16

To a solution of **1** (44 mg, 0.214 mmol, 2 eq.) and dipropargylated-hydroquinone **1o** (20 mg, 0.107 mmol, 1 eq.) in



water/*tert*-butanol (1 : 2, v/v, 2 mL), were added CuSO₄·5H₂O (5.3 mg, 0.0214 mmol, 0.2 eq.) and sodium ascorbate (8.5 mg, 0.0428 mmol, 0.4 eq.), and the resulting reaction mixture was stirred at room temperature for 48 h. TLC indicated the formation of a polar compound. The reaction mixture was filtered through a celite pad, which was washed with methanol (40 mL), and the solution adsorbed on silica gel. Purification by silica gel chromatography using ethyl acetate/methanol (80 : 20, v/v) as solvent afforded compound **16** (51 mg, 79% yield) as a white solid. Silica gel TLC demonstrated an *R*_f = 0.28 (with EtOAc/MeOH, 6 : 4, v/v, as solvent); ¹H NMR (D₂O, 600 MHz): δ 8.25 (s, 2H), 6.99 (s, 4H), 5.71 (d, *J* = 9 Hz, 2H), 5.20 (s, 4H), 3.96 (appt, *J* = 9 Hz, 2H), 3.87 (d, *J* = 12 Hz, 2H), 3.76–3.66 (m, 6H), and 3.59 (appt, *J* = 9 Hz, 2H); ¹³C NMR (D₂O, 150 MHz): 152.3, 143.7, 124.5, 116.9, 87.4, 79.9, 75.9, 72.2, 68.9, 62.0, 60.4, 54.4, and 42.5; HR-ESI-MS *m/z* [M + Na]⁺ calc'd for C₂₄H₃₂N₆NaO₁₂ 619.21975, found 619.19861.

Conflicts of interest

The authors have filed a patent application for the process described in this paper (TH1801005294), but otherwise have no conflicts of interest to declare.

Acknowledgements

The authors are thankful to Dr Ravi Shanker L. and Dr Yanling Hua for assistance with HRMS and NMR, respectively. Financial support from Suranaree University of Technology (SUT) and the Thailand Research Fund (TRF), Thailand, grant BRG5980015, and the National Research University project grant from the Commission on Higher Education of Thailand to SUT is gratefully acknowledged.

References

- 1 F. W. Lichtenthaler, *Angew. Chem., Int. Ed. Engl.*, 1994, **33**, 2364–2374.
- 2 R. Wombacher, *Angew. Chem., Int. Ed. Engl.*, 2006, **45**, 2469–2472.
- 3 J. R. Knowles, *Nature*, 1991, **350**, 121–124.
- 4 Q. Wang and S. G. Withers, *J. Am. Chem. Soc.*, 1995, **117**, 10137–10138.
- 5 Q. Wang, D. Trimbur, R. Graham, R. A. J. Warren and S. G. Withers, *Biochemistry*, 1995, **34**, 14554–14562.
- 6 R. Charoenwattanasatien, S. Pengthaisong, I. Breen, R. Mutoh, S. Sansenya, Y. Hua, A. Tankrathok, L. Wu, C. Songsiriritthigul, H. Tanaka, S. J. Williams, G. J. Davies, G. Kurisu and J. R. K. Cairns, *ACS Chem. Biol.*, 2016, **11**, 1891–1900.
- 7 Y. Hua, S. Sansenya, C. Saetang, S. Wakuta and J. R. Ketudat Cairns, *Arch. Biochem. Biophys.*, 2013, **537**, 39–48.
- 8 T. Tanaka, H. Nagai, M. Noguchi, A. Kobayashi and S.-I. Shoda, *Chem. Commun.*, 2009, 3378–3379.
- 9 D. Lim, M. A. Brimble, R. Kowalczyk, A. J. A. Watson and A. J. Fairbanks, *Angew. Chem., Int. Ed.*, 2014, **53**, 11907–11911.
- 10 S. Shoda, A. Kobayashi, M. Noguchi, T. Tanaka, H. Nagai and T. Matsumoto, JP2009292790A, 2009.
- 11 M.-L. Larabi, C. Frechou and G. Demailly, *Tetrahedron Lett.*, 1994, **35**, 2175–2178.
- 12 A. E. Meslouti, D. Beaupere, G. Demailly and R. Uzan, *Tetrahedron Lett.*, 1994, **35**, 3913–3916.
- 13 R. Beniazza, N. Bayo, F. Molton, C. Duboc, S. Massip, N. McClenaghan, D. Lastecoueres and J.-M. Vincent, *Beilstein J. Org. Chem.*, 2015, **11**, 1950–1959.
- 14 R. Beniazza, R. Lambert, L. Harmand, F. Molton, C. Duboc, S. Denisov, G. Jonusauskas, N. D. McClenaghan, D. Lastecoueres and J. M. Vincent, *Chem.-Eur. J.*, 2014, **20**, 13181–13187.
- 15 A. Pathigoolla, R. P. Pola and K. M. Sureshan, *Appl. Catal., A*, 2013, **453**, 151–158.
- 16 (a) B. L. Wilkinson, L. F. Bornaghi, S.-A. Poulsen and T. A. Houston, *Tetrahedron*, 2006, **62**, 8115–8125; (b) G. C. Feast, O. E. Hutt, X. Mulet, C. E. Conn, C. J. Drummond and G. P. Savage, *Chem.-Eur. J.*, 2014, **20**, 2783–2792; (c) S. Muthana, H. Yu, H. Cao, J. Cheng and X. Chen, *J. Org. Chem.*, 2009, **74**, 2928–2936.
- 17 S. Dedola, D. L. Hughes, S. A. Nepogodiev, M. Rejzek and R. A. Field, *Carbohydr. Res.*, 2010, **345**, 1123–1134.
- 18 N. Anand, N. Jaiswal, S. K. Pandey, A. K. Srivastava and R. P. Tripathi, *Carbohydr. Res.*, 2011, **346**, 16–25.
- 19 L. L. Rossi and A. Basu, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 3596–3599.
- 20 E. D. Chrysina, E. Bokor, K.-M. Alexacou, M.-D. Charavgi, G. N. Oikonomakos, S. E. Zographos, D. D. Leonidas, N. G. Oikonomakos and L. Somsak, *Tetrahedron: Asymmetry*, 2009, **20**, 733–740.
- 21 E. Bokor, T. Docsa, P. Gergely and L. Somsak, *Bioorg. Med. Chem.*, 2010, **18**, 1171–1180.
- 22 E. Bokor, E. Kyriakis, T. G. A. Solovou, C. Koppany, A. L. Kantsadi, K. E. Szabo, A. Szakacs, G. A. Stravodimos, T. Docsa, V. T. Skamnaki, S. E. Zographos, P. Gergely, D. D. Leonidas and L. Somsak, *J. Med. Chem.*, 2017, **60**, 9251–9262.
- 23 R. P. Tanpure, B. Ren, T. S. Peat, L. F. Bornaghi, D. Vullo, C. T. Supuran and S.-A. Poulsen, *J. Med. Chem.*, 2015, **58**, 1494–1501.
- 24 B. L. Wilkinson, A. Innocenti, D. Vullo, C. T. Supuran and S.-A. Poulsen, *J. Med. Chem.*, 2008, **51**, 1945–1953.
- 25 S. G. Gouin, E. Vanquelef, J. M. G. Fernandez, C. O. Mellet, F. Y. Dupradeau and J. Kovensky, *J. Org. Chem.*, 2007, **72**, 9032–9045.
- 26 A. Massarotti, S. Aprile, V. Mercalli, E. D. Grosso, G. Grosa, G. Sorba and G. C. Tron, *ChemMedChem*, 2014, **9**, 2497–2508.
- 27 P. Stefan, P. Oliver, T. Norbert, W. Ulrich and E. Frank, WO/2017/207754, 2017.
- 28 S. Cecioni, V. Oerthel, J. Iehl, M. Holler, D. Goyard, J.-P. Praly, A. Imbert, J. F. Nierengarten and S. Vidal, *Chem.-Eur. J.*, 2011, **17**, 3252–3261.
- 29 D. H. Kim, Y. S. Choe, K.-H. Jung, K.-H. Lee, J. Y. Choi, Y. Choi and B.-T. Kim, *Arch. Pharmacol. Res.*, 2008, **31**, 587–593.



- 30 S. Maschauer, R. Haubner, T. Kuwert and O. Prante, *Mol. Pharmaceutics*, 2014, **11**, 505–515.
- 31 Z. Lei, J. Wang, G. Mao, Y. Wen, Y. Tian, H. Wu, Y. Li and H. Xu, *J. Agric. Food Chem.*, 2014, **62**, 6065–6071.
- 32 C. J. Carroux, G. M. Rankin, J. Moeker, L. F. Bornaghi, K. Katneni, J. Morizzi, S. A. Charman, D. Vullo, C. T. Supuran and S.-A. Poulsen, *J. Med. Chem.*, 2013, **56**, 9623–9634.
- 33 M. Lopez, L. F. Bornaghi, A. Innocenti, D. Vullo, S. A. Charman, C. T. Supuran and S.-A. Poulsen, *J. Med. Chem.*, 2010, **53**, 5913–5926.
- 34 S. Ballet, C. Betti, A. Novoa, C. Tomboly, C. U. Nielsen, H. S. Helms, A. Lesniak, P. Kleczkowska, N. N. Chung, A. W. Lipkowski, B. Brodin, D. Tourwe and P. W. Schiller, *ACS Med. Chem. Lett.*, 2014, **5**, 352–357.
- 35 P. Thirumurugan, D. Matosiuk and K. Jozwiak, *Chem. Rev.*, 2013, **113**, 4905–4979.
- 36 V. K. Tiwari, B. B. Mishra, K. B. Mishra, N. Mishra, A. S. Singh and X. Chen, *Chem. Rev.*, 2016, **116**, 3086–3240.
- 37 G. J. Roth, B. Liepold, S. G. Muller and H. J. Bestmann, *Synthesis*, 2004, **1**, 59–62.
- 38 D. Goyard, T. Docsa, P. Gergely, J.-P. Praly and S. Vidal, *Carbohydr. Res.*, 2015, **402**, 245–251.
- 39 I. Delso, J. Valero-Gonzalez, F. Gomollon-Bel, J. Castro-Lopez, W. Fang, I. Navratilova, D. M. F. van Aalten, T. Tejero, P. Merino and R. Hurtado-Guerrero, *ChemMedChem*, 2018, **13**, 128–132.
- 40 D. Goyard, A. S. Chajistamatiou, A. I. Sotiropoulou, E. D. Chrysina, J.-P. Praly and S. Vidal, *Chem. - Eur. J.*, 2014, **20**, 1–11.
- 41 G. Yu, Y. Ma, C. Han, Y. Yao, G. Tang, Z. Mao, C. Gao and F. Huang, *J. Am. Chem. Soc.*, 2013, **135**, 10310–10313.



ARTICLE

<https://doi.org/10.1038/s41467-019-09691-z>

OPEN

Discovery of processive catalysis by an exo-hydrolase with a pocket-shaped active site

Victor A. Streltsov^{1,15}, Sukanya Luang^{2,15}, Alys Peisley¹, Joseph N. Varghese¹, James R. Ketudat Cairns³, Sebastien Fort⁴, Marcel Hijnen⁵, Igor Tvaroška⁶, Ana Ardá⁷, Jesús Jiménez-Barbero⁷, Mercedes Alfonso-Prieto⁸, Carme Rovira^{8,9}, Fernanda Mendoza^{10,11}, Laura Tiessler-Sala¹¹, José-Emilio Sánchez-Aparicio¹¹, Jaime Rodríguez-Guerra^{11,12}, José M. Lluch^{11,13}, Jean-Didier Maréchal¹¹, Laura Masgrau^{11,13} & Maria Hrmova^{2,14}

Substrates associate and products dissociate from enzyme catalytic sites rapidly, which hampers investigations of their trajectories. The high-resolution structure of the native *Hordeum* exo-hydrolase HvExol isolated from seedlings reveals that non-covalently trapped glucose forms a stable enzyme-product complex. Here, we report that the alkyl β -D-glucoside and methyl 6-thio- β -gentiobioside substrate analogues perfused in crystalline HvExol bind across the catalytic site after they displace glucose, while methyl 2-thio- β -sophoroside attaches nearby. Structural analyses and multi-scale molecular modelling of nanoscale reactant movements in HvExol reveal that upon productive binding of incoming substrates, the glucose product modifies its binding patterns and evokes the formation of a transient lateral cavity, which serves as a conduit for glucose departure to allow for the next catalytic round. This path enables substrate-product assisted processive catalysis through multiple hydrolytic events without HvExol losing contact with oligo- or polymeric substrates. We anticipate that such enzyme plasticity could be prevalent among exo-hydrolases.

¹Commonwealth Scientific and Industrial Research Organisation, Materials Science and Engineering, Parkville Victoria 3052, Australia. ²School of Agriculture, Food and Wine, University of Adelaide, Waite Campus, Glen Osmond South Australia 5064, Australia. ³School of Chemistry and Center for Biomolecular Structure, Function and Application, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand. ⁴University Grenoble Alpes, Centre de Recherches sur les Macromolécules Végétales, Grenoble cedex 9 38041, France. ⁵GE Healthcare Life Sciences, Paramatta NSW 2150, Australia. ⁶Institute of Chemistry, Slovak Academy of Sciences, Bratislava 84538, Slovak Republic. ⁷Centre for Cooperative Research in Biosciences, Derio-Bizkaia 48160, Spain. ⁸Departament de Química Inorgànica i Orgànica, Universitat de Barcelona, Barcelona 08028, Spain. ⁹Institució Catalana de Recerca i Estudis Avançats, Barcelona 08010, Spain. ¹⁰Departamento de Ciencias Químicas, Universidad Andrés Bello, Sede Concepción, Talcahuano 4260000, Chile. ¹¹Departament de Química, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain. ¹²Institute of Chemical Research of Catalonia, The Barcelona Institute of Science and Technology, Tarragona 43007, Spain. ¹³Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain. ¹⁴School of Life Sciences, Huaiyin Normal University, Huaian 223300, China. ¹⁵These authors contributed equally: Victor A. Streltsov and Sukanya Luang. Correspondence and requests for materials should be addressed to M.H. (email: maria.hrmova@adelaide.edu.au)

Enzymes are biological catalysts that are fundamental to life. Enzymes afford enormous accelerations to chemical reactions compared to uncatalysed reactions. It has become increasingly recognised that the largest contribution to the enzyme catalytic power arises from the electrostatic environment of polar active sites^{1,2}. Enzymes use protein architectures to precisely position a set of amino acid residues to catalyse interconversions of substrates into products. This structure-based view is supported by thermodynamic and kinetic models that treat enzymes in complex with substrates and transition states in succession³. Only rarely these models consider substrate and product diffusion as a part of their catalytic mechanisms, as these processes proceed rapidly, although on occasions products are seen entrapped in active sites⁴. Yet, the lack of descriptions of substrate associative and product dissociative pathways creates a knowledge gap⁵.

A vital conclusion drawn from the structural studies with the native β -D-glucan glucohydrolase, isoenzyme ExoI (HvExoI) isolated from barley seedlings, was that the glucose (Glc) product released from β -D-glucoside substrates remains entrapped in the enzyme active site until an incoming substrate binds^{6–9}, presumably lowering the energy barrier to facilitate Glc displacement. At this stage, the mechanism of Glc displacement and how it is linked to the catalytic cycle of HvExoI remained unanswered. Notably, no other native GH3 structures with entrapped Glc are available, because these recombinant enzymes not exposed to presumably high enough Glc concentration, do not entrap Glc during protein maturation and secretion. However, several GH3 enzyme complexes with in crystallo-perfused Glc^{10,11} or Glc-derivatives^{12,13} are available. Additionally, the native GH78 α -L-rhamnosidase with a deep pocket-shaped active site also holds the entrapped Glc molecule, which was not perfused in crystals¹⁴.

We have previously observed that mechanism-based inhibitors conduritol B epoxide and 2',4'-dinitrophenyl 2-deoxy-2-fluoro- β -D-glucoside dislodge Glc from the -1 subsite and form stable cyclitol esters via an α -anomeric linkage with Asp285⁷. Similarly, transition-state ion-like gluco-phenylimidazole mimics displace Glc and bind in the -1 subsite with their Glc component distorted in the ⁴E envelope conformation^{15,16}. Non-hydrolysable S-glycoside analogues also remove Glc from the -1 subsite and bind across the active site^{7,17}. Because HvExoI hydrolyses various positional isomers of β -D-glucosides, we rationalised this observation by the non-reducing Glc moiety being locked into a fixed position via hydrogen bonds (H-bonds) with the residues in the -1 subsite, whereas the reducing-end moiety is free to adopt multiple orientations in the $+1$ subsite. This allows the remainder of the saccharide substrates beyond the $+1$ subsite to project unencumbered^{7,9,17}.

The fundamental function of hydrolysis of oligo- and polysaccharides is essential for the understanding of the global carbon cycle, which forms the basis of a multibillion-dollar biotechnological industry. In this context, the HvExoI enzyme⁶ that belongs to the glycoside hydrolase (GH) family 3¹⁸, serves as one of the archetypal models for hydrolysis of oligo- and polysaccharides. Structural information of GH3 enzymes is available through 36 unique entries in the Protein Data Bank (PDB) from all phyla, although there is only one plant structure available (that of HvExoI), notably in complex with a variety of oligosaccharide substrate analogues and mimics^{7–9}. HvExoI folds in an $(\alpha/\beta)_8$ barrel (domain 1) and an $(\alpha/\beta)_6$ sandwich (domain 2) that at their interface cradle a 13 Å-deep pocket- or crater-shaped active site. This site accommodates the catalytic nucleophile Asp285 and catalytic acid/base Glu491 in the -1 subsite, while the aromatic clamp of Trp286 and Trp434 indole moieties constitutes the neighbouring $+1$ subsite^{6–9}. HvExoI releases a single Glc from

non-reducing termini of oligo- and polymeric β -D-glucosides to support seed germination and plant development⁹.

In this work, we examine product and substrate pathways along the catalytic cycle of a plant exo-hydrolase HvExoI, and how this enzyme utilises its plasticity and that of glycoside substrates. We use the high-resolution X-ray crystallography supported with enzyme kinetics, mass spectrometry, nuclear magnetic resonance (NMR) spectroscopy, multi-scale molecular modelling employing docking¹⁹, Molecular dynamics (MD) simulations, Genetic Algorithms with Unrestricted Descriptors for Intuitive Molecular Modelling (GaudiMM)²⁰ and Protein Energy Landscape Exploration (PELE)²¹ calculations, to reveal the Glc product displacement route, and to extract how each hydrolytic event including Glc release is precisely coordinated with the incoming substrate association and hydrolysis. This leads us to generate a variant enzyme and probe it for Glc entrapment and the S-glycoside analogue binding. Along this journey with HvExoI, we observe a remarkable phenomenon, coin the term 'substrate-product assisted processive catalysis', and describe how product and substrate trajectories are coordinated via this mechanism during catalysis. We discover that through chemical signalling, the incoming substrate evokes the formation of an autonomous and transient lateral cavity that serves as a conduit for the Glc product displacement from the active site. Crucially, these co-operative reactant pathways, where polymeric substrates never dissociate from the enzyme, sustain consecutive relocations of substrates and products via dedicated routes, ensuring that this exo-hydrolase adopts processive catalysis.

Results

GC/MS proves that Glc is bound to crystalline native HvExoI.

We purified²², crystallised²³ and chemically analysed crystals to reconcile our structural observations with Glc entrapment in native HvExoI purified from barley seedlings^{6–9}. Several crystals removed from the mother liquor were extracted and fractionated by normal-phase high performance liquid chromatography (HPLC) with evaporative light scattering detection (Fig. 1a; top panel, solid line). The mother liquor from these crystals was also analysed, both without and augmented with authentic Glc (Fig. 1a; top panel, dashed and dotted lines, respectively). The materials from these preparations eluting between 6.0 and 7.3 min during HPLC separation were pooled, reduced and peracetylated. The total ion chromatograms of the material extracted from crystals and from the mother liquor augmented with Glc showed a peak with the retention time of 23.26 min, corresponding to glucitol hexa-acetate (Fig. 1b; 1st and 3rd middle panels). Mass fragmentation spectra of these materials (Fig. 1c; bottom panels) showed identical profiles of total ions, demonstrating that the chemical substance eluting at 23.26 min was Glc. We could not detect Glc in the mother liquor, from which crystals had been removed (Fig. 1b; middle panel), while authentic Glc added to the same mother liquor was readily identifiable.

Crystal structures reveal the Glc product in native HvExoI.

