THE ROLES OF ACTIVE SITE AMINO ACID

RESIDUES OF *Tx*GH116 IN GLUCOSE

BINDING AND CATALYSIS

Huang Meng

A Thesis Submitted in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy in Biochemistry

Suranaree University of Technology

Academic Year 2020

บทบาทของกรดอะมิโนบริเวณเร่งปฏิกิริยาของ *TxGH116* ที่เกี่ยวข้องกับการจับและทำปฏิกิริยากับกลูโคส

นายหวง เหมิง

THE ROLES OF ACTIVE SITE AMINO ACID RESIDUES OF

*Tx*GH116 IN GLUCOSE BINDING AND CATALYSIS

Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for Degree of Doctor of Philosophy.

Thesis Examining Committee

_____
(Dr. Sakesit Chumnarnsilpa)
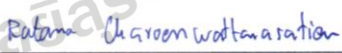
Chairperson

_____
(Prof. Dr. James R. Ketudat-Cairns)

Member (Thesis Advisor)

_____
(Assoc. Prof. Dr. Mariena Ketudat-Cairns)

Member

_____
(Dr. Chomphunuch Songsiriritthigul)
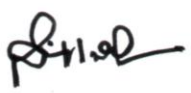
Member

_____
(Dr. Ratana Charoenwattanasatien)

Member

_____
(Assoc. Prof. Dr. Chatchai Jothityangkoon)

Vice Rector for Academic Affairs

and Quality Assurance

_____
(Prof. Dr. Santi Maensiri)

Dean of Institute of Science

หวง เหมิง: บทบาทของกรดอะมิโนบริเวณเร่งปฏิกิริยาของ *Tx*GH116 ที่เกี่ยวข้องกับการจับและทำปฏิกิริยากับกลูโคส (THE ROLES OF ACTIVE SITE AMINO ACID RESIDUES OF *Tx*GH116 IN GLUCOSE BINDING AND CATALYSIS) อาจารย์ที่ปรึกษา : ศาสตราจารย์ ดร.เจมส์ เกตุทัต-คาร์นส์, 160 หน้า.

เอนไซม์ *Tx*GH116 จากเชื้อแบคทีเรีย *Thermoanaerobacterium xylanolyticum* เป็นเบตากลูโคซิเดสที่ทนความร้อนได้ ซึ่งเอนไซม์นี้ถูกศึกษาโครงสร้างเป็นตัวแรกจากตระกูลไกลโคไซด์ไฮโดร-เลส GH116 การอนุรักษ์ของกรดอะมิโนบริเวณเร่งปฏิกิริยาของตระกูล GH116 ในระดับสูงบ่งชี้ถึงความสำคัญของกรดอะมิโนเหล่านี้ ในงานวิจัยนี้ได้ทำการกลายพันธุ์อย่างเป็นระบบของกรดอะมิโนบริเวณเร่งปฏิกิริยาที่น่าจะเกี่ยวข้องกับหน้าที่เร่งปฏิกิริยา เพื่อหาปริมาณการมีส่วนร่วมของแต่ละโซ่ข้างในการเร่งปฏิกิริยา

ประการแรก ทำการวิเคราะห์การทำงานอย่างเป็นระบบของกรดอะมิโนบริเวณเร่งปฏิกิริยาของ *Tx*GH116 ที่ตำแหน่งจับกับน้ำตาลไกลโคน ได้แก่ กรดอะมิโน D452 H507 T591 E730 W732 E777 R786 และ R792 ในการจับซับสเตรตและการเร่งปฏิกิริยา จากจลนพลศาสตร์ของเอนไซม์และซับสเตรท และการวิเคราะห์โครงสร้างของเอนไซม์ ได้อธิบายบทบาทของกรดอะมิโนแต่ละตัวโดยทั่วไปปฏิกิริยาการจับอย่างง่ายกับไกลโคนของซับสเตรท ไม่สามารถอธิบายการเปลี่ยนแปลงการทำงานได้อย่างชัดเมื่อมีการกลายพันธุ์ของกรดอะมิโนที่จับกับกลูโคส และการกระจายประจุ และความเสถียรของกรดอะมิโนที่สภาวะทรานซิชันดูเหมือนจะมีความสำคัญมากกว่า โดยพิจารณาจากผลกระทบต่อค่า k_{cat} มากกว่า K_M