The key observation that led us to explore the precise spatial disposition of the Glc product in the active site and the Glc displacement route in HvExoI, was based on our original observation that the Glc product remains entrapped in the enzyme's active site that is isolated from the native source of young seedlings^{6,8}. We anticipated that the last remaining Glc molecule originating from oligo- and polysaccharide substrates, stays associated with the enzyme in plant tissues. To this end, previous crystal structures of native HvExoI at 2.20 Å and 2.70 Å (in-house X-ray source data) resolution showed a single Glc molecule bound in the -1 subsite^{6,8}. In the current work we refined native

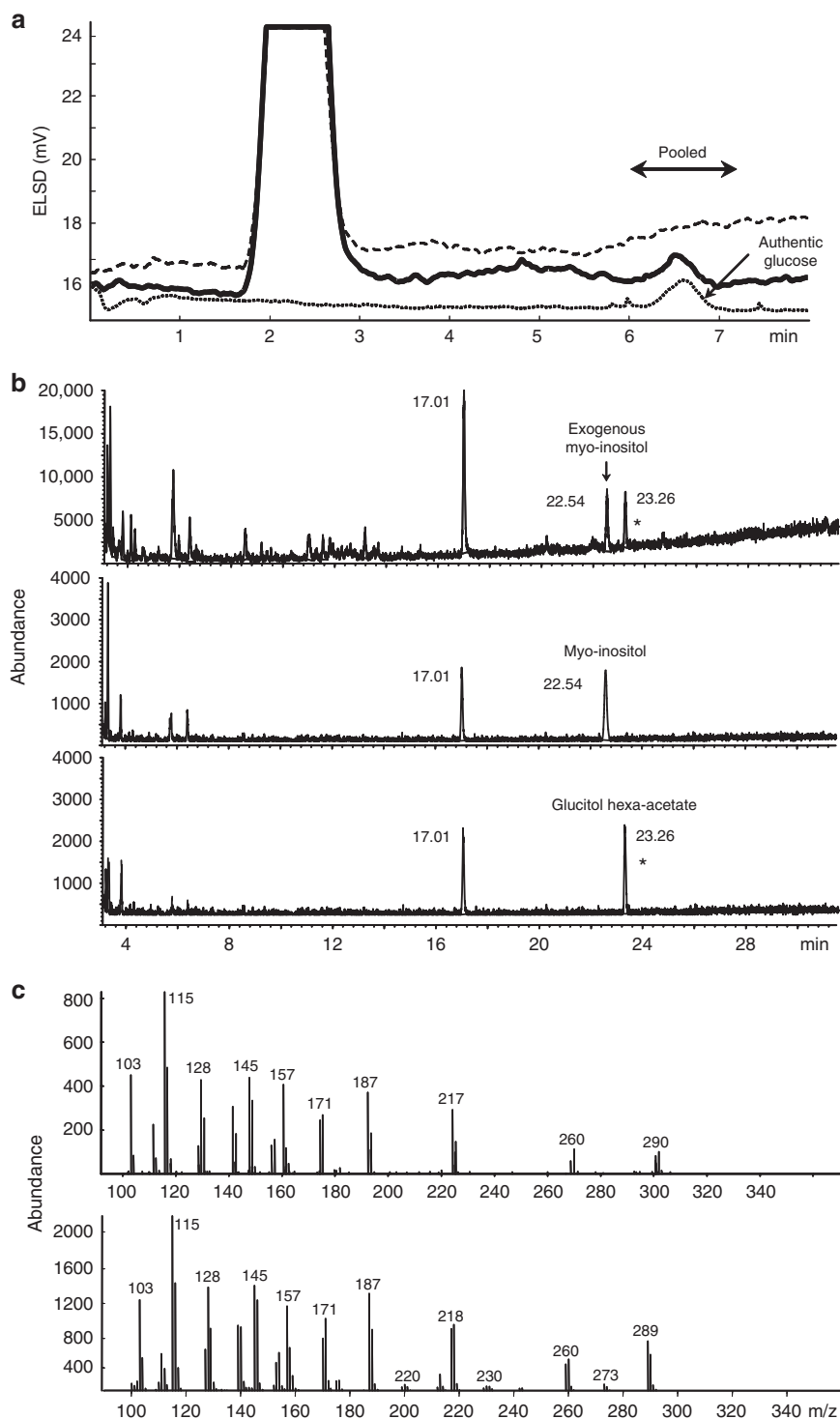


Fig. 1 Glc trapped in native HvExol is bound non-covalently in the 4C_1 conformation. **a** HPLC chromatogram of material extracted from crystals (solid line), mother liquor (dashed) and mother liquor augmented with authentic Glc (dotted). Arrow points to the peak of authentic Glc in augmented mother liquor. Fractions (left right arrow) were pooled for GC/MS analysis. **b** Total ion chromatograms obtained by GC/MS of HPLC-eluted materials (6–7.3 min) containing alditol acetates. Materials from crystals (top), mother liquor (middle) and mother liquor augmented with Glc (bottom). Myo-inositol (20 ng) served as an internal standard (top and middle panels). Numbers near peaks indicate retention times. Substance eluted in 17.01 min peak corresponds to a contaminating plasticiser. Material extracted from crystals and eluting at 23.26 min (top panel) is identical to that of glucitol hexa-acetate at 23.26 min (bottom panel) on a Prevail Carbohydrate ES column (peaks indicated by asterisks). **c** Fragmentation mass spectra of glucitol hexa-acetate: top, material extracted from crystals; bottom glucitol hexa-acetate eluting at 23.26 min

HvExoI at 1.65 Å resolution (synchrotron data), revealing that Glc is bound at 0.5 occupancy at each −1 and +1 subsites in the 4C_1 chair conformation, suggesting that it may be mobile between both subsites (Fig. 2a; Supplementary Fig. 1a).

Glc is absent in recombinant HvExoI but could be perfused in.

Although Glc has always been detected in the native enzyme isolated from barley seedlings^{6,8}, it was uncertain if it could be observed in recombinant HvExoI produced in *Pichia pastoris*. The 1.45 Å structure of the recombinant enzyme, which is kinetically and structurally nearly identical to the native one^{24,25}, failed to show the presence of Glc in the active site, where we identified glycerol and up to seven water molecules mimicking the positions of hydroxyl groups of Glc bound in native HvExoI (Fig. 3a; Supplementary Fig. 2a). However, after the crystals of recombinant HvExoI were perfused with Glc at a near saturating concentration, the 1.55 Å structure revealed two Glc molecules one each in the −1 and +1 subsites, adopting alternate 4C_1 chair (occupancy 0.8) and 1S_3 skew boat (occupancy 0.2) conformations with classical Cremer–Pople ring-puckering parameters²⁶. In the −1 subsite, we detected the network of 12–13 mono- and co-operative bi- and tri-dentate H-bonds^{27,28}, of interest were those forming short H-bonds of 2.5–2.7 Å between Oδ1 and Oδ2 of Asp95, and the C6-OH and C4-OH groups of the Glc moiety. We attempted to position into the electron density other skew boat conformers of Glc (3S_5 , 1S_5) with various occupancies, however, convergence in the refinement was reached with those of 4C_1 and 1S_3 . We concluded that during maturation in *P. pastoris*, Glc cannot be entrapped and thus observed, because the intracellular concentration in this host is not high enough during enzyme maturation or secretion. To demonstrate that Glc was not displaced from the active site by the cryoprotectant, data were collected from recombinant crystals at ambient temperatures without applying vitrification solutions. Here, no Glc was observed in the active site, but was readily observable after perfusion of crystals with Glc.

SPR analysis of Glc and substrate binding to HvExoI. To determine the strength of Glc binding and to compare it with that of the substrate thio-analogue methyl 6-thio-β-gentiobioside (G6SG-OMe), we used the recombinant HvExoI enzyme. The steady-state affinity²⁹ K_D value of 0.008×10^{-3} M for G6SG-OMe revealed that it bound tighter than the Glc hydrolytic product (K_D of 0.16×10^{-3} M) (Table 1; Supplementary Fig. 3).

NMR spectroscopy of Glc binding to recombinant HvExoI. To identify conformational states of bound Glc with the measured K_D value of 0.16×10^{-3} M (Table 1; Supplementary Fig. 3), we used 1H saturation transferred difference (STD)³⁰ and transferred nuclear Overhauser effect spectroscopy (trNOESY)³¹ in solution NMR. STD spectra recorded with Glc in a 60-fold excess relative to HvExoI detected Glc bound non-covalently. As only STDs for protons of β-anomeric Glc (and not for α-form) were observed, while one would expect an α:β anomeric ratio of ~2:3 in solution equilibrium, we concluded that the enzyme specifically binds β-anomeric Glc (Fig. 4a); this observation supports the retaining hydrolytic mechanism of HvExoI²². To identify the conformation of Glc, trNOESY experiments under a relatively low HvExoI versus Glc ratio revealed weak trNOEs between H1-H3 and H1-H5 protons of β-anomeric Glc (Fig. 4b), consistent with β-anomeric Glc bound in the low-energy 4C_1 chair conformation. To corroborate these findings, thiocellobiose which cannot be hydrolysed by HvExoI⁷, was added in excess relative to HvExoI. Here, clear negative NOEs between specific thiocellobiose proton pairs were observed (Fig. 4c), while strong trNOEs of the

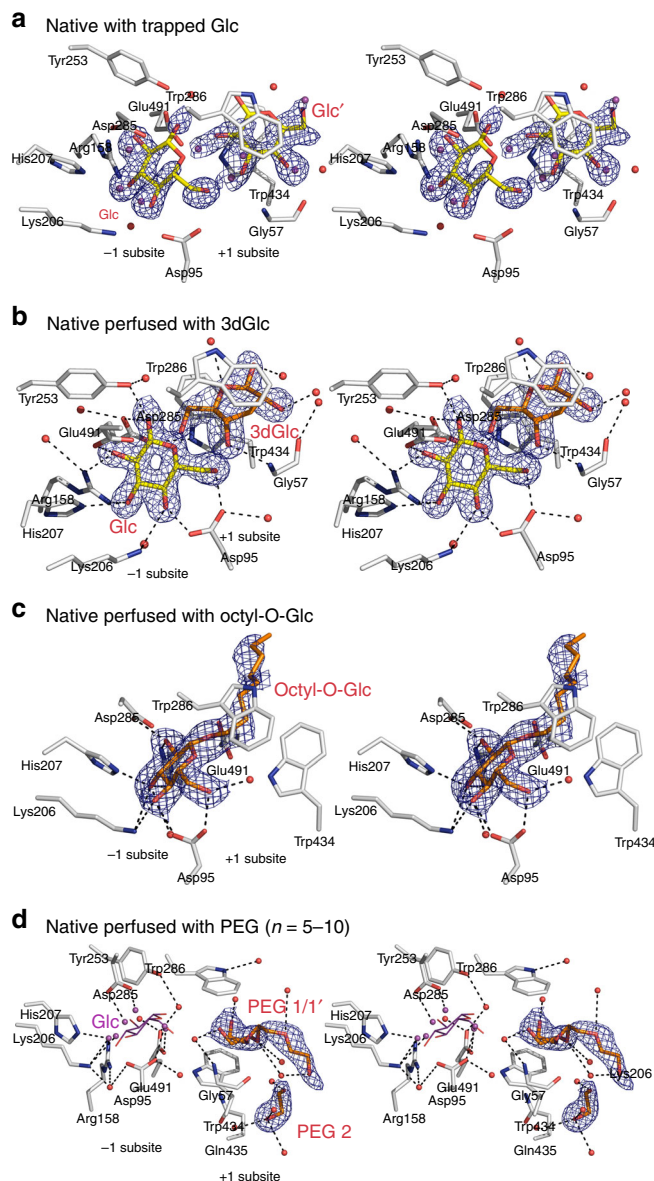


Fig. 2 Native HvExoI with trapped Glc, and with perfused 3dGlc, octyl-O-Glc and PEG. **a** Stereo view of native HvExoI with Glc (Glc and Glc', carbons: yellow sticks) at 0.5 occupancy, which oscillates between the −1 and +1 subsites. **b** Stereo view of native HvExoI with Glc (carbons: yellow sticks) and 3dGlc (carbons: orange sticks) at 1.0 occupancies, bound in the −1 and +1 subsites, respectively. **c** Stereo view of native HvExoI with octyl-O-Glc (carbons: orange sticks) at 1.0 occupancy, bound across the −1 and +1 subsites. **d** Stereo view of native HvExoI with two PEG ($n = 5-10$) molecules (carbons: orange sticks; PEG 1 in two alternate conformations at occupancies 0.5 each and PEG 2 at occupancy 1.0), bound in the +1 and putative +2 subsites. Grey, red, and blue represent carbon, oxygen and nitrogen atoms, respectively. Water molecules are shown as red or magenta (bound alternate water molecules when the ligand is missing and are not numbered) spheres in complexes with Glc and PEG. In the PEG complex magenta-coloured water molecules mimic positions of OH groups of Glc (added for illustration in magenta lines). Separations of less than 3.50 Å from active site residues (carbons: grey sticks) are shown as dashed lines. Derived $|2mF_{obs} - DF_{calc}|$ electron density maps are contoured at 1σ (blue mesh). In $|2mF_{obs} - DF_{calc}|$, F_{obs} and F_{calc} are observed and calculated X-ray structure factor amplitudes, where m is the figure of merit and D is the estimated coordinate error

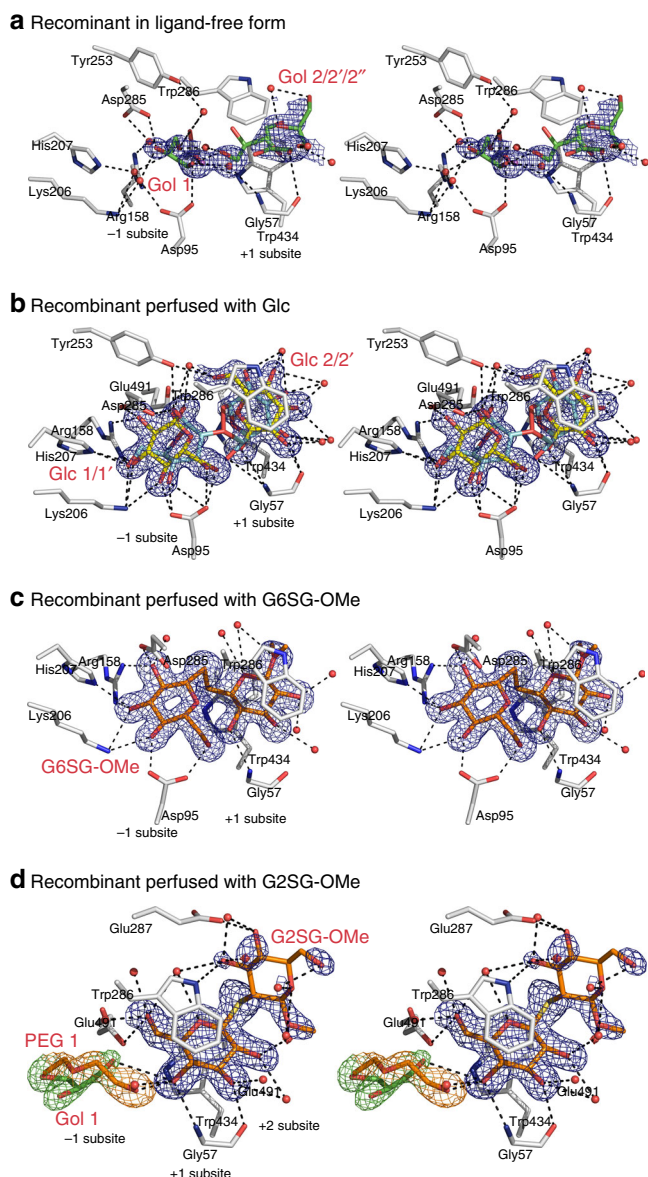


Fig. 3 Recombinant HvExoI and with perfused Glc, G6SG-OMe, or G2SG-OMe. **a** Stereo view of recombinant HvExoI in a ligand-free form. Two glycerol molecules (carbons: green sticks) (Gol 1 at occupancy 0.5; Gol 2 in three alternate conformations) are bound in the -1 and +1 subsites. **b** Stereo view of recombinant HvExoI with two Glc molecules in the 4C_1 conformation (carbons: yellow sticks) at 0.8 occupancy and in the 1S_3 conformation (carbons: cyan sticks) at 0.2 occupancy, bound in the -1 and +1 subsites. **c** Stereo view of recombinant HvExoI with G6SG-OMe (carbons: orange sticks) at 1.0 occupancy, bound across the -1 and +1. **d** Stereo view of recombinant HvExoI with G2SG-OMe (carbons: orange sticks) at 0.7 occupancy, bound across the +1 and putative +2 subsites. Water molecules are shown as red spheres. Separations of less than 3.50 Å from the active site residues (carbons: grey sticks) are shown as dashed lines. Derived $|2mF_{obs} - DF_{calc}|$ electron density maps are contoured at 1σ for G2SG-OMe and Gol (blue mesh) in (a-c). In (d), electron density maps are contoured at 1σ for PEG (orange mesh) and Gol (green mesh) in the -1 subsite

anomeric proton of non-reducing Glc (H-1') with intra-residual H-3' and H-5' (Supplementary Fig. 4; Supplementary Note 1) implied that the non-reducing β -D-glucopyranose ring of thio-cellobiose was bound to HvExoI in the 4C_1 conformation³².

Table 1 Binding (K_D) and inhibition (K_i) parameters for HvExoI by ligands and substrate analogues

Ligand/ inhibitor	K_D^a/K_i^b (M)	ΔG^c (kJ \times mol $^{-1}$)	Chemical formula ^d
Glc ^a	0.16×10^{-3}	-21.7	$C_6H_{12}O_6$
G6SG-OMe ^a	0.008×10^{-3}	-29.1	$C_{13}H_{24}O_{10}S$
3dGlc ^b	9.8×10^{-3}	-11.7	$C_6H_{12}O_5$
4dGlc ^b	9.4×10^{-3}	-11.8	$C_6H_{12}O_5$
Octyl-O-Glc ^b	0.13×10^{-3}	-22.6	$C_{14}H_{28}O_6$
Octyl-S-Glc ^b	1.1×10^{-3}	-17.2	$C_{14}H_{28}O_5S$
G2SG-OMe ^b	2.55×10^{-3}	-15.1	$C_{13}H_{24}O_{10}S$

^aUsing Surface Plasmon Resonance with recombinant HvExoI

^bUsing inhibition kinetics with the 4-nitrophenyl β -D-glucopyranoside substrate⁸ with native HvExoI

^cCalculated according to $\Delta G = -RT \ln [K_i]$ or $\Delta G = -RT \ln [K_D]$ ²⁹

^dChemical structures are shown in Supplementary Table 1

Conformational Free Energy Landscape of Glc bound to HvExoI. To further assess conformational states of bound Glc in the high-resolution structure of recombinant HvExoI in complex with Glc, quantum mechanics/molecular mechanics (QM/MM) metadynamics simulations were performed to assess Conformational Free Energy Landscape (FEL) maps of β -D-Glc bound in the active site of HvExoI (Fig. 5; Supplementary Fig. 2b). These analyses indicated that although Glc was engaged in extensive H-bond networks, the β -D-glucopyranose ring adopted 4C_1 chair and $^1S_3/B_{3,0}$ (-1 subsite) or 4C_1 chair and $B_{3,0}$ (+1 subsite) distorted skew boat/boat conformations (Fig. 5; left and middle panels). More precisely in the -1 subsite, H-bonds with Asp285 and Arg158 restrained the 2-OH group in the equatorial position, and stabilised the 4C_1 conformation by 8 kcal/mol more than that of 1S_3 or $B_{3,0}$, with a conformational energy barrier change of 11 kcal/mol. Conversely, hydroxyl groups of Glc in the +1 subsite interacted with Asp95, Arg158, Lys206, His207, Asp285, and Glu491, stabilising the 4C_1 Glc conformation by 2 kcal/mol more than that of $B_{3,0}$, with a conformational energy barrier change of 6 kcal/mol. Notably, these findings concurred with conformational FEL of isolated Glc³³, in which 4C_1 and $^1S_3/B_{3,0}$ conformers were most stable (Fig. 5; right panel). Nonetheless, FEL of Glc bound in the -1 and +1 subsites exhibited exclusively these two minima due to a restricted conformational space within the enzyme active site (Figs. 2a, 3b, 5) as demonstrated in other glycosyl hydrolases³⁴.

Inhibition kinetics of substrate analogues in native HvExoI. To shed light on the molecular basis of the Glc displacement route, we selected a series of deoxy-Glc (dGlc) and alkyl-glucoside derivatives, and substrate mimics that could potentially displace Glc⁶. To assess the strength of inhibition, we determined dissociation constants of enzyme-inhibitor complexes (K_i) for 3-deoxy-glucose (3dGlc), 4-deoxy-glucose (4dGlc), n-octyl β -D-glucopyranoside (octyl-O-Glc), n-octyl 1-thio- β -D-glucopyranoside (octyl-S-Glc), a hydrophilic polymer polyethylene glycol (PEG; putative fractional polysaccharide mimic³⁵) and the thio-analogue methyl 2-thio- β -sophorose (G2SG-OMe). All inhibitors showed competitive inhibition with K_i values between 9.8×10^{-3} M and 0.13×10^{-3} M (Table 1). The most effective was octyl-O-Glc with K_i of 0.13×10^{-3} M, while surprisingly G2SG-OMe was weak with K_i of 2.55×10^{-3} M, while PEG was non-inhibitory, as expected.

Glc is not removed from native HvExoI by dGlc derivatives. To test the hypothetical Glc displacement route by using dGlc derivatives, crystals were perfused with saturating concentrations of

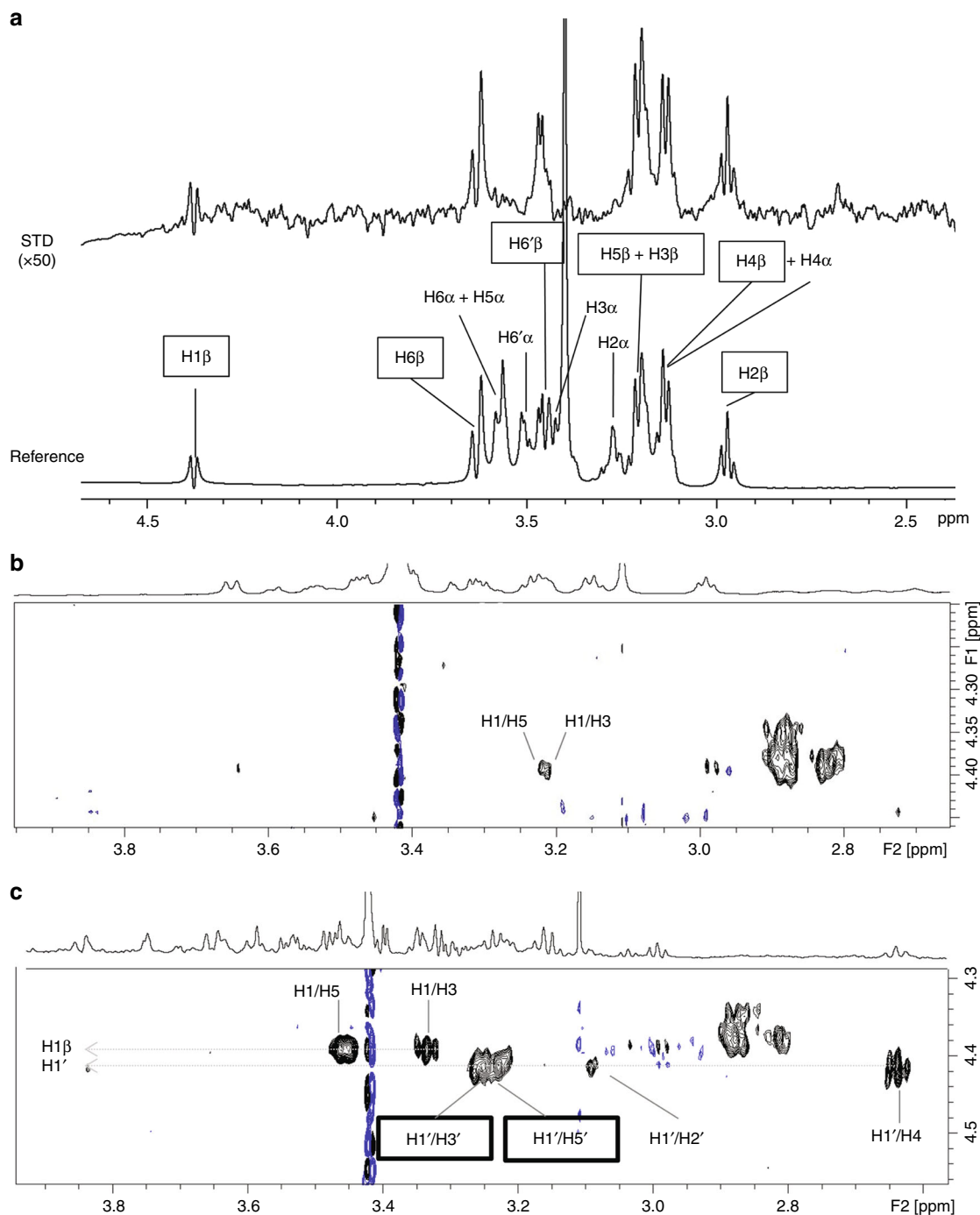


Fig. 4 Recombinant HvExoI recognises β -D-Glc in the 4C_1 conformation. **a** 1H STD NMR spectrum (top) and reference spectrum (bottom), acquired with recombinant HvExoI and Glc ($40\ \mu M$ and $2.4\ mM$, respectively) at $600\ MHz$ and $283\ K$. Only protons of β -D-Glc (squares) showed STDs. Off and on resonance frequencies were set at $100\ ppm$ and $0.65\ ppm$, respectively. The Gaussian shaped pulse ($30\ ms$) was used for selective irradiation with total saturation time of $2\ s$. **b** trNOESY spectrum of recombinant HvExoI and Glc ($57\ \mu M$ and $285\ \mu M$), acquired at $800\ MHz$, $283\ K$ and with $300\ ms$ mixing time. Only NOEs observed were weak trNOEs of Glc between $H1/H3$ and $H1/H5$. **c** trNOESY spectrum of recombinant HvExoI with Glc and thiocellobiose ($57\ \mu M$, $285\ \mu M$, and $171\ \mu M$, respectively), acquired at $800\ MHz$, $283\ K$ and with $300\ ms$ mixing time. All NOEs were negative. trNOEs defining the 4C_1 conformation of non-reducing Glc of bound thiocellobiose are in boxes. Blue lines in **(b)** and **(c)** refer to residual noise signals

3dGlc (Fig. 2b; Supplementary Fig. 1b) and 4dGlc. The structures of both complexes showed that dGlc derivatives could not displace Glc from the -1 subsite, but that Glc consolidated to the -1 subsite. The well-defined electron density map indicated that 3dGlc was orientated parallel to Trp286 and Trp434 in the $+1$ subsite, through π -system stacking contacts, forming H-bonds

via C6-OH with Ne1 of Trp434 and Oe2 of Glu491, and via C4-OH with the N atom of Gly57 (Fig. 2b; Supplementary Fig. 1b). Notably, as 3dGlc rotated to a different position relative to that of Glc in native HvExoI (such that the C4-OH group of 3dGlc overlapped the position of C3-OH of Glc in the native structure), it could establish H-bonds with Gly57 and Glc in the -1 subsite.

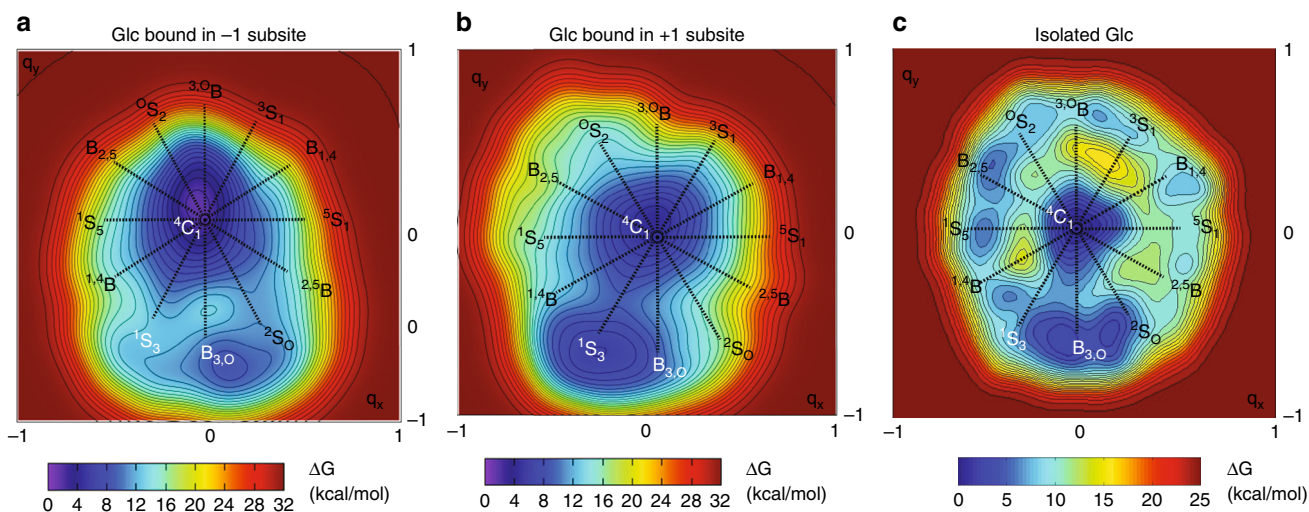


Fig. 5 Conformational FEL maps of β -D-Glc bound in the active site of HvExoI. Conformational FEL of β -D-Glc in the -1 (a) and $+1$ (b) subsites are based on ab initio QM/MM metadynamics. c Conformational FEL map of isolated Glc³⁴ (reproduced with permissions of the Journal of American Chemical Society; all rights reserved; <https://pubs.acs.org/doi/10.1021/jacs.5b01156>). Energy level lines are separated by 1 kcal/mol

Alkyl-glucosides and PEG remove Glc from the active site.

Contrary to what we observed with dGlc derivatives, *n*-octyl β -D-glucosides with tighter K_i values (Table 1) replaced Glc and bound across the active site (Fig. 2c; Supplementary Fig. 1c). While aliphatic chains of alkyl β -D-glucosides were threaded through the Trp286 and Trp434 aromatic clamp and did not interact with a surrounding environment, Glc moieties established a network of 10–11 mono- and co-operative bi- and tridentate H-bonds, like those observed in native (with trapped Glc) or recombinant Glc-perfused HvExoI (Figs. 2, 3). Similarly, after applying PEG at saturating concentrations to native HvExoI crystals, the 1.80 Å PEG-perfused structure revealed that Glc was replaced by five water molecules occupying the positions of the C1-OH to C6-OH groups of Glc in the -1 subsite (Fig. 2d; Supplementary Fig. 1d). The PEG molecule bound between Trp286 and Trp434 in two alternate conformations made a water-mediated H-bond with Gly57, which was reminiscent of the bond formed between C4-OH of 3dGlc and Gly57. The second PEG molecule bound near Trp434 did not contact the enzyme, but instead interacted with the first PEG via water-mediated H-bonds (Fig. 2d; Supplementary Fig. 1d).

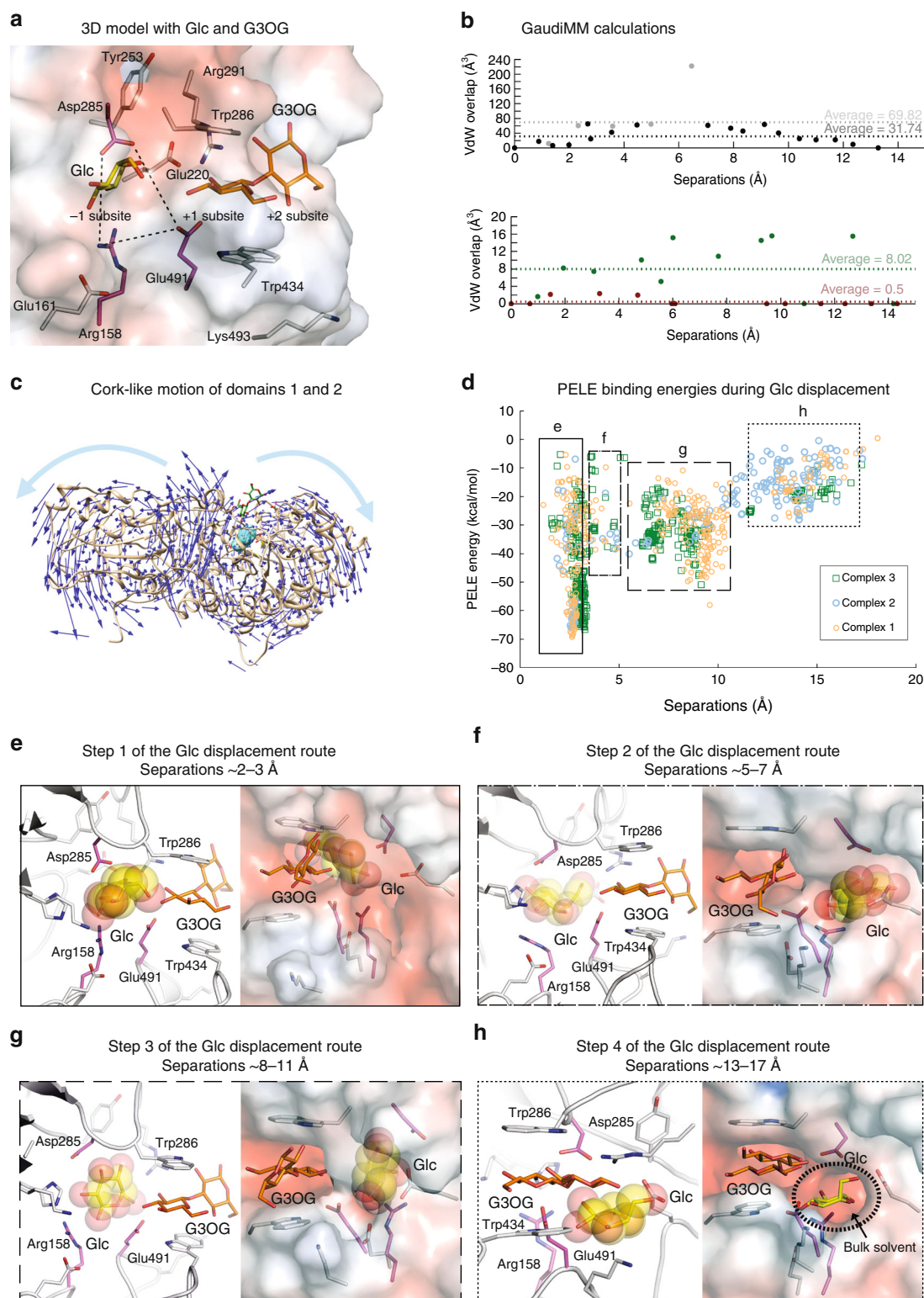
G6SG-OMe and G2SG-OMe bind to HvExoI in two distinct modes.

To understand how (1,6)- and (1,2)-linked substrate thioanalogues displace the Glc product from the active site, we perfused recombinant crystals of HvExoI and observed that the sugars adopted different poses (Fig. 3c, d; Supplementary Fig. 2c, d). It is worth noting that G6SG-OMe binds to the enzyme with K_D of 0.008×10^{-3} M, whereas G2SG-OMe binds about 320-fold weaker (K_i of 2.55×10^{-3} M) (Table 1). We observed that in the 1.57 Å structure of HvExoI with G6SG-OMe, the ligand formed well-defined electron densities for both Glc moieties across the active site (Fig. 3c; Supplementary Fig. 2c), and that the 1.68 Å structure with G2SG-OMe (refined at 0.7 occupancy at both subsites) had a less-defined density for the ligand, which bound to the enzyme in a markedly different position (Fig. 3d; Supplementary Fig. 2d). For G6SG-OMe, we observed short H-bonds in the -1 subsite, which resembled those observed in the native (Fig. 2a; trapped Glc) or recombinant (Fig. 3a; perfused with Glc) enzymes. In contrast, G2SG-OMe could not slide past the $+1$ subsite in the pocket, likely due to a thio-glycosidic bond

rigidity, and hence it projected with the reducing-end Glc moiety into the solvent at the putative $+2$ subsite, while the non-reducing-end was held in the adjacent $+1$ subsite. For the -1 subsite, we assigned glycerol and PEG molecules that formed H-bonds with the G2SG-OMe non-reducing end that also established the H-bond with Gly57; the reducing moiety of G2SG-OMe interacted via C4-OH with O ϵ 2 of Glu287 and via C3-OH with N ϵ 1 of Trp286 (Fig. 3d; Supplementary Fig. 2d). In the G6SG-OMe structure we noted significant elevations of B-factor values in two loops (Thr214-Glu228 and Glu491-Asn498) near the catalytic site entry (Supplementary Note 2); these B-factor elevations were not seen in the G2SG-OMe structure.

HvExoI-G2SG-OMe complex is an intermediate for Glc exit path. We considered the HvExoI in complex with G2SG-OMe (PDB 6MD6) (Fig. 3d; Supplementary Fig. 2d) to embody the attributes of an intermediate enzyme-substrate complex, and the disposition of G2SG-OMe that of an incoming substrate. Hence, we used this complex for investigations of the hypothetical Glc displacement route via multi-scale molecular modelling employing docking¹⁹ and MD simulations, followed by GaudiMM²⁰ and PELE²¹ pathway calculations.

Respective reciprocal docking of the Glc product and the β -D-glucopyranosyl-(1,2)-D-glucose (G2OG) or β -D-glucopyranosyl-(1,3)-D-glucose (G3OG) substrates into HvExoI:G2OG-OMe or HvExoI:Glc complexes, combined with MD simulations indicated that the existence of ternary HvExoI:Glc:G2OG or HvExoI:Glc:G3OG complexes (designated 1–6) was plausible, where Glc was bound in the -1 subsite, and G2OG or G3OG were attached at $+1$ and putative $+2$ subsites (Fig. 6a; Supplementary Figs. 5, 6, 8, Supplementary Notes 3–4, Supplementary Data 1). This suggested that if the incoming substrate binds at the $+1$ to $+2$ subsites, while Glc is still trapped in the -1 subsite, an alternative Glc exit path, other than through the $+1$ subsite needs to be considered. We also searched for substrate binding modes different to those at the $+1$ to $+2$ subsites, using docking of the G2OG or G3OG substrates in structures based on MD simulations of the native HvExoI:Glc complex (Supplementary Figs. 5–8; Supplementary Note 4; Supplementary Data 1–2). However, no stable binding sites were found other than those at $+1$ and $+2$ subsites. In addition, docking and MD simulations



unequivocally revealed that the HvExoI:Glc:G2OG and HvExoI:Glc:G3OG complexes (Fig. 6a) converged and were stable, and could be used for investigations of Glc displacement. MD simulations also revealed that stable binding of substrates required Trp434 adopting the parallel orientation to that of Trp286, where C-H/ π interactions mediated binding. Notably, binding of the G2OG or G3OG substrates triggered the conformational change of Tyr253, modifying the buried lateral

cavity adjacent to the –1 and +1 subsites. We observed that when only one Glc was bound at the –1 subsite, Tyr253 remained H-bonded to the carbonyl oxygen of the Trp286 backbone for the full extent of simulation. The sidechains of the catalytic nucleophile Asp285 and acid/base Glu491, together with Arg158 (Fig. 6a; dashed lines) physically separated this cavity from the –1 subsite. In complexes 1, 3 and 4, the Tyr253 side chain moved away from this lateral cavity enlarging it, while in complex 2,

Fig. 6 Displacement pathway of Glc, based on crystal structures and molecular modelling. **a** Converged HvExoI:Glc:G3OG complex obtained by docking. Selected residues (carbons: atomic and purple sticks), Glc (carbons: yellow sticks) and G3OG (carbons: orange sticks) are shown. Glc in the -1 subsite is separated from the lateral cavity by the Arg158-Glu491-Asp285 (carbons: purple sticks) toll-like barrier (dashed lines). Surface morphology coloured by electrostatic potentials: white, neutral; blue, $+5 \text{ kT e}^{-1}$; red, -5 kT e^{-1} . **b** Glc cannot exit, when crystallographic coordinates are considered by GaudiMM (grey symbols). Glc displacement only occurs when protein atoms are relocated to the lowest-energy normal mode (black). Clashes of Glc are shown as the function of separations from the Glc geometric centre in the -1 subsite. Displacement routes present fewer clashes, when MD structures are considered (green) and protein atoms move (dark red). **c** Lowest-energy normal mode shows a cork-like motion of domains 1 and 2. **d** Glc displacement routes calculated by PELE using ternary HvExoI:Glc:substrate complexes 1–3. PELE binding energies, plotted as a function of Glc separation from the -1 subsite, show that Glc traverses via the lateral cavity at the separation of 8 \AA from protein surface. Full, semi-dashed, dashed and dotted lines specify separation ranges of $2\text{--}3 \text{ \AA}$, $5\text{--}7 \text{ \AA}$, $8\text{--}11 \text{ \AA}$ and $13\text{--}17 \text{ \AA}$ (referring to line thickness in **e–h**) between respective centres of masses and C4-OH groups of Glc in initial and final structures. **e–h** Four steps along the Glc displacement route based on complexes 1–3 (Glc carbons: yellow spheres). Left and right images are rotated by 200° (**e–g**) and -50° (**h**) along y-axes. **e** Glc is in the -1 subsite. **f** Movements of Arg158, Asp285 and Glu491 allow Glc traversing into the lateral cavity. **g** Glc in the lateral cavity is partly exposed to the bulk solvent. **h** Glc is fully exposed to the bulk solvent. Black dotted ellipsoid indicates the transient solvent-exposed aperture of the lateral cavity that facilitates Glc displacement

this side chain rotated to the bottom of the cavity making it shallower (Supplementary Figs. 6, 8; Supplementary Notes 3–4).

To find out if binding of substrates by HvExoI lacking the bound Glc product would lead to tighter or weaker binding than that with Glc included, we carried out further calculations and compared binding affinities. Docking calculations of G2OG, G3OG and β -D-glucopyranosyl-(1,6)-D-glucose (G6OG) disaccharide substrates at the -1 and $+1$ subsites, predicted binding with higher affinities (Goldscore scoring function values of 66 for G2OG, 76 for G3OG and 74 for G6OG), when the Glc product was absent in the -1 subsite. However, when the Glc product was included in the -1 subsite, docking of G2OG, G3OG and G6OG in the $+1$ and putative $+2$ subsites predicted binding with lower affinities (Goldscore scoring function values of 60 for G2OG, 57 for G3OG and 61 for G6OG). These data indicated that bound Glc lowered binding energies for incoming substrates as they had no access to the higher affinity -1 subsite.

Discovery of substrate-product assisted processivity in HvExoI.

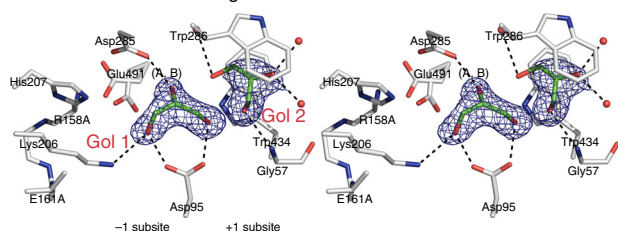
Initially, we used the GPathFinder extension of the GaudiMM platform²⁰ to reveal potential Glc exit routes from the active site into the bulk solvent, whereby we considered steric clashes. The initial set of simulations was performed with the HvExoI:Glc:G2OG ternary complex 1, based on the coordinates of HvExoI in complex with G2SG-OMe and Glc docked in the -1 subsite. No exit path was detected in this case, as Glc never overcame steric clashes to vacate the -1 subsite (Fig. 6b; top panel; Supplementary Data 3). However, further calculations with the protein backbone displaced by 2.2 \AA along the lowest-energy normal mode strikingly identified the Glc displacement path. As a bonus, these calculations also revealed a cork-like motion between domains 1 and 2, which was correlated with the open and closed states of the active site (Fig. 6c). The same ligand path was also obtained with the MD-derived HvExoI:Glc:G3OG ternary complex 3 (Fig. 6b; bottom panel), highlighting the importance of incorporating local and collective protein motions in these calculations.