ประการที่สอง ตรวจสอบกรดอะมิโนที่จับกับตัวเร่งปฏิกิริยากรด/เบส และนิวคลีโอไฟล์ การเปรียบเทียบลำดับของกรดอะมิโนเหล่านี้ แสดงให้เห็นว่ารูปแบบ "ทรีโอนีน-นิวคลีโอไฟล์-ไทโรซีน" ของตัวเร่งปฏิกิริยาถูกอนุรักษ์ไว้แบบทั่วไป ซึ่งทรีโอนีนมีพันธะไฮโดรเจนที่เป็นทั้งตัวให้และตัวรับกลูตามีน 727 ในรูปแบบ "กลูตามีน-กรด/เบส-ไทโรซีน" ไม่ได้มีการอนุรักษ์อย่างสูง ในขณะที่ไทโรซีน 523 มีบทบาทสำคัญในการปรับตัวเร่งปฏิกิริยากรด/เบสในการให้โปรตอน เอนไซม์กลายพันธุ์ T450A และ Y790F ที่เกี่ยวข้องกับตัวเร่งปฏิกิริยานิวคลีโอไฟล์สามารถทำงานได้ลดลงอย่างมากเมื่อเทียบกับเอนไซม์ดั้งเดิม เอนไซม์กลายพันธุ์ T450A มีพีเอชโปรไฟล์ที่แคบและมีค่าพีเอชที่เหมาะสมกับการทำงานเป็นกลาง แต่ Y790F มีพีเอชโปรไฟล์ที่กว้าง เอนไซม์กลายพันธุ์ Y523F Q727A และ Q727E ที่เกี่ยวข้องกับตัวเร่งปฏิกิริยากรด/เบสมีพีเอชโปรไฟล์ที่แคบ และเอนไซม์กลายพันธุ์ของ Q727 มีค่า k_{cat} สูงกว่าเอนไซม์ *Tx*GH116 ดั้งเดิม

ประการที่สาม อธิบายบทบาทของกรดอะมิโนบริเวณเร่งปฏิกิริยาของ $TxGH116$ ที่มีโซ่ข้างเป็นอะโรมาติกในตำแหน่งจับ +1 และ +2 จากจลนพลศาสตร์ของเอนไซม์และการวิเคราะห์โครงสร้างแสดงให้เห็นว่ากรดอะมิโน Y445 มีบทบาทสำคัญในการซ้อนจับกับสับสเตรตในตำแหน่งจับ +1 และยังช่วยรักษาเสถียรภาพของสภาวะทรานซิชันเหมือนออกโซคาร์บิเนียมไอออน (oxocarbinium ion-like transition state) การกำจัดโซ่ข้างอะโรมาติกของไทโรซีนส่งผลให้ปฏิกิริยาไฮโดรไลซิสของเอนไซม์ลดลงอย่างมาก ทรานส์ไกลโคซิเลชันของเอนไซม์กลายพันธุ์ Y445L ที่ลดลงแสดงให้เห็นถึงความสำคัญของ Y445 ในการจับกับโมเลกุลตัวรับในระหว่างกระบวนการกลูโคซิลทรานส์เฟอเรส กรดอะมิโน W525 ไม่เพียงแต่ให้การซ้อนจับซับสเตรตในตำแหน่งจับ +1 และ +2 แต่อาจจะทำหน้าที่ในการปลดปล่อยผลิตภัณฑ์และการจดจำซับสเตรท การกลายพันธุ์ W525F ส่งผลให้เกิดการจับซับสเตรตสังเคราะห์ดีขึ้น และเอนไซม์กลายพันธุ์ W525L มีค่า k_{cat} สูงกว่าเอนไซม์ดั้งเดิม ปฏิกิริยาไฮโดรไลซิสที่เพิ่มขึ้นของเอนไซม์กลายพันธุ์ W525L บ่งชี้ถึงบทบาทสำคัญของ W525 ในการกำหนดอัตราส่วนของการไฮโดรไลซิสต่อการทรานส์ไกลโคซิเลชันใน $TxGH116$

การศึกษานี้ช่วยทำให้เข้าใจมากขึ้นเกี่ยวกับการทำงานของเอนไซม์นี้ โดยการหาความสัมพันธ์ที่สำคัญของกรดอะมิโนบริเวณเร่งปฏิกิริยาในการจับกับซับสเตรท และตัวยับยั้ง ซึ่งอาจนำไปสู่ปรับ-เปลี่ยนเอนไซม์เบตากลูโคซิเดส $TxGH116$ เพื่อการประยุกต์ใช้ที่ดีขึ้น

สาขาวิชาเคมี   ลายมือชื่อนักศึกษา _____

ปีการศึกษา 2563   ลายมือชื่ออาจารย์ที่ปรึกษา _____

HUANG MENG: THE ROLES OF ACTIVE SITE AMINO ACID RESIDUES OF TxGH116 IN GLUCOSE BINDING AND CATALYSIS. THESIS ADVISOR : ASSOC. PROF. JAMES KETUDAT-CAIRNS, PhD., 160 PP.