To explore the Glc displacement route in detail, the ligand migration calculations of the PELE approach²¹ were used, whereby structures generated at each step along the path (by sequential ligand and protein geometric perturbations followed by energy minimisation using the OPLS-AA force field) were evaluated and accepted or rejected according to a Metropolis criterion at a given temperature. Using this approach, the HvExoI:Glc:G2OG and HvExoI:Glc:G3OG complexes 1–3 converged, and revealed that Glc egressed from the -1 subsite to the adjacent lateral cavity that was enlarged after Tyr253 rotated. This cavity was defined by an ensemble of 14 residues: Trp156, Arg158, His207, Phe208-backbone, Asp211, Asn219, Glu220,

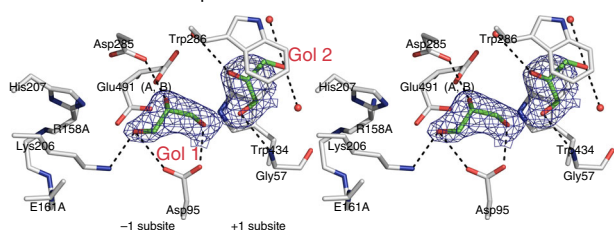
Ser252, Tyr253-backbone, Ser254, Asp285, Arg291, Glu491 and Thr492 (Fig. 6; Supplementary Fig. 9). This cavity was partially separated from the solvent by the Asn219, Glu220 and Arg291 sidechains, with the two latter residues forming a salt bridge with Oe1 and Oe2 of Glu220 to NH1 of Arg291 at separations of $3.10 \text{ \AA}/2.65 \text{ \AA}$, $2.77 \text{ \AA}/3.56 \text{ \AA}$ and $4.65 \text{ \AA}/2.84 \text{ \AA}$ for complexes 1, 2, and 3, respectively. The abundance of data points in the PELE plot at separations between $8\text{--}11 \text{ \AA}$ (Fig. 6d) indicated that there was a local energy minimum for Glc binding in this cavity (Fig. 6d; subpanel g). For complex 2, PELE calculations suggested a more transient passage of Glc through the lateral cavity, most likely due to a shallower profile resulting from the specific Tyr253 conformation, compared to other complexes. Nevertheless, in all cases, if Glc was to advance to the lateral cavity from the -1 subsite, it must traverse the space between Asp285, Glu491 and Arg158, which likely represent a toll-like barrier (Fig. 6a, dashed lines; Fig. 6e); this Glc passage corresponded to fewer data points at separations between $5\text{--}7 \text{ \AA}$ (Fig. 6d; Supplementary Data 4). With Glc bound in the -1 subsite, the shortest separations between the Arg158-Asp285, Arg158-Glu491 and Glu491-Asp285 sidechains were those at $5.10 \text{ \AA}/5.42 \text{ \AA}/5.17$, $3.09 \text{ \AA}/3.16 \text{ \AA}/4.42 \text{ \AA}$ and $5.39 \text{ \AA}/6.32 \text{ \AA}/6.65 \text{ \AA}$ in respective complexes 1, 2 and 3. Notably, conformational changes of Arg158, Asp285 and Glu491 sidechains facilitated Glc movement (Fig. 6e, f), and altered the separations specified above to $7.07 \text{ \AA}/5.33 \text{ \AA}/6.21 \text{ \AA}$, $3.00 \text{ \AA}/2.97 \text{ \AA}/6.82 \text{ \AA}$ and $6.96 \text{ \AA}/7.23 \text{ \AA}/7.71 \text{ \AA}$, when Glc bypassed the toll-like barrier. Once in the lateral cavity, Glc was free to exit from the cavity into the bulk solvent via a transient and autonomous aperture. This transient opening was formed through rotations and backbone fluctuations of Glu220, Arg291, Thr492 and Lys493, and surrounding residues in the vicinity of the bound β -D-glucoside molecule (Fig. 6h). As Glc migrated across the lateral cavity (Fig. 6e, f), it established H-bonds³⁶ with the protein residues and incoming substrates to maintain energetic favorability (Supplementary Tables 2–4). This suggested that the hydrophilic environment of the toll-like barrier and the lateral cavity may have evolved for this exit route and may be evolutionarily conserved in GH3 enzymes (Supplementary Fig. 9). Finally, and most importantly, the Glc displacement route raised the possibility that this trajectory may facilitate processive catalysis in HvExoI and other GH3 exo-hydrolases with pocket-shaped active sites.

Proof of concept using the R158A/E161A variant. Ligand migration calculations using PELE (Fig. 6) and the conservation patterns of the toll-like barrier (Supplementary Fig. 9) led us designing the double non-conservative R158A/E161A variant to critically assess the roles of these residues in binding Glc and the

a Variant R158A/E161A in ligand-free form



b Variant R158A/E161A perfused with Glc



C Variant R158A/E161A perfused with G6SG-OMe

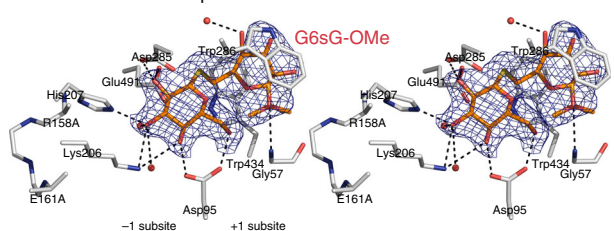


Fig. 7 R158A/E161A of HvExol and with perfused Glc or G6SG-OMe.

a Stereo view of a ligand-free form of R158A/E161A. Two glycerol molecules (carbons: green sticks) at 1.0 occupancies are bound in the -1 and $+1$ subsites. **b** Stereo view of R158A/E161A with perfused Glc, which failed to bind, instead two glycerol molecules (carbons: green sticks) are bound at 1.0 occupancies in the -1 and $+1$ subsites. **c** Stereo view of R158A/E161A with G6SG-OMe (carbons: orange sticks) at 1.0 occupancy, bound across the -1 and $+1$ subsites. Separations of less than 3.30 Å from electronegative atoms of active site residues (carbons: grey) and water oxygens, are shown as dashed lines. Derived $|2mF_{\text{obs}} - DF_{\text{calc}}|$ electron density maps are contoured at 1σ (blue mesh)

G6SG-OME substrate thio-analogue (Fig. 7; Supplementary Fig. 10). The 1.65 Å and 2.21 Å structures without and with perfused Glc showed two glycerol molecules, one each in the −1 and +1 subsites (Fig. 7a, b; Supplementary Fig. 10a, b), and the absence of Glc despite perfusing crystals at saturating concentrations. On the other hand, the 2.30 Å structure of R158A/E161A perfused with G6SG-OME strikingly revealed this sugar bound across the active site, but in a different pose compared to that of the wild-type (WT) enzyme (cf. Figures 3c, 7c; Supplementary Figs. 2c and 10c). In the variant structure the position of the Glc moiety in the −1 subsite matched that of WT, but the disposition of the reducing-end Glc moiety was flipped in its position between the indole moieties of Trp286 and Trp434, such that intra-ring oxygen of the β-D-glucopyranose moiety pointed to Trp434. Obviously, this G6SG-OME pose was due to Ala replacements for Arg158 and Glu161.

Discussion

Substrates and products in enzyme active sites bind and unbind at fast rates that challenge investigations of their trajectories, leading to the lack of a deep understanding of these hallmarks of enzyme catalysis^{37–39}. Here, we capitalise on the Glc product entrapment

observation in native HvExoI isolated from seedlings⁶, which renders it an archetype model to examine product/substrate dissociations/associations in a pocket-shaped active site. Most approaches towards the descriptions of reactant movements are based on MD simulations of substrate and product binding and unbinding, including random collisions or diffusion⁴⁰. Other studies reveal substrate/product migration pathways that accompany conformational changes of protein backbones or secondary structures^{41,42}.

When we first described the HvExoI structure⁶⁻⁹, we were unable to explain why the trapped Glc product remained bound in the active site and has not diffused away, and what was the implication of this unassuming observation. The obvious explanation for not seeing naturally bound products or co-factors in structures of other GH enzymes is that these proteins are generated in recombinant hosts, where intracellular concentrations of potential enzyme reactants are not high enough during protein maturation. To this end, solving crystal structures of enzymes purified from native sources offers an additional information that could be beneficial for the understanding of catalytic cycles.

When we first detected the Glc product entrapment in native HvExoI, we hypothesised that the role of Glc may be linked to the pre-organised state of the active site to maintain efficient catalysis required for the growth of a plant embryo⁹. One way of preserving this pre-organised state would be to keep entropic costs of catalysed reactions low, where an entrapped product from a previous hydrolytic cycle could lower overall entropic demands for binding of incoming substrates^{1,43}. The Glc product retention in the active site may also disfavour binding of incorrect substrates through product dissociation rates that would govern the selectivity of substrate binding⁴⁴. Others may argue that although most enzymes have evolved to use conformational adjustments to favour tight binding of correct substrates and product release, some enzymes may have not acquired this asset. This assumption was confirmed by perfusing dGlc derivatives that akin of reaction products could not remove Glc, but substrate analogues and mimics could.

As mentioned above, despite the plethora of structural data for HvExoI^{6-9,15-17}, we could not explain, how Glc egresses from the 13 Å-deep pocket-shaped active site, although we detected cross-talk between residues entrapping Glc in the -1 subsite, and those in the +1 subsite delineated by Trp286 and Trp434. Function of Trp residues in retaining Glc or substrates is unsurprising, as these residues have extended heterocyclic indole ring systems with amphipathic characteristics, high de-localised electron densities and permanent dipole moments due the intra-ring nitrogens, that form H-bonds²⁹. In all 22 structures of HvExoI we so far resolved (Supplementary Table 5), Trp286 and Trp434 at the +1 subsite showed no structural heterogeneity, and at separations of about 4 Å from ligands were compliant with the provision of two sets of C-H/ π interactions⁴⁵ depending on the stereochemistry of bound ligands^{7,17}. Crystalline HvExoI was also perfused with the (3)- β -D-Glc-S-(1,3)- β -D-Glc-(1,)₁₄₋₂₂ polymeric substrate⁴⁶, but we did not observe potential binding sites beyond the -1 and +1 subsites.

The time-space averaged vision of crystal structures presented here was supported by the multi-scale molecular modelling, based on docking¹⁹, MD simulations, QM/MM, GaudiMM²⁰ and PELE²¹, that provided the first clear view of succession of events during Glc displacement. This approach also hinted that trapped Glc may play a surprising role in catalysis and led to the discovery of processivity by this exo-hydrolase. In the light of this discovery we coin the term ‘substrate-product assisted processive catalysis’, due to the key role that a substrate and a product play in the evocation of the catalytic pathway.

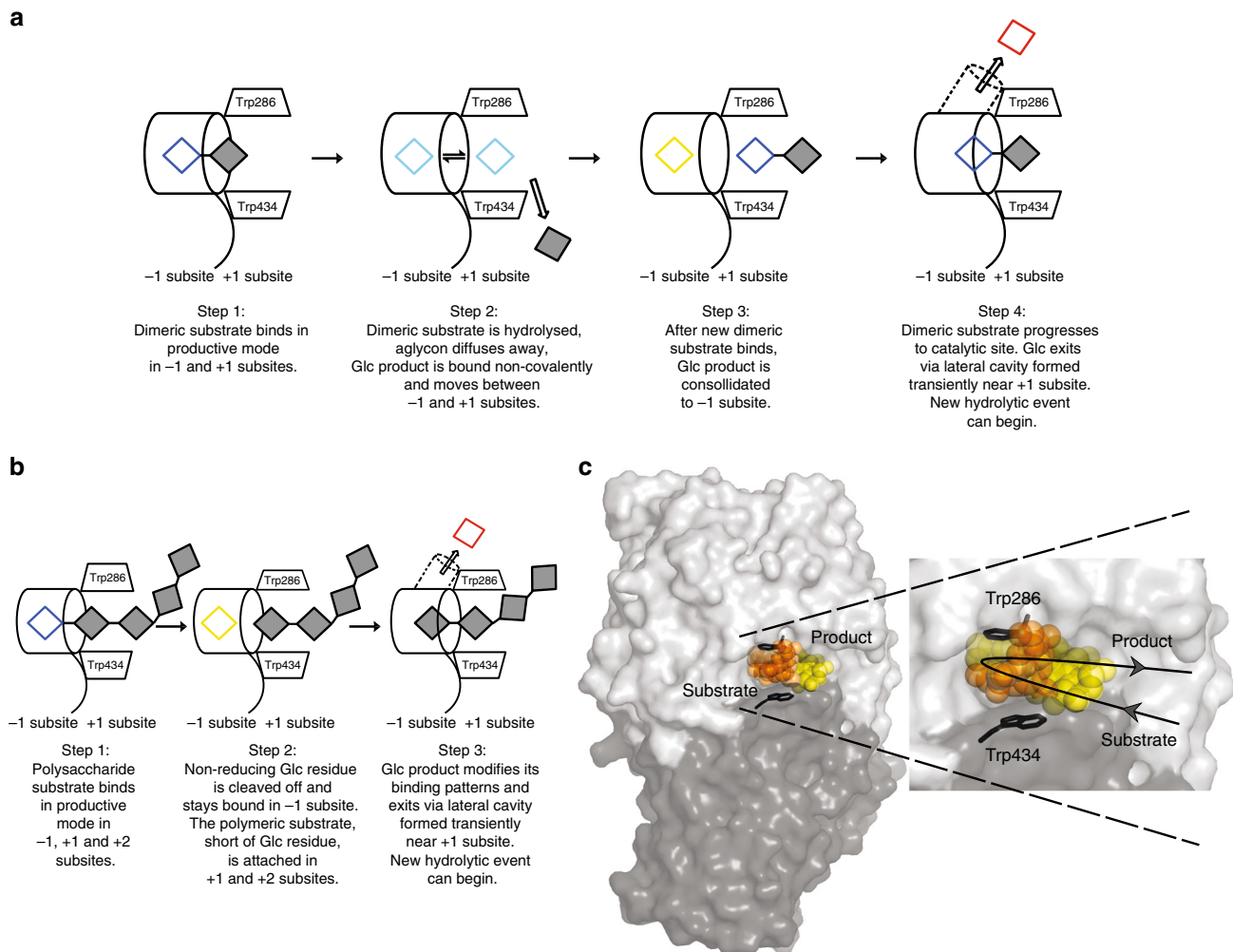


Fig. 8 The mechanism of substrate-product assisted processive catalysis by HvExoL. **a** Mechanism of Glc displacement with a disaccharide. After the disaccharide (empty blue and filled grey squares) bound in -1 and +1 subsites (step 1) is hydrolysed and an aglycon diffuses away (step 2), Glc (cyan square) remains non-covalently trapped and oscillates (double arrow) between the -1 and +1 subsites. Glc (yellow square) is consolidated to the -1 subsite after an incoming substrate binds (step 3), which later advances to the catalytic site. Glc (red square) modifies its binding patterns and exits (large arrow) via an autonomous and transient lateral cavity (cylinder in dashed lines) formed near the catalytic site (step 4); a next hydrolytic cycle can begin. **b** Mechanism of Glc displacement with a polysaccharide to facilitate substrate-product assisted processive catalysis. After non-reducing (blue square) and penultimate (filled square) residues bind in a productive mode at the -1 and +1 subsites (step 1), the non-reducing Glc residue (yellow square) is cleaved off, with the remainder of the substrate attached (step 2). Glc (red square) modifies its binding patterns and is released (large arrow) via a lateral cavity (cylinder in dashed lines, step 3). Here, the next hydrolytic cycle continues with the same polysaccharide, where the polysaccharide, short of one Glc (filled squares) advances into the catalytic site after uninterrupted binding. Non-reducing (blue) and hydrolysed (cyan-yellow-red in (a), yellow-red in (b)) Glc is shown. **c** Structural basis of substrate-product assisted processive catalysis. Left: The substrate (orange spheres) and hydrolysed (cyan-yellow-red in (a), yellow-red in (b)) Glc is shown. Right: detail of the active site. Circle line with arrows indicates directionality during processive catalysis. The image was generated using the coordinates of HvExoL:G6SG-OMe and HvExoL:G2SG-OMe complexes (Fig. 3c-d; orange spheres), and the Glc displacement snapshots based on PELE calculations (Fig. 6e-h; yellow spheres)

We demonstrate that the substrate binding and product displacement routes proceed through stages that are carefully orchestrated in succession (Fig. 8). We suggest that a disaccharide (Fig. 8a) binds near the catalytic site at the +1 and putative +2 subsites most likely through random collisions, from where it progresses to the -1 and +1 subsites. Following binding in a productive mode (step 1), the disaccharide (empty blue and filled grey squares) is poised for hydrolysis. After the non-reducing Glc is cleaved off and aglycon diffuses away, the Glc product is entrapped non-covalently, whereby it may oscillate between the -1 and +1 subsites (cyan square) (step 2). After the next disaccharide binds at +1 and +2 subsites, the Glc product

(yellow square) is consolidated to the -1 subsite. This is linked to Tyr253 rotation, which enlarges the lateral cavity adjacent to the -1 and +1 subsites (step 3). Next, the Glc product (red square) traverses from the -1 subsite through rotations of the Arg158 and Asp285 sidechains and associated backbone atoms, and bypasses the toll-like barrier shaped by Asp285, Glu491 and Arg158, via this autonomous lateral cavity (step 4, cavity shown as a barrel in dashed lines). This cavity is formed transiently near the +1 subsite, through which the Glc product exits (arrow) via an aperture into the bulk solvent (Fig. 8a; step 4). Strikingly, when the toll-like barrier is disrupted in the R158A/E161A variant, Glc no longer binds in the active site (Fig. 7), which underlines the

vital role of the barrier in Glc entrapment. The sequence of events involving incoming substrate binding and Glc displacement are captured in Supplementary Movie 1.

We propose a similar substrate/product progression route (Fig. 8b) takes place with a polysaccharide (empty blue and filled grey squares), often with differently linked β -D-glucosyl residues^{8,19}, through processive catalysis. Here, (step 1) after Glc (blue square) is cleaved off from the polysaccharide, (step 2) the Glc product (yellow square) remains bound at the -1 subsite, while the remainder of the polysaccharide (grey squares) stay attached at the $+1$ and putative $+2$ subsites. As the Glc product (red square) exits (arrow) through the lateral cavity (step 3, shown as a barrel in dashed lines), the polysaccharide can advance to the -1 and $+1$ subsites, such that an uninterrupted hydrolysis of the polymer continues until it is completely hydrolysed. In both instances, after Glc egresses (Fig. 8a, b), the next hydrolytic cycle that is facilitated by the dedicated substrate binding and product displacement routes, can continue. Consequently, this coordinated system with oligo- and polymeric substrates, resembles a loop or a conveyor belt that directionally shuffles substrates and ensures efficient catalysis (Fig. 8c). Although in all computations, we modelled Glc in 4C_1 conformation, we do not rule out that the Glc distortion to a higher energy of the 1S_3 conformer, as it occurs in crystal structures, may contribute to displacement of the Glc product or its progression.

To our knowledge, this type of processive catalysis^{47–49} has not been described in any exo-acting hydrolase with a pocket-shaped active site like that of HvExoI. In this work, we define processive catalysis in broader terms, rather than those based on hydrolytic profiles of products⁵⁰. Considering that HvExoI operates on polysaccharides such as (1,3;1,4)- β -D-glucans contained in plant cell walls^{9,51}, this finding is unsurprising, as the enzyme's efficient hydrolysis is the prerequisite for seed germination and the development of a plant embryo^{9,23,51}. Processive catalysis (non-dissociative sequential degradation) has been accepted for endo-acting hydrolases with cleft- or groove-shaped catalytic sites, such as (1,3)- and (1,3–1,4)- β -D-glucan endohydrolases⁵², and (1,4)- β -D-glucan endohydrolases⁵³, and for exo-acting chitinases⁵⁰ and cellobiohydrolases⁵⁴ with ridge- or tunnel-shaped sites. Conversely, β -D-glucosidases and exo-acting hydrolases with funnel-, crater- or pocket-shaped^{6,55} active sites are deemed to be non-processive^{48,49}.

Our concept of substrate and product processive co-operative routes with pocket-shaped exo-hydrolases^{6–9} is associated with chemical signalling, where entrapped products, bound substrates and enzymes co-operatively create pathways dedicated to product dissociation. The hallmark and advantage of processive catalysis is a high efficiency, where oligo- or polysaccharides are directionally threaded without the catalyst losing contact with substrates to keep binding entropic cost low, while performing multiple hydrolytic events on the same substrate molecule. While this work was focussed on a plant GH3 hydrolase, we suggest that substrate-product assisted processive catalysis may be more prevalent among exo-hydrolases with pocket-shaped active sites, irrespective of their substrate specificity. To this end, our findings could promote further investigations of enzymes involved in biomass degradation, where one of the challenges is to circumvent the bottleneck of product inhibition, a problem that could be addressed through the better understanding of enzyme catalytic cycles. More broadly, substrate-product assisted processive catalysis in enzymes with pocket-shaped catalytic sites is likely to have significance in other enzymes due to the plasticity of protein structures. It is exactly this plasticity that accounts in HvExoI for product dissociation, although in other enzymes the atomic details of reactants movements may be different.

The discovery of substrate-product assisted processive catalysis in HvExoI prompts investigations of the evolutionary origin of this mechanism. The availability of the presented experimental and computational tools alongside a rich source of information for the GH3 family enzymes that originate from all phyla^{9,18} (currently around 23,000 entries), will now allow the study of evolution of this catalytic mechanism. Preliminary analyses of 500 sequences related to HvExoI revealed that the residues underlying this mechanism are conserved in land plants, but absent in red and green algae, suggesting that this mechanism is rather ancient and has evolved in land plants about 470 million years ago.

We conclude that through the descriptions of associative and dissociative reactant trajectories that we explored using our interdisciplinary approach, we are now in a better position to understand how reactants advance in active sites, and how to improve enzyme catalytic rates, stability, product inhibition and drug discovery.

Methods

Materials. 3-Deoxy-glucose (3dGlc) and 4-deoxy-glucose (4dGlc) were obtained from GlycoTeam GmbH (Hamburg, Germany). N-octyl β -D-glucopyranoside (octyl-O-Glc) and n-octyl 1-thio- β -D-glucopyranoside (octyl-S-Glc) were from Anatrace Inc. (Maume, OH, USA). 4-Nitrophenyl β -D-glucopyranoside (4NPGlc), glycerol, myo-inositol and other chemicals were obtained from Sigma-Aldrich (USA) or as described previously^{6,7,17}. CH_3CN was from BDH Laboratory Supplies (UK), and polyethylene glycol 400 (PEG400; $n = 5–10$) was from Fluka Biochemica (Germany).

Synthesis of methyl 6-thio- β -gentiobioside. The synthesis proceeded in two steps as follows.