GLYCOSIDE HYDROLASE/KINETICS/MUTAGENESIS/CRYSTALLOGRAPHY

TxGH116 from *Thermoanaerobacterium xylanolyticum* is a thermostable β-glucosidase that was the first structurally characterized enzyme from glycoside hydrolase family GH116. High conservation of the amino acid residues in the GH116 family active-site pocket indicate the importance of these residues. Here, we systematically mutated the active-site amino acids residues that appeared to be related with catalytic function to quantify the contribution of each side chain to catalysis.

Firstly, we conducted the systematic functional analysis of TxGH116 active site glycone sugar binding residues: D452, H507, T591, E730, W732, E777, R786 and R792 in substrate binding and catalysis. From enzyme substrate kinetics and structural analysis, we clarified the roles of each amino acid residue. In general, simple binding interactions with the glycone of the substrate cannot fully explain the activity changes upon mutation of the glucose binding molecule and their charge-distribution and stabilization of the transition state appear to be more critical, based on larger effects on k_{cat} than K_M.

Secondly, our work investigated the amino acid residues that interact with the catalytic acid/base and nucleophile. Sequence alignment of these residues shows that the "Thr-nucleophile–Tyr" pattern of catalytic residues is generally conserved, in

which threonine has both hydrogen-bond dornor and acceptor functions. The glutamine 727 residue in "Gln-acid/base-Tyr" is not highly conserved, while the tyrosine 523 residue is and plays an essential role in modulating acid/base residue protonation. Catalytic nucleophile-binding residue mutants T450A and Y790F had large decreases in activity, while T450A had a narrow pH profile with an optimum close to neutral, but Y790F had a broad pH profile. Catalytic acid/base-binding residue mutants Y523F, Q727A and Q727E all show narrow pH optima profile and the Q727 mutants have higher k_{cat} values than wild type TxGH116.

Thirdly, we elucidated the role of aromatic residues in the TxGH116 slot like active site subsites +1 and +2. Enzyme kinetics and structural analysis showed that the Y445 plays a pivotal role in stacking the substrate moiety in subsite +1 and also helps to stabilize the oxocarbinium ion-like transition state. Elimination of the tyrosine aromatic platform resulted in a large reduction in hydrolytic activity. The decreased transglycosylation in the TxGH116 Y445L mutant indicated the importance of the Y445 in acceptor molecule binding during the glucosyl transferase process. The W525 residue not only provides substrate stacking in subsites +1 and +2, but may also act in product release and substrate recognition. The W525F mutation resulted in improved synthetic substrate binding and W525L has a higher k_{cat} value than WT. The increased hydrolysis activity from the W525L mutant indicated the pivotal role of W525 in determining the ratio of hydrolysis to transglycosylation in TxGH116.

These studies improve our understanding of how this enzyme functions which may allow the modification of TxGH116 β-glucosidase for improved application.

School of Chemistry

Academic Year 2020

Student's Signature _____

Advisor's Signature _____

ACKNOWLEDGEMENTS

active brain that bring me so much critical idea and thinking during my work, and also the most Chinese conversation I got around the lab. At the same time, I would like to express my heartfelt thanks to all the teachers who have helped and taught me in my study in SUT, and sincerely thanks to my lab members, colleagues and staff in the School of Chemistry, Suranaree University of Technology for their help in my study and living in Thailand.

I would like to thank Assoc. Prof. Dr. Chun-Jung Chen and the protein crystallography beamline staff of the National Synchrotron Radiation Research Center (NSRRC), Hsinchu, Taiwan, for data collection and processing assistance.

Finally, I would like to express the most heartfelt gratitude to my family members who have provided me the unconditional great help and support from the beginning on my way to explore science. Though the most unfortunate thing happened to me that I could not meet them frequently that made me missed so much from them, while I survived and became stronger. Their love is always the source of my strength during the period of my study abroad. And I love you, Dad and Mom.

Huang Meng

CONTENTS

CONTENTS (Continued)

CONTENTS (Continued)

CONTENTS (Continued)

CONTENTS (Continued)

CONTENTS (Continued)

CONTENTS (Continued)

CONTENTS (Continued)

Page

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF ABBREVIATIONS

Abs	Absorbance
ABTS	2,2'-azinobis-3-ethylbenthaiazolinesulfonic acid
(m, μ)l	(milli, micro) Liter(s)
(m, μ)g	(milli, micro) Gram(s)
(m, μ)M	(milli, micro) Molar(s)
(μ)mol	(micro) Mole(s)
°C	Degrees Celsius
bp	Base pairs
BSA	Bovine serum albumin
BGC	Beta glucose
CV	Column Volume
DNA	deoxynucleic acid
dNTPs	Deoxynucleoside triphosphates
EDTA	Ethylenediamine tetraacetate
ESI-MS	Electrospray ionization-mass spectrometry
EtOAc	Ethyl acetate
G2F	2-deoxy-2-fluoro-α-D-glucosyl moiety
GH	Glycoside hydrolase
GH116	Glycoside hydrolase 116
Glc	Alpha glucose
GIM	Glucosyl imidazole