First, methyl 6-thio- β -gentiobioside heptaacetate (chemical structure in Supplementary Table 1): 2,3,4,6-Tetra-O-acetyl-1-thio- β -D-glucopyranoside (845 mg, 2.32 mmol)⁵⁶, 1,4-dithioerythritol (357 mg, 2.32 mmol) and cysteamine (1179 mg, 2.32 mmol) were successively added to a solution of methyl 2,3,4-tri-O-acetyl-6-deoxy-6-iodo- β -D-glucopyranoside (1 g, 2.32 mmol) in hexamethylphosphoramide (HMPA) (7 mL)⁵⁷. The mixture was kept for 1 h at room temperature then precipitated into ice-water (100 mL). The solid was collected on Celite, washed with water and then dissolved in CH_2Cl_2 . The organic phase was washed with water, dried over Na_2SO_4 and concentrated. Column chromatography in silica gel (EtOAc-light petroleum 1:1) afforded methyl 6-thio- β -gentiobioside heptaacetate (1.3 g, 84% yield). $[\alpha]_D^{25} -20.0$ (c 0.73, CHCl_3); ESI-HRMS(+) calcd for $\text{C}_{27}\text{H}_{38}\text{O}_{17}\text{NaS}$ 689.17219; found 689.17081; ${}^1\text{H}$ NMR (CDCl_3): δ 5.16 (t, 1 H, $J = 9.5$ Hz), 5.15 (t, 1 H, $J = 9.5$ Hz), 5.0 (t, 1 H, $J = 9.8$ Hz), 4.95 (t, 1 H, $J = 10.0$ Hz), 4.93 (dd, 1 H, $J = 8.0$ Hz, $J = 10$ Hz), 4.88 (t, 1 H, $J = 9.5$ Hz), 4.65 (d, 1 H, $J = 10.0$ Hz), 4.39 (d, 1 H, $J = 8.0$ Hz), 4.22 (dd, 1 H, $J = 5.0$ Hz, $J = 12.5$ Hz), 4.12 (dd, 1 H, $J = 2.0$ Hz, $J = 12.5$ Hz), 3.65 (m, 2 H), 3.51 (s, 3 H), 2.79 (m, 2 H), 2.08–1.97 (m, 21 H); ${}^{13}\text{C}$ NMR (CDCl_3): δ 169.6–169.3, 101.6, 83.9, 76.0, 74.4, 73.8, 72.7, 71.9, 71.3, 70.2, 68.2, 62.1, 57.1, 31.2, 20.77–20.72.

Second, methyl 6-S- β -D-glucopyranosyl-6-thio- β -D-glucopyranoside (methyl 6-thio- β -gentiobioside; G6SG-OMe) (chemical structure in Supplementary Table 1): sodium methylate 1 M (300 μL) was added to a solution of methyl 6-thio- β -gentiobioside heptaacetate (300 mg, 0.45 mmol) in MeOH (30 mL). After 4 h at room temperature, the mixture was neutralised with Amberlite IR 120 H^+ resin, filtered and concentrated. The residue was dissolved with water, and freeze dried affording methyl 6-thio- β -gentiobioside (166 mg, 99% yield). $[\alpha]_D^{25} -26.2$ (c 1.13, H_2O); ESI-HRMS(+) calcd for $\text{C}_{13}\text{H}_{24}\text{O}_{10}\text{NaS}$ 395.09824; found 395.09750; ${}^1\text{H}$ NMR (D_2O): δ 4.69 (d, 1 H, $J = 10.0$ Hz), 4.43 (d, 1 H, $J = 8.0$ Hz), 3.94 (dd, 1 H, $J = 2$ Hz, $J = 12.5$ Hz), 3.75 (dd, 1 H, $J = 5.5$ Hz, $J = 12.0$ Hz), 3.68 (m, 1 H), 3.62 (s, 3 H), 3.54–3.24 (m, 8 H), 2.97 (dd, 1 H, $J = 8.0$ Hz, $J = 14.5$ Hz); ${}^{13}\text{C}$ NMR (D_2O): δ 103.3, 86.0, 79.8, 77.2, 76.0, 75.5, 73.2, 72.5, 72.4, 69.5, 60.9, 57.4, 31.3.

Synthesis of methyl 2-thio- β -sophoroside. The synthesis proceeded in three steps as follows.

First, methyl 3,4,6-tri-O-acetyl-2-O-trifluoromethanesulfonyl- β -D-mannopyranoside (chemical structure in Supplementary Table 1): Distilled pyridine (0.5 mL) and trifluoromethanesulfonic anhydride (113 μL , 3.6 equiv) were successively added to an ice-cold solution of methyl 3,4,6-tri-O-acetyl- β -D-mannopyranoside⁵⁸ (60 mg, 0.19 mmol) in anhydrous CH_2Cl_2 (3 mL). The mixture was stirred at 0 °C for 30 min and at room temperature for 1 h. The solution was diluted with water and extracted with CH_2Cl_2 . The organic phase was washed with aqueous KHSO_4 (20%, v/v), saturated aqueous NaHCO_3 , dried over anhydrous Na_2SO_4 and concentrated. The crude triflate (84 mg) was obtained in a quantitative yield and directly used without purification.

Second, methyl 3,4,6-tri-*O*-acetyl-2-*S*-(2,3,4,6-tetra-*O*-acetyl- β -D-glucopyranosyl)-2-thio- β -D-glucopyranoside (methyl 2-thio- β -sophorose heptaacetate) (chemical structure in Supplementary Table 1): diethylamine (1 mL) under Argon was added to a stirred solution of 2,3,4,6-tetra-*O*-acetyl-1-thio- β -D-glucopyranose⁵⁹ (91 mg, 0.22 mmol), to which crude methyl 3,4,6-tri-*O*-acetyl-2-*O*-trifluoromethanesulfonyl- β -D-mannopyranoside (84 mg, 0.19 mmol) in anhydrous DMF (10 mL) was added. The mixture was stirred overnight at room temperature and then concentrated. A solution of the residue in CH_2Cl_2 was washed in water, dried over anhydrous Na_2SO_4 and concentrated. Flash chromatography on silica gel (ethyl acetate/petroleum ether, 1:1) afforded methyl 2-thio- β -sophorose heptaacetate (26 mg, 21%). Analytical data agreed with those reported previously⁶⁰. ESI-MS: $m/z = 689 [\text{M}^+\text{Na}]^+$; ^1H NMR (CDCl_3) δ 5.12 (t, 1 H, $J = 9.3$ Hz), 5.04 (t, 1 H, $J = 9.4$ Hz), 4.97 (m, 3 H), 4.73 (d, 1 H, $J = 10.2$ Hz), 4.36 (d, 1 H, $J = 8.5$ Hz), 4.30 (dd, 1 H, $J = 12.3$ Hz), 4.26 (dd, 1 H, $J = 4.8$, $J = 12.5$ Hz), 4.13 (m, 2 H), 3.66 (2ddd, 2 H), 3.56 (s, 3 H), 3.06 (dd, 1 H, $J = 8.7$ Hz, $J = 10.7$ Hz), 2.08–1.99 (7 s, 21 H); ^{13}C NMR (CDCl_3) δ 170.57–169.22 (CO), 103.79, 83.79, 76.03, 74.02, 73.11, 71.77, 71.48, 69.51, 68.25, 62.25, 62.15, 57.40, 49.90, 20.86–20.77 (CH_3).

Third, methyl 2-*S*- β -D-glucopyranosyl-2-thio- β -D-glucopyranoside (methyl 2-thio- β -sophorose; G2SG-OME) (chemical structure in Supplementary Table 1): 1 M sodium methylate (200 μL) was added to a solution of methyl 2-thio- β -sophorose heptaacetate (26 mg, 0.04 mmol) in MeOH (5 mL) and stirred at ambient temperature for 2 h. The mixture was neutralised with the Amberlite IR 120 H^+ resin, filtered and concentrated. After lyophilisation, pure methyl 2-thio- β -sophorose was isolated with 97% yield (14.5 mg). $[\alpha]_{\text{D}} -15.02$ (c 0.3, H_2O). ESI-MS: $m/z = 395 [\text{M}^+\text{Na}]^+$; ^1H NMR (CDCl_3) δ 4.76 (d, 1 H, $J = 10.0$ Hz), 4.57 (d, 1 H, $J = 9.0$ Hz), 3.96 (m, 2 H), 3.78 (m, 2 H), 3.66 (dd, 1 H, $J = 8.8$ Hz, $J = 10.8$ Hz), 3.61 (s, 3 H), 3.56–3.33 (m, 6 H), 2.85 (dd, 1 H, $J = 9.2$ Hz, $J = 10.6$ Hz); ^{13}C NMR (CDCl_3) δ 102.36, 84.43, 79.76, 77.02, 75.61, 74.97, 72.82, 70.63, 69.32, 60.77 (2 C), 57.23, 51.26.

Steady-state affinity binding with recombinant HvExol. Interaction between recombinant HvExol and analytes Glc (molecular mass 180.2 g), G6SG-OME (372.4 g) and PEG ($n = 5$ –10); (380–420 g) were performed using Surface Plasmon Resonance (SPR) at 25 °C using the Biacore X100 V2.0.1 instrument (GE Healthcare, USA) equipped with the plus package. Purified recombinant HvExol at 99% homogeneity (detected by SDS-PAGE) was covalently attached to a CM5 chip (GE Healthcare) in 10 mM sodium acetate buffer, pH 5.0. The carboxymethyl groups present on the dextran layer of the CM5 chip were activated with 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) and *N*-hydroxysuccinimide (NHS) (GE Healthcare). Any unoccupied reactive sites were blocked by a 7 min injection of 1 M ethanolamine hydrochloride-NaOH, pH 8.5. A ligand density of ~15,000 RU of recombinant HvExol was achieved. Due to the fast rate constants of interactions, it was impossible to obtain individual rate constants k_a and k_d . Hence equilibrium analyses were performed for all protein-analyte interactions. For each interaction serial dilutions of analytes covering the $10 \cdot K_D - 0.1 \cdot K_D$ concentration ranges were prepared in 10 mM HEPES-NaOH buffer, pH 7.4 with 150 mM NaCl. Samples were injected onto the surface of a chip at a rate of 30 $\mu\text{L}/\text{min}$ for 90 s, after which the formed enzyme/analyte complexes could dissociate for 90 s. As all analytes dissociated completely from the surface of the recombinant enzyme immobilised on a chip, no regeneration steps were needed. All responses were double referenced using a blank-immobilised reference surface (NHS + EDC-activated/deactivated) and a series of blank injections conducted to account for the system and injection artefacts. For each analyte at least one concentration was injected in duplicate. Triplicate experiments were performed, and the double-referenced data were globally fitted using a 1:1 equilibrium interaction model and the Biacore T200 v2.0 analysis software (GE Healthcare). Steady-state affinity binding constants K_D (Supplementary Fig. 3) were calculated 30 s after the onset of injections using a 5 s window in an average response in the fit.

Inhibition constants K_i with native HvExol. K_i values of deoxy-glucose and alkyl β -D-glucoside derivatives, and G2SG-OME with native HvExol were derived from hydrolytic rates. Enzyme inactivation was monitored at 30 °C by incubating 8 μmoles of the enzyme in 100 mM sodium acetate buffer, pH 5.25, containing 0.8–6 mM 4NPGlc as a substrate, 160 $\mu\text{g}/\text{mL}$ BSA, and 3dGlc, 4dGlc, octyl-*O*-Glc, octyl-*S*-Glc and G2SG-OME. Each inhibitor was tested at six concentrations at 0.4–3 times the K_i values in duplicate. We used the $v = V_{\text{max}} S / [K_M \cdot (1 + I/K_i) + S]$ equation to calculate K_i constant; K_i is defined as a dissociation constant of the enzyme-inhibitor complex. The enzyme activity was monitored at 410 nm and Dixon plots were used to determine K_i under the pseudo-first-order rate limiting conditions by a non-linear regression analysis, as described previously²².

Expression and crystallisation of recombinant HvExol. The optimised HvExol cDNA (GenBank Accession GU441535; was subcloned into the pPICZaBNH₈ vector, from which the protein was expressed in *Pichia pastoris* strain SMD1168H (Invitrogen, Carlsbad, CA, USA). The R158A/E161A variant (primers listed in Supplementary Table 6) was prepared by site-directed mutagenesis and expressed in *Pichia*²⁴. Recombinant HvExol enzymes were purified using SP-Sephacrose

cation-exchange and immobilised metal affinity chromatography^{24,25}. Purified recombinant wild-type (WT) and variant HvExol were crystallised via a hanging-drop vapour-diffusion method, using macro- and cross-seeding with WT native crystal seeds. The fully-grown crystals of recombinant HvExol reached up to $500 \times 250 \times 375 \mu\text{m}$ sizes after 5 to 14 days and were used for diffraction²⁵.

Crystal structure determination. Enzyme crystals with lengths of up to 400 μm in the longest dimensions were transferred in 100 mM HEPES-NaOH buffer, pH 7.0, containing 1.2% (w/v) PEG ($n = 5$ –10) and 1.7 M ammonium sulphate (solution A), where 3-deoxy-glucose (3dGlc), 4-deoxy-glucose (4dGlc), *n*-octyl β -D-glucopyranoside (octyl-*O*-Glc), *n*-octyl 1-thio- β -D-glucopyranoside (octyl-*S*-Glc), methyl 2-thio- β -sophorose (G2SG-OME), methyl 6-thio- β -gentiobioside (G6SG-OME) or Glc were supplied during ligand perfusions. The final concentrations of ligands were between 5–20 mM and perfusion proceeded during 5–720 min. Alternatively, the crystals were deposited in a solution A with the concentration of PEG increased to the 35% (v/v). After various perfusion times at 4 ± 2 °C crystals were cryo-protected with 15% (v/v) glycerol in the solution A and mounted on a synchrotron goniometer in a stream of N_2 gas at 100 K (Oxford Instruments, UK). X-ray diffraction data (native, and with 3dGlc and 4dGlc) were collected on the undulator beamline BioCARS 14-ID-B at the University of Chicago Advanced Photon Source (USA) equipped with a bent cylindrical Si-mirror (Rh coating) Diamond(111) double-bounce monochromator and focussing to the MARCCD-165 detector. X-ray diffraction data sets of crystals with PEG and with octyl-*O*-Glc and octyl-*S*-Glc were collected using the multipole wiggler beamline BL5 at the Photon Factory (Japan), which was fitted with a collimating mirror, double-crystal Si(111) monochromator and a focusing to the ADSC Quantum 315 CCD detector. X-ray diffraction data from recombinant WT and variants were collected at the MX1 and MX2 beamlines of the Australian Synchrotron (Australia) at 100 K (Oxford Instruments) or at ambient temperature (291 K) with a collimating mirror, double-crystal Si(111) monochromator 03BM1 dipole/bending magnet and ADSC Quantum 210r Detector. In the latter case, crystals were mounted using transparent polymer capillaries from the RT Screening kit (MiTeGen). All data were collected at 0.5 – 1° oscillations throughout 180–720°. Data were processed using the DENZO/SCALEPACK HKL 2000 suite of programmes⁶¹. Autoindexing determined that space group $P4_32_12$ of crystals were consistent with the tetragonal space group $P43212$ in all instances. The structures of the enzyme-ligand complexes were refined using CCP4 REFMAC5⁶² and PHENIX⁶³. We used the solved structures (PDB accessions: 1EX1, 1IEQ, 1IEV, 1IEW, 1IEX and 1J8V) without ligands, ions, glycerol and water molecules, as starting models for refinements of HvExol in complex with ligands^{7,15–17}. The iterative model building using XtalView/MIFIT⁶⁴, Coot⁶⁵ and refinements with REFMAC5⁶² allowed tracing of all the residues. Following convergence in a standard refinement, a further improvement of about 2% in the Rwork/Rfree factors ratio was achieved by refining domain 1, (residues 1–357) and domain 2, (residues 374–559), as two independent anisotropic entities with translation-libration-screw (TLS) motion^{66,67}. Electron densities for the ligands were well-defined in the active site regions at 3σ level in $m|F_o| - |F_c|$ maps and water molecules were located automatically with CCP4 ARP at levels higher than 3σ . Water molecules were retained if they satisfied H-bond criteria and if their $2m|F_o| - |D|F_c|$ electron density maps were confirmed after refinements, where m is the figure of merit and D is an estimated coordinate error. During model building and refinement, 5% of the data were flagged for cross-validation to monitor the progress of refinements, using R_{free} statistics^{68,69}. The PROCHECK programme⁶⁹ was used to check the geometrical quality of the models. Ramachandran plots^{70,71} of all structures showed that 99.8% of residues were found in the most favourable, additionally allowed and generously allowed regions of the plot, with well-defined electron density for Ile432, the only residue located in disallowed regions. Data collection, cell parameters and refinement statistics of all structures are summarised in Supplementary Tables 7–10. Graphics images were prepared with PyMol (Schrödinger, USA); electrostatic potentials were generated with a solvent probe radius at 1.40 Å, and protein and solvent dielectric constants 2 and 78, respectively.

GC/MS with evaporative light scattering detection. Six crystals of native HvExol with well-defined edges in a longest axis dimension of ~80–120 μm were collected from a hanging drop in a cat whisker, deposited in a test tube, washed in solution A, pelleted by low speed centrifugation ($500 \times g$, 1 min) and dissolved in 100 μL of 90% (v/v) ethanol. The mixture was treated for 5 min at 80–85 °C, held on ice for 1 h and centrifuged ($4000 \times g$, 10 min). The supernatant was recovered, the pellet washed twice in 100 μL of re-distilled water (conductance 24 μS), and the supernatants were combined and freeze dried. The extract obtained from crystals, a freeze-dried mother liquor, and a mother liquor with authentic Glc (100 ng) being added, were dissolved in 25 μL 65% (v/v) CH_3CN and separated by high performance liquid chromatography (HPLC) coupled with the Evaporative Light Scattering Detector (ELSD model 800, Alltech Associates Inc.). The samples were eluted isocratically with 65% (v/v) CH_3CN at a flow rate of 0.6 mL/min. The fractions eluting between 6–7.3 min were collected and freeze dried. A model 1090 HPLC liquid system controlled by ChemStation software (Agilent Technologies) was used for chromatography, and separation was carried out on a Prevail

Carbohydrate ES column, 5 μ m, 250 \times 4.6 mm (Alltech Associates Inc.). The column temperature during separations was 21 °C. The column eluent was split in 6.5 (collect): 1 (detector) ratio. The detector drift tube was at 40 °C, the nitrogen inlet pressure was 1.5 bars and the signal GAIN setting was set at 4. The HPLC-eluted fractions (6–7.3 min) were reduced with 1 M NaBD₄ in 2 M NH₄OH for 18 h at 4 °C and acetylated with acetic anhydride. Myo-inositol (20 ng) was added as an internal standard. Acetylated alditols were analysed on a low polarity 25 m \times 0.25 mm (internal diameter) Chrompack Capillary Column CP-Sil 5 CB LB/MS (Varian Inc., CA, USA) using He as a gas flow with a Hewlett-Packard 6890 Series GC System and a Hewlett-Packard 5973 mass selective detector (Agilent Technologies)⁷².

NMR spectroscopy of Glc bound to HvExoI. NMR spectra were acquired at 283 K on the Bruker AVANCE III 600 MHz and 800 MHz spectrometers (Bruker, MA, USA). For Saturation Transfer Difference (STD) experiments, a stock solution of recombinant HvExoI deglycosylated by endoglycosidase H²⁵ in milliQ water (143 μ M, 20 mM sodium acetate buffer, pH 5.25, 159 mM NaCl) was diluted to 40 μ M in the same buffer in D₂O, and 60 equivalents (2.4 mM) of Glc from a stock solution in D₂O were added. For transferred Nuclear Overhauser Effect Spectroscopy (trNOESY) experiments, the same recombinant enzyme stock solution was diluted to 57 μ M with a deuterated 20 mM sodium acetate buffer, pH 5.25, to which five equivalents of Glc (0.285 mM) were added followed by 300 msec mixing time. In a separate experiment, to the 57 μ M recombinant enzyme stock solution, five equivalents of Glc and three equivalents of thiocellobiose (0.171 mM) were added, both from a stock solution in D₂O, followed by 300 msec mixing time. STD spectra were acquired with 2048 scans, using 30 ms Gaussian shaped pulses for selective protein saturation at 0.65 ppm; the reference spectrum employed off-resonant saturation at 100 ppm. Different saturation times (0.5, 1, 2, 3 and 5 s) were used. Prior to Free Induction Decay acquisition, an excitation sculpting double echo for water suppression and a T2 filter (100 ms XY16 with 500 μ s echo times to minimise dephasing from JHH coupling evolution) protein signal suppression was inserted. STD spectra for free Glc were acquired under the same conditions to show that no direct ligand saturation occurred. trNOESY spectra were acquired with 64 scans, 4 K points in the direct dimension, 256 points in the indirect dimension and 300 ms mixing time. For water suppression we employed a WATERGATE scheme using 3–9–19 binomial pulses. ¹H signal assignments of ligands were derived from two-dimensional Total Correlation Spectroscopy (TOCSY), NOESY and ¹³C-Heteronuclear Single Quantum Coherence (HSQC) spectra. The NOESY spectrum of free thiocellobiose was acquired at 313 K with 700 ms mixing time to obtain positive NOEs. A NOESY spectrum acquired under the conditions used for the trNOESY spectrum (800 MHz, 283 K, 300 ms mixing time) yielded no observable NOEs, indicating NOE zero-crossing for thiocellobiose (for $\tau_c = 1.12/\omega_0$).

Classical MD simulations. The structure of native HvExoI in complex with Glc (defined as β -D-glucopyranose) (PDB 3WLH) with occupancy 0.5 for –1 and +1 subsites, suggests that the two subsites are alternatively occupied. We set up two enzyme-product complexes, with Glc bound in the –1 subsite, whereas the +1 subsite was occupied by water molecules, and vice versa. The protonation state of the titratable residues was selected based on their hydrogen bonding environment. Namely, all Arg, Lys, Asp and Glu residues were in ionised form, except for the acid-base residue Glu491, which was protonated (i.e. non-ionised) to represent the catalytic acid/base form required in the first step of the reaction, and the predominant form at the optimal pH of the enzyme. Histidine residues 98, 111, 234, 331, 419, 487 and 586 were considered as N_ε-protonated and His207, His262 and His377 as N_δ-protonated. Assigned titration states agreed with pK_a values predicted by PROPKA3.1 and with the optimised hydrogen placement algorithm of MolProbity. Disulfide bridges were considered between Cys151 and Cys159, and between Cys513 and Cys518. Three sodium ions were included to achieve neutrality, in addition to 150 mM NaCl to mimic the physiological ionic strength. The system size consisted of approximately 150,000 atoms. The initial structures were equilibrated by classical MD simulations. Protein and Glc were described via the parm99SB⁷³ and the Glycam06⁷⁴ force fields, respectively, whereas the TIP3P model⁷⁰ was used for water molecules and Åqvist parameters for the sodium and chloride ions. The classical MD simulations were performed with NAMD2.9^{76,77}. The SHAKE/RATTLE⁷⁸ algorithm was used to constrain the bonds involving hydrogen atoms in both the solvent and the solute. Multiple time step integration was carried out in r-RESPA; the production simulations used a base time step of 2 fs and a secondary time step of 4 fs for long-range interactions. Non-bonded interactions were computed at every step with a cut-off separation of 9 Å. Periodic boundary conditions and the particle mesh Ewald method were employed to evaluate long-range electrostatic interactions. Simulations were carried out in the NPT (N, number of particles; P, system pressure; T, temperature constant) ensemble, with a Nosé-Hoover thermostat and a Langevin piston Nosé-Hoover barostat^{79,80} to maintain the temperature at 300 K and the pressure at 1 atm, respectively. Each of the two enzyme-product complex simulations was run for 40 ns without restraints, ensuring that the system reached equilibrium. Analyses of trajectories used VMD and in-house scripts. One representative snapshot from the last time window (\approx 5 ns) of each simulation was taken as the starting structure for further QM/MM MD simulations. Crystallographically observed ⁴C₁ Glc

conformations and protein-sugar interactions were consistently maintained during classical MD.

QM/MM MD simulations. Classical MD snapshots of two enzyme-product complexes were submitted to QM/MM MD simulations to re-equilibrate the system within the QM/MM Hamiltonian. The method by Laio and co-workers⁸¹ was used, which combines Car-Parrinello MD⁸² with force field-based MD. Only the atoms of Glc were included in the QM region (24 atoms), whereas the remainder of the protein and solvent were in the MM region. The MM system was treated with the Amber force field, for consistency with the previous classical MD simulations. The QM system was treated with Density Functional Theory (DFT), using a plane wave (PW) basis set with a kinetic energy cut-off of 70 Ry, Martins-Troullier pseudopotentials⁸³ and the Perdew-Burke-Ernzerhoff (PBE) exchange-correlation functional⁸⁴; this is at the same level theory as that described for sugar conformational Free Energy Landscapes (FELs)^{33,34}. The QM system was enclosed in an isolated orthorhombic box of size 14.0 Å \times 13.2 Å \times 13.5 Å (–1 subsite) or 12.0 Å \times 15.3 Å \times 13.2 Å (+1 subsite). The electronic fictitious was set at 700 au and the simulation time step was at 0.12 fs. Both complexes were equilibrated at 300 K for around 5 ps before starting metadynamics simulations of Glc conformational FEL. A similar procedure was used in previous studies of conformational preferences of sugars in carbohydrate-active enzymes³⁴.

Ab initio metadynamics simulations. QM/MM metadynamics^{85,86} simulations were performed to explore the conformational FEL of Glc bound in the –1 or the +1 subsites. We employed the metadynamics driver provided by the Plumed2 plugin, combined with the well-tempered metadynamics approach⁸⁷. Ring-puckering coordinates qx and qy were used as collective variables, as described for isolated Glc³³. Metadynamics parameters were selected from the analysis of evolution of the puckering coordinates during the unbiased QM/MM MD simulation and from expected energy barrier heights based on our previous work on both isolated Glc and a β -D-glucoside bound in the active site of 1,3;1,4- β -D-glucanase³⁴. The height and the width of Gaussian terms were initially set at 1 kcal mol^{–1} and 0.1 Å, respectively, and a bias factor of $\gamma = 10$ was used for the well-tempered approach. The deposition time was set at 24 fs (200 MD steps). The same set up was used for both enzyme-product complexes. Simulations were stopped after adding 7,668 Gaussians (enzyme-product complex with Glc bound in the –1 subsite) and 7725 Gaussians (Glc bound in +1 subsite). Analyses of evolution of puckering coordinates and FEL showed that the metadynamics simulation reached a diffusive regime and that the energy differences converged.

Computational methods to explore the Glc exit path in HvExoI. MD simulations in complex with Glc trapped in the –1 subsite (PDB 3WLH) proceeded as follows. Hydrogen atoms were added and protonation states of titratable residues at pH 7.0 were assigned with PROPKA3.1⁸⁸ through PDB2PQR version 2.0.0⁸⁹. In this complex, Glu491 was modelled in a non-ionised form; this is the required protonation state for the acid/base catalyst in substrate/product complexes of retaining GH3 enzymes that catalyse hydrolysis by the generally accepted double-displacement reaction mechanism²². The entire system was solvated in a water box of 103.0 \times 88.0 \times 101.0 Å³ volume, which ensured that there was a 5 Å thick layer of water molecules in each direction from any atom. Sodium ions were added to neutralise the total charge of the system forming a fully solvated model of 83,428 atoms. The system was energy-minimised to remove steric clashes and heated through 4 ns equilibration (integration steps of 2 fs) until reaching the temperature of 300 K at 1 Bar. This process also involved the application of positional restraints on the protein and ligand atoms that were gradually released until the whole system was fully equilibrated. Equilibration of the system was stopped when root-mean-square deviation (RMSD) values of the main chain reached the stable value lower than 1.60 Å. Two independent simulations of 100 ns each were run under NPT conditions, whereas temperature and pressure were controlled through Langevin dynamics and the Nosé-Hoover algorithm combined with the Langevin piston method^{79,80}, respectively. Periodic boundary conditions were applied, and the SHAKE algorithm⁷⁸ was used to adjust O-H separations of water molecules. The cut-off value of 12 Å was used for non-bonded interactions. All calculations were carried out with NAMD2.9^{76,77} using the CHARMM22 force field^{84,85}, for the protein and TIP3P for water molecules⁷⁵, while the CHARMM36 all-atom carbohydrate force field⁹⁰ was used for Glc.

Docking of Glc and disaccharides in HvExoI. Molecular docking calculations of Glc and β -D-glucopyranosyl-(1,2)-D-glucose (G2OG) or β -D-glucopyranosyl-(1,3)-D-glucose (G3OG) to HvExoI for selected protein structures were performed with Gold5.2⁹¹ applying a search cavity of the 20 Å radius and using the Goldscore scoring function. Side-chain flexibility for Glu491 was allowed, whereas Trp434 and Trp286 were flexible as specified below. Complex 1 was generated by docking Glc to the HvExoI:G2OG complex derived from the HvExoI:G2SG-OME crystal structure (PDB 6MD6). Complex 2 was generated by docking G2OG to the HvExoI:Glc complex derived from the native HvExoI crystal structure (PDB 3WLH). Both complexes contain Glc bound at the –1 subsite and G2OG at +1 and putative +2 subsites. Complexes 3–6 were obtained by docking G2OG or G3OG to the HvExoI:Glc structure (PDB 3WLH) subjected to MD simulations,

with the major conformation of Trp434 (CA-CB-CG-CD1 dihedral angle of around -50°). MD simulations were carried out under periodic boundary conditions of NPT ensembles, as described for MD simulations of HvExoI in complex with Glc. Docking of G2OG, G3OG and β -D-glucopyranosyl-(1,6)-D-glucose (G6OG) lacking the Glc product in the active site was carried out as described above.

GaudiMM calculations. Ternary complexes 1–3 before MD simulation (as described above) and complex 3 after MD simulation with substrates bound in a productive mode, were used as starting points to investigate Glc displacement pathways from the -1 subsite of HvExoI. The protein contained Glc at the -1 subsite, and G2OG or G3OG attached at $+1$ and putative $+2$ subsites. GaudiMM²⁰ (<https://github.com/insilichem/gaudi>), a recently developed modular multi-objective genetic algorithm platform allows conformational exploration of defined genes (defined below) with multiple evaluation operators of fitness or objectives. The GPathFinder extension was used, defining as genes protein and substrates molecules, rotational bonds of ligands and rotamers of the protein, and as objectives, minimisation of van der Waals contacts and maximisation of separations between Glc geometric centres at the -1 subsite. This afforded the low-cost computational pre-identification of putative exit channels. The ProDy (Protein Dynamics)⁹² Normal Mode Algorithm (NMA) was used to generate alternative starting structures for the GPathFinder. Normal mode calculations were performed with Tangram implementation on the UCSF Chimera suite⁹³, using residues grouped in clusters of 100 (to avoid protein distortion and achieve acceptable conformations), generating 10 modes with a cut-off of 15.0 Å and a Gamma LJ of 1.0. In the case of the crystal structure (PDB 3WLH), the second frame of the mode with frequency 31.2 cm^{-1} was chosen, while the second frame of the mode with frequency 28.7 cm^{-1} was selected for complex 3. In both cases, the displacement parameter in the Tangram interface was set to 150.

PELE calculations. The PELE software (<https://pele.bsc.es>)^{21,94} was used to simulate Glc displacement and to identify the Glc migration route from the -1 subsite. PELE is a Monte Carlo based algorithm that generates configurations along a path through a sequential ligand and protein geometric perturbation scheme, side-chain conformational prediction and minimisation steps. The ligand (Glc) is perturbed by random translations and rotations, and by internal rotations of protein rotatable bonds, while C α carbons of the protein are displaced following one of the six lowest Anisotropic Normal Modes⁹⁵. The algorithm optimally arranges sidechains surrounding perturbed atoms using a rotameric library side-chain optimisation and energy minimisation with the OPLS-AA force field⁹⁶. The move of the system is accepted, defining a minimum, or rejected according to a Metropolis criterion for a given temperature. The PELE scheme performs an effective exploration of the protein energy landscape^{97–99}. Here, the exit path is defined by separations between the centre of masses of Glc molecules at the beginning (-1 subsite) and at the end of simulations (in solvent). Complexes 1–3 with Glc at the -1 subsite and G2OG or G3OG at $+1$ and putative $+2$ subsites were selected from MD simulations to identify potential exit pathways. One representative structure from each MD simulation of complexes 1–3 was selected as these complexes show G2OG or G3OG bound in the productive modes for subsequent catalysis. PELE simulations were terminated, when respective separations between centres of masses and C4-OH groups of Glc molecules in initial and final structures were above 15 Å.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The atomic coordinates and structure factors were deposited in the Protein Data Bank (www.pdb.org) with the PDB accessions: native HvExoI, and in complex with 3dGlc, 4dGlc, octyl-O-Glc, octyl-S-Glc and PEG400 are 3WLH, 3WLJ, 3WLK, 3WLM, 3WLN and 3WLL, respectively. The PDB accessions of WT recombinant ligand-free HvExoI and in complex with Glc, G2SG-OMe and G6SG-OMe are 3WLI, 3WLO, 6MD6 and 3WLP, respectively. The PDB accessions of recombinant R158A/E161A ligand-free HvExoI and in complex with Glc and G6SG-OMe, are 3WLQ, 3WLR and 6MI1, respectively. A reporting summary for this Article is available as a Supplementary Information file. All other data generated and/or analysed during the current study are available from the corresponding author upon reasonable request.

Code availability

Software applications described in Methods with associated references were used without code modifications. For analyses of geometrical parameters along MD simulations, we have used an in-house script, which is available upon request from the corresponding author.

Received: 10 December 2018 Accepted: 22 March 2019

Published online: 20 May 2019

References

- Warshel, A. et al. Electrostatic basis for enzyme catalysis. *Chem. Rev.* **10**, 3210–3235 (2006).
- Adamczyk, A. J., Cao, J., Kamerlin, S. C. & Warshel, A. Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proc. Natl Acad. Sci. USA* **108**, 14115–14120 (2011).
- Fried, S. D. & Boxer, S. G. Thermodynamic framework for identifying free energy inventories of enzyme catalytic cycles. *Proc. Natl Acad. Sci. USA* **110**, 12271–12276 (2013).
- Marc, A. & Engasser, J. M. Influence of substrate and product diffusion on heterogeneous kinetics of enzymic reversible reactions. *J. Theor. Biol.* **94**, 179–189 (1982).
- Elber, R. Long-timescale simulation methods. *Curr. Opin. Struct. Biol.* **15**, 151–156 (2005).
- Varghese, J. N., Hrmova, M. & Fincher, G. B. Three-dimensional structure of a barley β -D-glucan exohydrolase, a family 3 glycosyl hydrolase. *Structure* **7**, 179–190 (1999).
- Hrmova, M. et al. Catalytic mechanisms and reaction intermediates along the hydrolytic pathway of plant β -D-glucan glucosylhydrolase. *Structure* **9**, 1005–1016 (2001).
- Hrmova, M. & Fincher, G. B. Barley β -D-glucan exohydrolases substrate specificity and kinetic properties. *Carbohydr. Res.* **305**, 209–221 (1998).
- 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.* **342**, 1613–1623 (2007).
- Yoshida, E. et al. Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3 β -glucosidase from *Kluyveromyces marxianus*. *Biochem. J.* **431**, 39–49 (2010).
- McAndrew, R. P. et al. From soil to structure, a novel dimeric β -glucosidase belonging to glycoside hydrolase family 3 isolated from compost using metagenomic analysis. *J. Biol. Chem.* **288**, 14985–14992 (2013).
- Bacik, J.-P., Whitworth, G. E., Stubbs, K. A., Voadlo, D. J. & Mark, B. L. Active site plasticity within the glycoside hydrolase NagZ underlies a dynamic mechanism of substrate distortion. *Chem. Biol.* **19**, 1471–1482 (2012).
- Suzuki, K. et al. Crystal structures of glycoside hydrolase family 3 β -glucosidase 1 from *Aspergillus aculeatus*. *Biochem. J.* **452**, 211–221 (2013).
- Pachl, P. et al. Crystal structure of native α -L-rhamnosidase from *Aspergillus terreus*. *Acta Crystallogr.* **D74**, 1078–1084 (2018).
- Hrmova, M. et al. Three-dimensional structure of the barley β -D-glucan glucosylhydrolase in complex with a transition state mimic. *J. Biol. Chem.* **279**, 4970–4980 (2004).
- Hrmova, M. et al. Structural rationale for low nanomolar binding of transition state mimics to a family GH3 β -D-glucan glucosylhydrolase from barley. *Biochemistry* **44**, 16529–16539 (2005).
- Hrmova, M. et al. Structural basis for a broad specificity in higher plant β -D-glucan glucosylhydrolases. *Plant Cell* **14**, 1–22 (2002).
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
- Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**, 727–748 (1997).
- Rodríguez-Guerra Pedregal, J., Sciortino, G., Guasp, J., Municoy, M. & Maréchal, J.-D. GaudiMM: a modular multi-objective platform for molecular modeling. *J. Comput. Chem.* **38**, 2118–2126 (2017).
- Borrelli, K. W., Vitalis, A., Alcantara, R. & Guallar, V. PELE: protein energy landscape exploration. a novel Monte Carlo based technique. *J. Chem. Theory. Comput.* **1**, 1304–1311 (2005).
- Hrmova, M. et al. 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 (1996).
- 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 Crystallogr.* **D54**, 6887–689 (1998).
- Luang, S., Hrmova, M. & Ketudat Cairns, J. R. High-level expression of barley β -D-glucan exohydrolase HvExoI from a codon-optimized cDNA in *Pichia pastoris*. *Prot. Exp. Purif.* **73**, 90–98 (2010).
- Luang, S., Ketudat Cairns, J. R., Streltsov, V. A. & Hrmova, M. Crystallisation of wild-type and variant forms of a recombinant β -D-glucan glucosylhydrolase from barley (*Hordeum vulgare* L.) by macroseeding with wild-type native microcrystals and preliminary X-ray analysis. *Int. J. Mol. Sci.* **11**, 2759–2769 (2010).
- Cremer, D. & Pople, J. A. A general definition of ring puckering coordinates. *J. Am. Chem. Soc.* **97**, 1354–1358 (1975).
- Quirocho, F. A. Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions. *Ann. Rev. Biochem.* **55**, 287–315 (1986).

28. Fernandez-Alonso, M. C., Canada, F. J., Jimenez-Barbero, J. & Cuevas, G. Molecular recognition of saccharides by proteins. Insights on the origin of the carbohydrate-aromatic interactions. *J. Am. Chem. Soc.* **127**, 7379–7386 (2005).
29. Fersht, A. *Structure and Mechanism in Protein Science*, pp. 1–631, (W. H. Freeman and Co, New York, 1999).
30. Mayer, M. & Meyer, B. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* **38**, 1784–1788 (1999).
31. Clore, G. M. & Gronenborn, A. M. Theory and applications of the transferred Nuclear Overhauser Effect to the study of the conformations of small ligands bound to proteins. *J. Magn. Reson.* **48**, 402–417 (1982).
32. Montero, E. et al. NMR studies of the conformation of thiocellobiose bound to a beta-glucosidase from *Streptomyces* sp. *FEBS Lett.* **421**, 243–248 (1998).
33. Biarnés, X., Ardèvol, A., Iglesias-Fernández, J., Planas, A. & Rovira, C. Catalytic itinerary in 1,3-1,4- β -glucanase unraveled by QM/MM metadynamics. Charge is not yet fully developed at the oxocarbenium ion-like transition state. *J. Am. Chem. Soc.* **133**, 20301–20309 (2011).
34. Ardèvol, A. & Rovira, C. Reaction mechanisms in carbohydrate-active enzymes: glycoside hydrolases and glycosyltransferases. Insights from ab initio quantum mechanics/molecular mechanics dynamic simulations. *J. Am. Chem. Soc.* **137**, 7528–7547 (2015).
35. Lee, P. I. & Peppas, N. A. Prediction of polymer dissolution in swellable controlled release systems. *J. Control Release* **6**, 207–215 (1987).
36. Mills, J. E. & Dean, P. M. Three-dimensional hydrogen-bond geometry and probability information from a crystal survey. *J. Comput. Aided. Mol. Des.* **10**, 607–622 (1996).
37. Nashine, V. C., Hammes-Schiffer, D. & Benkovic, S. J. Coupled motions in enzyme catalysis. *Curr. Opin. Chem. Biol.* **14**, 644–651 (2010).
38. Buyong, M. & Nussinov, R. Enzyme dynamics point to stepwise conformational selection in catalysis. *Curr. Opin. Chem. Biol.* **14**, 652–659 (2010).
39. Lampe, J. N., Brandman, R., Sivaramakrishnan, S. & de Montellano, P. R. Two-dimensional NMR and all-atom molecular dynamics of cytochrome P450 CYP119 reveal hidden conformational substates. *J. Biol. Chem.* **285**, 9594–9603 (2010).
40. Pieraccini, S., Sironi, M. & Colombo, G. A molecular simulation analysis of the origins of regioselectivity. *Chem. Physics Lett* **418**, 373–376 (2006).
41. Sundareshan, V., Charton, J., Yamaguchi, M. & Stout, C. D. Conformational diversity in NAD(H) and interacting transhydrogenase nicotinamide nucleotide binding domains. *J. Mol. Biol.* **346**, 617–629 (2004).
42. Yang, L.-W. & Bahar, I. Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Structure* **13**, 893–904 (2005).
43. Gutteridge, A. & Thornton, J. Conformational changes observed in enzyme crystal structures upon substrate binding. *J. Mol. Biol.* **346**, 21–28 (2005).
44. Hammes-Schiffer, S. & Benkovic, S. J. Relating protein motion to catalysis. *Annu. Rev. Biochem.* **75**, 519–541 (2006).
45. Díaz, M. D., Fernández-Alonso, M. C., Cuevas, G., Cañada, F. J. & Jiménez-Barbero, J. On the role of aromatic-sugar interactions in the molecular recognition of carbohydrates: A 3D view by using NMR. *Pure Appl. Chem.* **80**, 1827–1835 (2008).
46. Hrmova, M. et al. Barley (1,3)- β -D-glucan endohydrolase mutants synthesise crystalline (1,3)- β -D-glucans. *J. Biol. Chem.* **277**, 30102–30111 (2002).
47. Davies, G. & Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853–859 (1995).
48. Sørensen, A., Lübeck, M., Lübeck, P. S. & Ahring, B. K. Fungal beta-glucosidases: a bottleneck in industrial use of lignocellulosic materials. *Biomolecules* **3**, 612–631 (2013).
49. van Dongen, S. F. M., Elemans, J. A., Rowan, A. E. & Nolte, R. J. Processive catalysis. *Angew. Chem. Int. Ed.* **52**, 11420–11428 (2014).
50. Sørle, M., Zakariassen, H., Norberg, A. L. & Eijsink, V. G. H. Processivity and substrate-binding in family 18 chitinases. *Biotransform.* **30**, 353–365 (2012).
51. Hrmova, M. & Fincher, G. B. Structure-function relationships of β -D-glucan endo- and exohydrolases from higher plants. *Plant Mol. Biol.* **47**, 73–91 (2001).
52. Varghese, J. N. et al. Three-dimensional structures of two plant β -glucan endohydrolases with distinct substrate specificities. *Proc. Natl Acad. Sci. USA* **91**, 2785–2789 (1994).
53. Kleywegt, G., Zou, J.-Y., Divne, C. & Davies, G. J. & Sinning, I. et al. The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 Å resolution, and a comparison with related enzymes. *J. Mol. Biol.* **273**, 1–15 (1997).
54. Divne, C., Stahlberg, J., Teeri, T. T. & Jones, T. A. High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from *Trichoderma reesei*. *J. Mol. Biol.* **275**, 309–325 (1998).
55. Isorna, P. et al. 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 (2006).
56. Horton, D. 1-Thio-D-glucose. *Methods Carbohydr. Chem.* **2**, 433–437 (1963).
57. Classon, B., Liu, Z. & Samuelson, B. New halogenation reagent systems useful for the mild one-step conversion of alcohols into iodides or bromides. *J. Org. Chem.* **53**, 6126–6130 (1998).
58. Dong, H., Pei, Z., Angelin, M., Byström, S. & Ramström, O. Efficient synthesis of β -D-mannosides and β -D-talosides by double parallel or double serial inversion. *J. Org. Chem.* **72**, 3694–3701 (2007).
59. Horton, D., Wolfrom, M. L. & Thiosugars, I. Synthesis of derivatives of 2-amino-2-deoxy-1-thio-D-glucose. *J. Org. Chem.* **27**, 1794–1800 (1962).
60. Johnston, B. D. & Pinto, B. M. Synthesis of thio-linked disaccharides by 1 \rightarrow 2 intramolecular thioglycosyl migration: oxacarbenium versus episulfonium ion intermediates. *J. Org. Chem.* **65**, 4607–4617 (2000).
61. Otwinowski, Z. & Minor, W. A processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1996).
62. Marshudov, G. N. P. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr.* **D67**, 355–367 (2011).
63. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr.* **D66**, 213–221 (2010).
64. McRee, D. E. XtalView/Xfit - a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165 (1996).
65. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of coot. *Acta Crystallogr.* **D66**, 486–501 (2010).
66. Winn, M. D., Isupov, M. N. & Marshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr.* **D57**, 122–133 (2001).
67. Painter, J. & Merritt, E. A. A molecular viewer for the analysis of TLS rigid-body motion in macromolecules. *Acta Crystallogr.* **D61**, 465–471 (2005).
68. Brünger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475 (1992).
69. Brünger, A. T. *X-PLOR Version 3.851: a system for crystallography and NMR*. (Yale University Press, New Haven, CT, 1992).
70. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99 (1963).
71. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK—a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
72. Lau, E. & Bacic, A. Capillary gas chromatography of partially methylated alditol acetates on a high polarity, cross-linked, fused silica BPX70. *Column. J. Chrom.* **637**, 100–103 (1993).
73. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
74. Kirschner, K. N. et al. GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.* **29**, 622–655 (2008).
75. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
76. Kale, L. et al. NAMD2: greater scalability for parallel molecular dynamics. *Comput. Phys.* **151**, 283–312 (1999).
77. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *Comput. Chem.* **26**, 1781–1802 (2005).
78. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *Comput. Phys.* **23**, 327–341 (1977).
79. Nosé, S. J. A unified formulation of the constant temperature molecular dynamics methods. *Chem. Phys.* **81**, 511–519 (1984).
80. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).
81. Laio, A., VandeVondele, J. & Rothlisberger, U. A. Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations. *J. Chem. Phys.* **116**, 6941–6947 (2002).
82. Car, R. & Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **55**, 471–2474 (1985).
83. Troullier, N. & Martins, J. L. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B* **43**, 1993–2006 (1991).
84. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
85. Barducci, A., Bonomi, M. & Parrinello, M. Metadynamics. *WIREs Comput. Mol. Sci.* **1**, 826–843 (2011).
86. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl Acad. Sci. USA* **99**, 12562–12566 (2002).
87. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100**, 020603 (2008).
88. MacKerell, A. D. Jr. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).

89. MacKerell, A. D. Jr., Feig, M. & Brooks, I. I. C. L. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400–1415 (2004).
90. Guvench, O. et al. CHARMM additive all-atom force field for carbohydrate derivatives and their utility in polysaccharide and carbohydrate-protein modeling. *J. Chem. Theor. Comp* **7**, 3162–3180 (2011).
91. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**, 727–748 (1997).
92. Bakan, A., Meireles, L. M. & Bahar, I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* **27**, 1575–1577 (2011).
93. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
94. Madadkar-Sobhani, A. & Guallar, V. PELE web server: atomistic study of biomolecular systems at your fingertips. *Nucleic Acids Res.* **41**, W322–W328 (2013).
95. Cossins, B. P., Hosseini, A. & Guallar, V. Exploration of protein conformational change with PELE and metadynamics. *J. Chem. Theory. Comput.* **8**, 959–965 (2012).
96. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. & Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **105**, 6474–6487 (2001).
97. Hernández-Ortega, A. et al. Substrate diffusion and oxidation in GMC oxidoreductases: an experimental and computational study on fungal aryl-alcohol oxidase. *Biochem. J.* **436**, 341–350 (2011).
98. Lucas, F. et al. Molecular determinants for selective C25-hydroxylation of vitamins D2 and D3 by fungal peroxxygenases. *Catal. Sci. Technol.* **6**, 288–295 (2016).
99. Gygli, G., Lucas, M. F., Guallar, V. & van Berkel, W. J. H. The ins and outs of vanillyl alcohol oxidase: Identification of ligand migration paths. *PLoS Comput. Biol.* **13**, e1005787 (2017).

Acknowledgements

We thank M. Raab (Slovak Academy of Sciences), B.J. Smith (La Trobe University) and G.B. Fincher (University of Adelaide) for interest in this research. H. Tong (Advanced Photon Source), N. Matsugaki and S. Wakatsuki (Photon Factory), and F. Pettolino (University of Melbourne) are thanked for advice and S. Pradeau for technical assistance. Funding was supported by the Huaiyin Normal University and the Australian Research Council (DP120100900) to M.Hrmova., the Australian Synchrotron (MX1 and MX2 beamlines) to M.Hrmova. and V.A.S., the Advanced Photon Source (14-ID-B beamline) to J.N.V. and M.Hrmova., the Photon Factory (BL5 beamline) to V.A.S., the Suranaree University of Technology and the Thailand Research Fund (BRG5980015) to J.R.K.C. and S.L., Generalitat de Catalunya (Commissioner for Universities and Research, Department of Innovation, Universities and Enterprise) and the European Union through a Beatriu de Pinós fellowship (BP-B 2013) to M.A.-P., Generalitat de Catalunya (SGR2017–1189), MICINN (CTQ2017–85496-P) and Spanish Structures of Excellence María de Maeztu (MDM-2017–0767) to C.R., Spanish Ministerio de Ciencia e Innovación to J.M.L. (CTQ2017–83745-P) and J.-D.M. (CTQ2017–87889-P), and Generalitat de Catalunya (SGR2017–13234) to J.-D.M. L.M. thanks to the Universitat Autònoma de Barcelona Talent Program, F.M. to CONICYT (PFCHA/Doctorado Becas CHILE/2012

72130118) and J.R.-G. to Generalitat de Catalunya and the European Social Fund (2017FL_B2_00168). S.F. was supported by Glyco@Alps (ANR-15-IDEX-02), Carnot Institut PolyNat, Labex ARCANE and CBH-EUR-GS (ANR-17-EURE-0003), and thanks to Chimie Moléculaire de Grenoble for access to facilities. M.A.-P. and C.R. acknowledge MareNostrum and Minotauro and BSC-CNS (RES-QCM-2017–2–001) resources. Use of synchrotrons was supported by the Australian Synchrotron Research Program, which is funded by the Commonwealth of Australia under the Major National Research Facilities Program.

Author contributions

S.F. synthesised thio-oligosaccharides; M.Hrmova. performed enzyme inhibition kinetics; M.Hijnen. conducted SPR; S.L. and J.R.K.C. carried out mutagenesis; S.L. and M.H. performed crystallisation; V.A.S., J.N.V. and M.Hrmova. collected X-ray data; V.A.S., S.L., A.P. and M.Hrmova. refined crystal structures; M.Hrmova. executed GC/MS and HPLC; A.A. and J.J.-B. performed NMR spectroscopy; I.T. conducted initial MD calculations; M.A.-P. and C.R. performed ab initio QM/MM metadynamics and conformational FEL calculations; F.M., L.T.-S., J.-E.S.-A., J.R.-G., J.M.L., J.-D.M. and L.M. implemented docking, classical MD simulations, GaudiMM and PELE calculations; M.Hrmova. wrote the manuscript with contributions from co-authors; M.Hrmova. oversaw the project.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-019-09691-z>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Journal peer review information: *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© Crown 2019

Research Article

การตกผลึกและการวิเคราะห์การเลี้ยวเบนรังสีเอกซ์ของเอนไซม์ Os4BGlu18 จากข้าว

Crystallization and X-ray diffraction analysis of rice Os4BGlu18

สุภาภรณ์ ไบยา^{1*} สติลา เพ็งไธสง² และเจมส์ เกตุทัต คาร์นส์²

Supaporn Baiya^{1*} Salila Pengthaisong² and James Ketudat Cairns²

¹คณะวิทยาศาสตร์ ศรีราชา มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตศรีราชา อำเภอศรีราชา จังหวัดชลบุรี 20230

¹Faculty of Science at Sriracha, Kasetsart University, Sriracha Campus, Sriracha, Chonburi 20230

²สาขาวิชาเคมี สำนักวิชาวิทยาศาสตร์ มหาวิทยาลัยเทคโนโลยีสุรนารี อำเภอเมือง จังหวัดนครราชสีมา 30000

²School of Chemistry, Institute of Science, Suranaree University of Technology, Muang, Nakhon Ratchasima 30000

*E-mail: supaporn.bai@ku.ac.th

Received: 16/07/2017; Accepted: 3/10/2017

บทคัดย่อ

เอนไซม์ Os4BGlu18 จากข้าวเป็นเอนไซม์ในกลุ่มมอโนลิแกนด์โอลิโกเปปไทด์-กลูโคซิเดส ที่สามารถย่อยสลายในกลุ่มมอโนลิแกนด์โอลิโกไซไคลด์ยับยั้งศัตรูพืช ซึ่งประกอบด้วย coniferin, syringin, และ *p*-coumatyl alcohol glucoside ได้อย่างมีประสิทธิภาพมากกว่ายับยั้งศัตรูพืชจากธรรมชาติชนิดอื่น งานวิจัยนี้ผลิตเอนไซม์ Os4BGlu18 ใน *Escherichia coli* ด้วยการรวมเข้ากับ thioredoxin-His₆ tag และทำให้บริสุทธิ์ด้วยคอลัมน์ immobilized metal affinity chromatography (IMAC) มีการตัดปลายแอมิโนของเอนไซม์ Os4BGlu18 ที่มี thioredoxin-His₆ tag ออกและทำให้บริสุทธิ์อีกครั้งด้วยคอลัมน์ IMAC การตกผลึกเอนไซม์ Os4BGlu18 ด้วยวิธี microbatch ประสบความสำเร็จโดยพบผลึกรูปร่างกลมยาวจำนวนมาก เมื่อใช้สารตกผลึกที่ประกอบด้วย Tris-HCl 0.1 โมลาร์, pH 8.5 และ polyethylene glycol 3350 25% ที่อุณหภูมิ 15 องศาเซลเซียสเป็นเวลา 3 สัปดาห์ ผลึกของเอนไซม์ Os4BGlu18 แบบอิสระและแบบที่รวมอยู่กับ 2-deoxy-2-fluoroglucose (G2F) สามารถเลี้ยวเบนรังสีเอกซ์ที่ความยาวคลื่น 1.0 อังสตรอม และใช้เครื่องตรวจจับ ADSC Quantum 315 CCD ได้ด้วยความละเอียด 6.5 และ 7.0 อังสตรอม ตามลำดับ การวิเคราะห์

ทางด้านผลึกศาสตร์ของเอนไซม์ Os4BGlu18 อีสระพบว่ามิตซ์นีออร์โธโรมบิกสเปซกลุ่ม $P2_12_12_1$ และขนาดของยูนิตเซลล์มีค่า a เท่ากับ 163.8 อังสตรอม, b เท่ากับ 167.4 อังสตรอม และ c เท่ากับ 240.2 อังสตรอม สำหรับผลึกของเอนไซม์ที่รวมอยู่กับ G2F พบว่ามีมิตซ์นีเตตราโกนัลสเปซกลุ่ม $P4_22_2$ และขนาดของยูนิตเซลล์มีค่า a เท่ากับ 164.5 อังสตรอม, b เท่ากับ 164.5 อังสตรอม และ c เท่ากับ 241.0 อังสตรอม

คำสำคัญ: มอนอลิกลินอลบีตา-กลูโคซิเดส, 2-ดีออกซี-2-ฟลูออโรกลูโคส, การตกผลึกด้วยวิธีไมโครแบทช์

Abstract

Rice Os4BGlu18 was proposed to be a monolignol β -glucosidase and has been shown to hydrolyze the monolignol glucoside substrates, coniferin, syringin, and *p*-coumatyl alcohol glucoside, with higher catalytic efficiency than other natural substrates. The Os4BGlu18 protein was produced as a thioredoxin-His₆ fusion protein in *Escherichia coli* and purified by immobilized metal affinity chromatography (IMAC). The N-terminal thioredoxin-His₆ tag was cutoff and the protein was further purified by a second IMAC step. Screening in microbatch crystallization showed that Os4BGlu18 could be successfully crystallized to form many crystals with rod shape. The rod shaped crystals were grown in a precipitant solution containing 0.1 M Tris-HCl, pH 8.5, and 25% polyethylene glycol 3350 after 3 weeks of incubation at 15°C. The crystals of free Os4BGlu18 and its complex with 2-deoxy-2-fluoroglucose (G2F) could diffract X-rays at a 1.0 Å wavelength x-ray beam and an ADSC Quantum 315 CCD detector with nominal resolutions of 6.5 and 7.0 Å, respectively. The crystallographic analysis of free Os4BGlu18 showed that it was indexed to the orthorhombic $P2_12_12_1$ space group, and had the unit cell parameters: a = 163.8 Å, b = 167.4 Å, c = 240.2 Å. For the G2F complex, it was indexed to the tetragonal $P4_22_2$ space group, and had the unit cell parameters: a = 164.5 Å, b = 164.5 Å, c = 241.0 Å.

Keywords: monolignol β -glucosidase, 2-deoxy-2-fluoroglucose, microbatch crystallization

บทนำ

เอนไซม์บีตา-กลูโคซิเดสจากพืชจัดอยู่ในกลุ่มเอนไซม์ไกลโคไซด์ไฮโดรเลส โดยเป็นเอนไซม์เร่งปฏิกิริยาการสลายตัวด้วยน้ำที่ตำแหน่งบีตา-ไกลโคซิดิกของสารประกอบหมู่แอริลและแอลคิลบีตา-ดี-กลูโคไซด์และสารประกอบกลูโคโอลิโกแซ็กคาไรด์ เพื่อปลดปล่อยโมเลกุลกลูโคสและอะไกลโคนออกมา เอนไซม์ประเภทนี้พบในสิ่งมีชีวิตหลายชนิดได้แก่ อาร์เคีย แบคทีเรีย เห็ดรา พืช และสัตว์ ในพืชมีการรายงานว่าเอนไซม์บีตา-กลูโคซิเดสมีหน้าที่สำคัญหลายอย่าง เช่น การย่อยโอลิโกแซ็กคาไรด์ที่ผนังเซลล์ของพืช การปลดปล่อยสารป้องกันตนเอง

ของพืชจากรูปที่ไม่สามารถทำงานได้ให้กลายเป็นรูปที่สามารถทำงานได้ การควบคุมสารในกลุ่มฮอร์โมนพืช การเกิดปฏิกิริยาการแยกสลายด้วยน้ำกับสารในกลุ่มแอลคาลอยด์แอนาบอลิกอินเทอร์มีเดียตกลูโคไซด์สำหรับการสร้างมอโนเทอร์พีนแอลคาลอยด์ในวิถีเมแทบอลิซึม การปลดปล่อยสารระเหยง่ายภายในเซลล์พืช รวมถึงการปล่อยสารในกลุ่มมอโนลิกนอลซึ่งเป็นหน่วยการสร้างลิกนินในพืช (Opassiri et al., 2006; Ketudat Cairns & Esen, 2010)

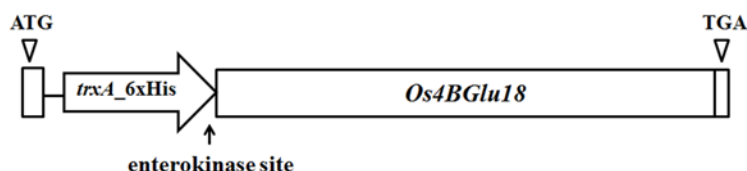
ในธรรมชาติสารกลุ่มมอโนลิกนอลกลูโคไซด์ที่ประกอบด้วย coniferin, syringin และ *p*-coumaryl alcohol glucoside จัดเป็นสารประกอบสำหรับการเคลื่อนที่หรือสำหรับการเก็บสะสมไว้ของมอโนลิกนอลเพื่อใช้ในการสร้างลิกนิน เมื่อพืชต้องการสร้างลิกนิน เอนไซม์มอโนลิกนอลบีตา-กลูโคซิเดสที่จำเพาะต่อการปลดปล่อยสารกลุ่มมอโนลิกนอลกลูโคไซด์จะเข้ามาตัดเอากลูโคสออกทำให้ได้มอโนลิกนอลที่สามารถนำไปสร้างลิกนินได้ต่อไป ความจำเพาะต่อสารตั้งต้นของเอนไซม์มอโนลิกนอลบีตา-กลูโคซิเดสเคยมีการศึกษาในพืชหลายชนิดเช่น ต้นสน ถั่ว ลูกไก่ อะราบิดอพซิส และข้าว (Hösel et al., 1978; Dharmawardhana & Ellis, 1998; Escamilla-Treviño et al., 2006; Baiya et al., 2014) จากการเทียบลำดับความเหมือนของกรดแอมิโนของเอนไซม์บีตา-กลูโคซิเดสในข้าวและอะราบิดอพซิส พบว่า Os4BGlu14, Os4BGlu16, และ Os4BGlu18 ในข้าวจัดอยู่ในกลุ่มเดียวกันกับมอโนลิกนอลบีตา-กลูโคซิเดส (BGlu45 และ BGlu46) ในอะราบิดอพซิส (Opassiri et al., 2006) เมื่อทำการศึกษาความจำเพาะต่อสารตั้งต้นของเอนไซม์ Os4BGlu16 และ Os4BGlu18 พบว่าเอนไซม์ทั้งสองสามารถทำปฏิกิริยาการสลายด้วยน้ำกับสารกลุ่มมอโนลิกนอลกลูโคไซด์ทั้ง 3 ชนิดได้อย่างมีประสิทธิภาพยิ่งกว่าสารตั้งต้นจากธรรมชาติชนิดอื่น ๆ แสดงให้เห็นถึงหน้าที่ที่จำเพาะของเอนไซม์ทั้ง 2 ชนิดนี้ต่อกระบวนการสร้างลิกนินภายในเซลล์พืช (Baiya et al., 2014)

เพื่อให้เข้าใจถึงโครงสร้างและหน้าที่การทำงานของเอนไซม์มอโนลิกนอลบีตา-กลูโคซิเดสให้มากขึ้น คณะผู้วิจัยจึงเลือกเอนไซม์ Os4BGlu18 มาทำการดักผลึกเพื่อให้เหมาะต่อการศึกษาโครงสร้างโปรตีนด้วยเทคนิคทางผลึกวิทยา

วิธีการทดลอง

1. การผลิตและการแยกเอนไซม์ Os4BGlu18 ให้บริสุทธิ์

ดีเอ็นเอของเอนไซม์ Os4BGlu18 ขนาด 1,500 กิโลเบสแพร์ ได้มาจากปฏิกิริยาถูกโซ่พอลิเมอร์เรส (PCR) ของ cDNA จากข้าวหอมดอกมะลิ 105 อายุ 7 วัน และโคลนเข้าไปในเวกเตอร์ pET32a(+) โดยด้านปลายแอมิโนของเอนไซม์ถูกเชื่อมเข้ากับ thioredoxin-His₆ tag ซึ่งมีตำแหน่งตัดจำเพาะสำหรับเอนไซม์ enterokinase ดังแสดงในรูปที่ 1 จากนั้นดีเอ็นเอสายผสมนี้ถูกย้ายเข้าไปในแบคทีเรียชนิด *Escherichia coli* สายพันธุ์ Origami (DE3) ตามการรายงานของ Baiya et al. (2014)



รูปที่ 1. แผนภาพดีเอ็นเอสายผสมของเอนไซม์ Os4BGlu18

การชักนำให้เอนไซม์ Os4BGlu18 แสดงออกทำได้โดยเลี้ยงแบคทีเรียที่มีดีเอ็นเอสายผสม pET32a/Os4BGlu18 อยู่ในเซลล์ด้วยอาหารชนิด Luria broth (LB) ซึ่งเติมแอมพิซิลลินความเข้มข้น 50 ไมโครกรัมต่อมิลลิลิตร กานามัยซินความเข้มข้น 15 ไมโครกรัมต่อมิลลิลิตร และเตตราไซคลีนความเข้มข้น 12.5 ไมโครกรัมต่อมิลลิลิตร ลงไปในอาหารและเลี้ยงเซลล์ภายในเครื่องเขย่าชนิดควบคุมอุณหภูมิที่ 37 องศาเซลเซียส ความเร็ว 200 รอบต่อนาที จนกระทั่งวัดค่าการดูดกลืนแสงที่ 600 นาโนเมตรได้ประมาณ 0.6 จึงทำการเหนี่ยวนำให้แบคทีเรียสร้างเอนไซม์ Os4BGlu18 ด้วย isopropyl β -D-thiogalactopyranoside (IPTG) ความเข้มข้น 0.1 มิลลิโมลาร์ ที่อุณหภูมิ 18 องศาเซลเซียส เป็นเวลา 16-18 ชั่วโมง หลังจากนั้นจึงทำการเก็บเซลล์แบคทีเรียแยกออกจากอาหาร LB โดยการปั่นเหวี่ยงที่ความเร็ว 4,500 รอบต่อนาที เป็นเวลา 15 นาที แล้วแช่แข็งไว้ที่อุณหภูมิ -80 องศาเซลเซียส เป็นเวลา 1 คืน ทำการสลายผนังเซลล์ของแบคทีเรียเพื่อให้เอนไซม์ Os4BGlu18 ที่ถูกสร้างภายในไซโทพลาซึมของเซลล์แบคทีเรียออกมาด้วยสารสกัดบัฟเฟอร์ที่ประกอบด้วย Tris-HCl pH 8.0 ความเข้มข้น 20 มิลลิโมลาร์, lysozyme ความเข้มข้น 200 ไมโครกรัมต่อมิลลิลิตร, 1% (v/v) Triton-X 100, phenylmethylsulfonylfluoride ความเข้มข้น 1 มิลลิโมลาร์, Dnase I ความเข้มข้น 5 ไมโครกรัมต่อมิลลิลิตร และ soy bean trypsin inhibitor ความเข้มข้น 0.1 มิลลิกรัมต่อมิลลิลิตร ในอัตราส่วน 5 มิลลิลิตรของสารสกัดบัฟเฟอร์ต่อ 1 กรัมของเซลล์แบคทีเรีย หลังจากนั้นจึงบ่มไว้ที่อุณหภูมิห้องเป็นเวลา 30 นาทีแล้วจึงแยกเอนไซม์ Os4BGlu18 ออกจากเซลล์ที่แตกนี้ด้วยการปั่นเหวี่ยงที่ความเร็ว 12,000 รอบต่อนาที อุณหภูมิ 4 องศาเซลเซียส เป็นเวลา 20 นาที

งานวิจัยในครั้งนี้ได้ปรับเปลี่ยนวิธีการแยกเอนไซม์ Os4BGlu18 ให้บริสุทธิ์ต่างออกไปจากเดิม (Baiya et al., 2014) เพื่อให้ได้เอนไซม์ที่เหมาะสมสำหรับการตกผลึก ขั้นตอนที่ 1 นำเอนไซม์ที่ได้จากการปั่นแยกออกจากเซลล์แบคทีเรียในขั้นตอนก่อนหน้านี้มาแยกให้บริสุทธิ์ด้วยคอลัมน์ immobilized metal affinity chromatography (IMAC) ที่ชาร์จให้มีประจุด้วย Co^{2+} โปรตีนปนเปื้อนที่ไม่สามารถจับกับ IMAC เรซินได้ถูกชะล้างออกไปจากคอลัมน์ด้วย imidazole ความเข้มข้น 0-10 มิลลิโมลาร์ ในบัฟเฟอร์ 20 มิลลิโมลาร์ Tris-HCl, pH 8.0 ที่มีเกลือ NaCl อยู่ 150 มิลลิโมลาร์ (บัฟเฟอร์ปรับสมดุล) จากนั้นเอนไซม์ Os4BGlu18 ที่จับอยู่กับ IMAC เรซินถูกชะออกมาด้วย imidazole ความเข้มข้น 250 มิลลิโมลาร์ ทำการตรวจสอบเอนไซม์ Os4BGlu18 ขนาดประมาณ 75 กิโลดาลตัน

โดยเทคนิค Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) แล้วจึงรวมส่วนของเอนไซม์ Os4BGlu18 ที่ต้องการเข้าด้วยกันแล้วกำจัด imidazole ออกไปโดยการกรองด้วย 30 kDa Amicon® Ultra-15 centrifugal filters ด้วยบัฟเฟอร์ปรับสมดุล ขั้นตอนที่ 2 ตัดส่วนปลายแอมิโน thioredoxin-His₆ tag ออกโดยใช้เอนไซม์ enterokinase ในอัตราส่วน 1 ไมโครกรัม enterokinase ต่อ 1 มิลลิกรัมเอนไซม์ Os4BGlu18 ทำการบ่มเอนไซม์ Os4BGlu18 กับเอนไซม์ enterokinase ที่อุณหภูมิ 23 องศาเซลเซียสเป็นเวลา 16 ชั่วโมง เมื่อบ่มเสร็จแล้วจึงทำการแยกเอนไซม์ Os4BGlu18 ออกจาก thioredoxin-His₆ tag อีกครั้งด้วยคอลัมน์ IMAC เนื่องจากเอนไซม์ Os4BGlu18 ถูกตัดเอาส่วนที่สามารถจับกับ IMAC เรซินออกไปแล้วทำให้การทำบริสุทธิ์ในครั้งนี้เอนไซม์ Os4BGlu18 ที่ต้องการจะถูกชะออกมาก่อนในขั้นตอนการล้างด้วยบัฟเฟอร์ปรับสมดุลที่มี imidazole อยู่ 0-5 มิลลิโมลาร์ ทำการตรวจสอบหาเอนไซม์ Os4BGlu18 ขนาดประมาณ 56 กิโลดาลตันอีกครั้งด้วยเทคนิค SDS-PAGE และทดสอบเพื่อยืนยันว่าเอนไซม์ Os4BGlu18 ที่ได้นี้ยังสามารถทำหน้าที่ได้ด้วยการทดสอบกับสารตั้งต้น *p*-nitrophenol β -D-glucopyranoside ที่ความเข้มข้น 1 มิลลิโมลาร์ ใน sodium acetate ความเข้มข้น 100 มิลลิโมลาร์ pH 5.0 จากนั้นจึงทำการกำจัด imidazole ออกไปอีกครั้งด้วย 10 kDa Amicon® Ultra-15 centrifugal filters

2. การตกผลึกเอนไซม์ Os4BGlu18

ก่อนทำการตกผลึกเอนไซม์ Os4BGlu18 ได้ทำการปรับความเข้มข้นของโปรตีนให้อยู่ที่ 3 มิลลิกรัมต่อมิลลิลิตรและกำจัดจุลินทรีย์ ฟัน และเอนไซม์บางส่วนที่อาจตกตะกอน ซึ่งปนเปื้อนอยู่นี้ออกไปด้วยการกรองผ่าน Ultrafree-MC filter ขนาด 0.22 ไมโครเมตร จากนั้นจึงเริ่มตกผลึกด้วยการใช้ชุดสารสำหรับตกผลึกโปรตีนชนิด HR2-130 (Hampton Research, CA, USA), HR2-134 (Hampton Research), และ Morpheus™46 (Laboratory of Molecular Biology, Cambridge, UK) เทคนิคที่ใช้ในการตกผลึกเป็นแบบ Microbatch โดยใช้เพลทขนาด 96 หลุม เริ่มจากการเติม 100% Paraffin oil ลงไปในหลุม 10 ไมโครลิตร จากนั้นใส่สารสำหรับตกผลึก 1 ไมโครลิตร ลงไปบริเวณก้นหลุม แล้วจึงใส่เอนไซม์ Os4BGlu18 (ความเข้มข้น 3 มิลลิกรัมต่อมิลลิลิตร) 2 ไมโครลิตร ลงไปที่ก้นหลุม ให้รวมเป็นหยดเดียวกันกับสารตกผลึก ปิดฝาเพลทเพื่อป้องกันฝุ่นและเก็บไว้ในกล่องพลาสติกที่มีฟองน้ำเปียกชื้นอยู่ แล้วจึงนำไปบ่มไว้ที่ตู้บ่มอุณหภูมิ 15 องศาเซลเซียส ตรวจสอบผลึกโปรตีนที่จะเกิดขึ้นด้วยกล้องจุลทรรศน์แบบสเตอริโอ

3. การเลี้ยงเบเนรงส์อี็กซ์และการวิเคราะห์ด้านผลึกศาสตร์

ผลึกที่เกิดขึ้นถูกเกี่ยวขึ้นมาแช่ในสารป้องกันการแช่แข็งซึ่งประกอบด้วย glycerol 18% (v/v) ในสารตกผลึกที่พบผลึกของเอนไซม์ Os4BGlu18 โดยเพิ่มความเข้มข้นจากเดิม 15% ผลึกของเอนไซม์ Os4BGlu18 บางส่วนจะถูกนำมาแช่กับสารยับยั้งการทำงานของเอนไซม์นี้ ได้แก่ 2-deoxy-2-fluoroglucose (G2F) ที่ความเข้มข้น 1 และ 10 มิลลิโมลาร์ เป็นเวลา 20 นาที แล้วจึงเกี่ยวผลึกไปแช่แข็งในไนโตรเจนเหลว ผลึกของเอนไซม์

Os4BGlu18 และผลึกที่แช่ใน G2F ถูกนำไปทดสอบการเลี้ยวเบนของรังสีเอ็กซ์ที่ National Synchrotron Radiation Research Center (NSRRC, Hsinchu, Taiwan) ที่มีความยาวคลื่นในการทดสอบ 1 อังสตรอม ที่ BL15A beamline และตรวจจับด้วยเครื่อง ADSC Quantum 315 CCD detector ผลึกถูกรักษาความเย็นไว้ที่อุณหภูมิ 110 องศาเซลเซียสด้วยแก๊สไนโตรเจนในระหว่างที่ทำการทดสอบ ชุดข้อมูลที่เก็บได้ทั้งหมดถูกนำมาวิเคราะห์ด้วยโปรแกรม HKL-2000 package (Otwinowski and Minor, 1997) และ CCP4 suite

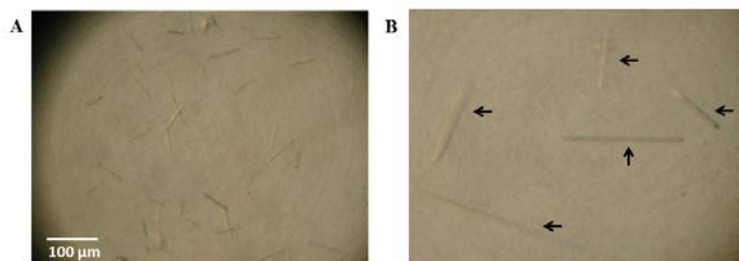
ผลการทดลอง

การผลิตเอนไซม์ Os4BGlu18 ในแบคทีเรีย *E. coli* สายพันธุ์ origami (DE3) โดยการเหนี่ยวนำด้วย IPTG 0.1 มิลลิโมลาร์ ที่อุณหภูมิ 18 องศาเซลเซียสเป็นเวลา 16-18 ชั่วโมง สามารถผลิตเอนไซม์ Os4BGlu18 ในปริมาณ 4 มิลลิกรัม ตามการรายงานของ Baiya et al. (2014) เอนไซม์ Os4BGlu18 ที่ผ่านการทำให้บริสุทธิ์ด้วย 2 ขั้นตอน ด้วยคอลัมน์ IMAC ทำให้ได้โปรตีนที่มีความบริสุทธิ์มากกว่า 98% โดยการประเมินจาก SDS-PAGE ตามรูปที่ 2 และมีขนาดประมาณ 56 กิโลดาลตัน



รูปที่ 2. SDS-PAGE ของเอนไซม์ Os4BGlu18 หลังจากผ่านการทำให้บริสุทธิ์ด้วยคอลัมน์ IMAC ในขั้นตอนที่ 2 ช่องที่ 1 คือโปรตีนมาตรฐาน (หน่วยกิโลดาลตัน) และช่องที่ 2 คือเอนไซม์ Os4BGlu18 ขนาด 56 กิโลดาลตัน

การตกผลึกเอนไซม์ Os4BGlu18 โดยใช้ชุดสารสำหรับตกผลึกโปรตีน 3 ชนิดคือ HR2-130, HR2-134, และ Morpheus™46 พบว่า ชุดสารสำหรับตกผลึกโปรตีนชนิด HR2-134 (D11) ซึ่งประกอบด้วย Tris pH 8.5 0.1 M และ PEG 3350 25% (w/v) สามารถตกผลึกเอนไซม์ Os4BGlu18 ในรูปร่างกลมยาวขนาดต่าง ๆ ได้โดยใช้เวลาในการบ่ม 3 อาทิตย์ ดังแสดงในรูปที่ 3



รูปที่ 3. ผลึกรูปร่างกลมยาวตามลูกศรของเอนไซม์ Os4BGlu18 ที่พบในสารตกผลึกโปรตีน
ชนิด HR2-134 (D11) ภาพถ่ายมุมกว้าง (A) และภาพถ่ายมุมแคบ (B)

ผลึกเอนไซม์ Os4BGlu18 ถูกนำไปวิเคราะห์การเลี้ยวเบนด้วยรังสีเอ็กซ์ ข้อมูลที่เก็บมาได้จากการเลี้ยวเบนรังสีเอ็กซ์ของผลึกเอนไซม์ Os4BGlu18 และผลึกเอนไซม์ Os4BGlu18 ที่แช่ใน G2F ความเข้มข้น 10 มิลลิโมลาร์ ถูกนำมาวิเคราะห์ดังแสดงในตารางที่ 1 แต่ผลึกเอนไซม์ Os4BGlu18 ที่แช่ใน G2F ความเข้มข้น 1 มิลลิโมลาร์ไม่สามารถเลี้ยวเบนรังสีเอ็กซ์ได้ หลังจากวิเคราะห์ข้อมูลแล้วปรากฏว่าผลึกของเอนไซม์ Os4BGlu18 สามารถเลี้ยวเบนรังสีเอ็กซ์ได้ที่ความละเอียด 6.5 องศาจัดอยู่ในกลุ่มดัชนีออร์โธโรมบิกสเปซ $P2_12_1$ ด้วยยูนิตเซลล์พารามิเตอร์ดังนี้ a เท่ากับ 163.8 องศา, b เท่ากับ 167.4 องศา และ c เท่ากับ 240.2 องศา ส่วนผลึกของโปรตีน Os4BGlu18 ที่แช่ใน G2F ความเข้มข้น 10 มิลลิโมลาร์ สามารถเลี้ยวเบนรังสีเอ็กซ์ได้ที่ความละเอียด 7.0 องศา อยู่ในกลุ่มดัชนีเตตราโกนัลสเปซ $P4_22_2$ และยูนิตเซลล์พารามิเตอร์มีขนาดดังนี้ a เท่ากับ 164.5 องศา, b เท่ากับ 164.5 องศา และ c เท่ากับ 241.0 องศา

ตารางที่ 1. ข้อมูลการวิเคราะห์การเลี้ยวเบนรังสีเอ็กซ์ของผลึกโปรตีน Os4BGlu18 และผลึกโปรตีน Os4BGlu18 ที่ถูกแช่ใน G2F

	Os4BGlu18	Os4BGlu18 ที่แช่ใน G2F
ความละเอียด (องศา)	6.5	7.0
ระบบกลุ่มสเปซ	$P2_12_1$	$P4_22_2$
ขนาดของยูนิตเซลล์ (องศา)	a = 163.8	a = 164.5
	b = 167.4	b = 164.5
	c = 240.2	c = 241.0
	$\alpha = 90.0$	$\alpha = 90.0$
	$\beta = 90.0$	$\beta = 90.0$
	$\gamma = 90.0$	$\gamma = 90.0$

วิจารณ์และสรุปผลการทดลอง

เอนไซม์ Os4BGlu18 ถูกโคลนขึ้นมาโดยรวมเอาส่วนปลายเอมิโนเชื่อมเข้ากับ thioredoxin-His₆ tag เพื่อความสะดวกในการทำให้บริสุทธิ์ด้วยคอลัมน์ IMAC อย่างไรก็ตามส่วนปลายเอมิโนนี้อาจรบกวนการตกผลึกของโปรตีนได้ ทำให้ต้องมีการตัดเอาส่วนปลายนี้ออกโดยใช้เอนไซม์ enterokinase ซึ่งวิธีนี้ได้รับผลสำเร็จอย่างดีสำหรับเอนไซม์บีตา-กลูโคซิเดสในโอโซไซม์อื่น ๆ จากข้าว เช่น Os3BGlu6, Os3BGlu7, และ Os7BGlu26 (Chuenchor et al., 2008; Seshadri et al., 2009; Tankrathok et al., 2013) จากผลการผลิตและทำให้เอนไซม์ Os4BGlu18 บริสุทธิ์ด้วยคอลัมน์ IMAC 2 ขั้นตอน ทำให้สามารถผลิตเอนไซม์ได้ปริมาณมากพอสำหรับการตกผลึก เนื่องจากการศึกษาโครงสร้างของเอนไซม์โดยเทคนิคผลึกวิทยา จำเป็นต้องใช้เอนไซม์ในปริมาณที่สูง จึงจะทำให้เอนไซม์กลายเป็นผลึกได้ แตกต่างจากการศึกษาหน้าที่ของเอนไซม์โดยการตรวจหาความจำเพาะของเอนไซม์ต่อสารตั้งต้นด้วยการวัดอัตราเร็วของปฏิกิริยา ซึ่งคุณสมบัติในการเป็นตัวเร่งที่มีประสิทธิภาพสูงของเอนไซม์ จึงไม่จำเป็นต้องใช้เอนไซม์ในปริมาณมากก็สามารถตรวจหาอัตราเร็วของปฏิกิริยาได้ และด้วยความบริสุทธิ์ของเอนไซม์ที่มากถึง 98% จากการประเมินด้วยเทคนิค SDS-PAGE จึงทำให้การตกผลึกของเอนไซม์ Os4BGlu18 ประสบผลสำเร็จ นอกจากนี้ชุดตกผลึกโปรตีนที่นำมาใช้มีความหลากหลายของสารที่ใช้ในการตกผลึก ช่วยให้ผู้วิจัยค้นพบสภาวะที่เหมาะสมที่ทำให้โปรตีนตกผลึกได้ในเวลาอันสั้น โดยพบผลึกของเอนไซม์ Os4BGlu18 หลังจากบ่มที่อุณหภูมิ 15 องศาเซลเซียสเป็นเวลา 3 อาทิตย์ ในสารตกผลึกที่ประกอบด้วย Tris pH 8.5 0.1 M และ PEG 3350 25% (w/v) และเพื่อป้องกันการละลายของผลึกในอุณหภูมิที่สูงกว่า 15 องศาเซลเซียสในระหว่างการเลี้ยวเบนด้วยรังสีเอ็กซ์ ผลึกของเอนไซม์ Os4BGlu18 จะแช่แข็งไว้ที่อุณหภูมิ 110 องศาเคลวินด้วยแก๊สไนโตรเจนและมี glycerol 18% ซึ่งเป็นสารป้องกันการแช่แข็งของผลึกเอนไซม์ ผลึกของเอนไซม์ Os4BGlu18 ที่ทำการแช่ใน G2F สามารถใช้ความเข้มข้นได้สูงถึง 10 มิลลิโมลาร์โดยไม่ทำให้ผลึกเกิดความเสียหาย ในขณะที่ผลึกของเอนไซม์ Os4BGlu18 ที่ทำการแช่ใน G2F ความเข้มข้น 1 มิลลิโมลาร์ซึ่งเป็นความเข้มข้นที่ต่ำกว่า ส่งผลต่อความสามารถในการจับของเอนไซม์กับสารยับยั้งได้ลดลง ทำให้หลังจากวิเคราะห์แล้วไม่พบ G2F ในบริเวณเร่งของเอนไซม์ ผลึกของเอนไซม์ Os4BGlu18 แบบอิสระและที่แช่ใน G2F ความเข้มข้น 10 มิลลิโมลาร์ สามารถเลี้ยวเบนรังสีเอ็กซ์ได้ที่ความละเอียด 6.5 และ 7.0 อังสตรอมตามลำดับ Os4BGlu18 แบบอิสระจัดอยู่ในกลุ่มดัชนีออร์โธโรมบิกสเปซ $P2_12_12_1$ ส่วน Os4BGlu18 ที่แช่ใน G2F จัดอยู่ในกลุ่มดัชนีเตตราโกนัลสเปซ $P4_22_2$ แสดงให้เห็นถึงหน้าที่ของเอนไซม์ Os4BGlu18 ว่ามีการเปลี่ยนแปลงรูปร่างให้เหมาะสมพอดีกับสารตั้งต้นขณะทำปฏิกิริยา ซึ่งสนับสนุนทฤษฎีเหนี่ยวนำให้พอดีของการเกิดปฏิกิริยาระหว่างเอนไซม์กับสารตั้งต้น ดังนั้นวิธีการผลิตโดยทำให้เอนไซม์ Os4BGlu18 บริสุทธิ์และสารตกผลึกที่กล่าวมาข้างต้นเป็นวิธีการที่เหมาะสม ซึ่งทำให้เอนไซม์สามารถตกผลึกและเลี้ยวเบนรังสีเอ็กซ์ได้

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับสนับสนุนทุนวิจัยจากสำนักงานกองทุนสนับสนุนการวิจัย สำนักงานคณะกรรมการการอุดมศึกษา และมหาวิทยาลัยเทคโนโลยีสุรนารี คณะผู้วิจัยขอขอบพระคุณมา ณ ที่นี้ด้วย

เอกสารอ้างอิง

- Baiya, S., Hua, Y., Ekkhara, W. & Ketudat Cairns, J. R. (2014). Expression and enzymatic properties of rice (*Oryza sativa* L.) monolignol beta-glucosidases. *Plant Science*, 227(1), 101-109. doi:10.1016/j.plantsci.2014.07.009
- Chuenchor, W., Pengthaisong, S., Robinson, R. C., Yuvaniyama, J., Oonant, W., Bevan, D. R., et al. (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
- Dharmawardhana, D. P. & Ellis, B. E. (1998). β -Glucosidases and glucosyltransferase in lignifying tissues. *Journal of the American Chemical Society*, 697(6), 76-83. doi:10.1021/bk-1998-0697.ch006
- Escamilla-Treviño, L. L., Chen, W., Card, M. L., Shih, M.-C., Cheng, C. L. & Poulton, J. E. (2006). *Arabidopsis thaliana* β -glucosidases BGLU45 and BGLU46 hydrolyse monolignol glucosides. *Phytochemistry*, 67(15), 1651-1660. doi:10.1016/j.phytochem.2006.05.022
- Hösel, W., Surholt, E. & Borgmann, E. (1978). Characterization of β -glucosidase isoenzymes possibly involved in lignifications from chick pea (*Cicer arietinum* L.) cell suspension culture. *European Journal of Biochemistry*, 84(2), 487-492. doi:10.1111/j.1432-1033.1978.tb12190.x
- Ketudat Cairns, J. R. & Esen, A. (2010). β -Glucosidases. *Cellular and Molecular Life Sciences*, 67(20), 3389-3405. doi:10.1007/s00018-010-0399-2
- Opassiri, R., Pomthong, B., Onkoksoong, T., Akiyama, T., Esen, A. & Ketudat Cairns, J. R. (2006). Analysis of rice glycosyl hydrolase family 1 and expression of Os4bglu12 β -glucosidase. *BMC Plant Biology*, 6(1), 33-51. doi:10.1186/1471-2229-6-33
- Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In C. W. Carter & R. M. Sweet (Eds.), *Methods in Enzymology* (Volume 276) pp. 307-326. New York: Academic Press.
- 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 Physiology*, 151(1), 47-58. doi: 10.1104/pp.109.139436

Tankrathok, A., Iglesias-Fernández, J., Luang, S., Robinson, R. C., Kimura, A., Rovira, C., et al. (2013). Structural analysis and insights into the glycon specificity of the rice GH1 Os7BGlu26 β -D-mannosidase. *Acta Crystallographica Section D Biological Crystallography*, 69(10), 2124-2135. doi:10.1107/ S0907444913020568