LIST OF ABBREVIATIONS (Continued)

GOL	Glycerol
hr	Hour(s)
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl thio-β-D-galactoside
kDa	Kilo Dalton(s)
LB	Luria-Bertani lysogeny broth
min	Minute(s)
MW	Molecular weight
nm	nanometer(s)
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PMSF	Phenylmethylsulfonylfluoride
pNP	para-Nitrophenyl
pNPAra	pNP-α-L-arabinopyranoside
pNPC2	pNP-β-D-cellobioside
pNPGal	pNP-β-D-galactopyranoside
PVDF	Polyvinylidene flouride
RNA	Ribonucleic acid
rpm	rotations per minute
s	second(s)
SDS	Sodium dodecyl sulfate

LIST OF ABBREVIATIONS (Continued)

TEMED	Tetramethylenediamine
TLC	Thin-layer chromatography
Tris	Tris-(hydroxymethyl)-aminoethane
UV	Ultraviolet
v/v	Volume/volume
w/v	Weight/volume
pNPGlc	pNP-β-D-glucopyranoside
pNPXyl	pNP-β-D-xylopyranoside

CHAPTER I

INTRODUCTION

1.1 General introduction

An enzyme's catalytic reaction normally occurs in a buried pocket like active site, where a multiple of functional groups concert to comprise a unique environment that enables the catalysis of reactions with unparalleled efficiency and specificity. In the active site and surrounding sites, a relatively small number of specific amino acid residues are involved in substrate binding and an even smaller subset of these directly act in the catalytic reaction. These catalytic residues provide various reactive groups that promote catalysis or improve substrate and transition state binding through the provision of specific amino acid residues traits (Holliday et al., 2005). Functions performed by amino acid residues during catalysis have been combined into seven basic classes: stabilization roles, steric roles, activation roles, proton shuttling residues, hydrogen shuttling residues, electron shuttling residues and residues acting through covalent catalysis (Holliday et al., 2009).

Glycoside hydrolases (GHs) are ubiquitous in nature and catalyze the removal of carbohydrates from a range of biomolecules (McCarter and Withers, 1994). GHs exhibit exceptional catalytic power, based on their efficient working mechanism (Wolfenden and Snider, 2001). Based on their amino acid sequences, GHs enzymes can be classified in families and clans that show conserved catalytic mechanism, structure, and active site residues, but may vary in substrate specificity.

Their powerful catalytic action originates from the exquisite positional binding of substrates, in which the catalytic residues stabilize the charges formed during the passage through the transition state (Jencks, 1975). Unraveling the fine details of substrate and transition state binding, especially the geometry and energetics involved, is essential for understanding the catalytic efficiency of enzymes.

Beta-glucosidases (β-D-glucopyranoside glucohydrolases, E.C. 3.2.1.21) are enzymes which remove the nonreducing terminal β-D-glucosyl residue from glucoconjugates, including glucosides, 1-O-glucosyl esters, and oligosaccharides (Ketudat Cairns et al., 2015). β-Glucosidases have been categorized into the protein sequence-based GHs families GH1, GH2, GH3, GH5, GH9, GH30, GH116, etc (http://www.cazy.org) (Lombard et al., 2014), and different families show their own specific traits in either hydrolysis and transglycosylation. The specificity of β-glucosidase toward different substrates varies mostly depending on the enzyme source which is related with their function. Previous theoretical and experimental developments have revealed that hydrolysis of the glycosidic bond can occur with one of two possible stereochemical outcomes: inversion or retention of anomeric configuration (McCarter and Withers, 1994; Zechel and Withers, 2000), which helped us understand their working mechanism and to make use of β-glucosidase for potential applicatons.

Many studies elucidate the role of β-glucosidases catalytic residues that perform the chemistry, including general acid/base and nucleophile (McCarthy et al., 2004), electrostatic (Geronimo et al., 2018) contributions to catalysis and even determine the mechanism to convert between glycosidase with transglycosidase (Frutuoso and Marana, 2013; Lundemo et al., 2017; Teze et al., 2015). Hydrogen

bonding and stacking interactions are the dominant interactions in protein–carbohydrate complexes (Quiocho et al., 1989). The sugar-hydroxyl groups can serve simultaneously as hydrogen donors and acceptors and may potentially be involved in as many as three hydrogen bonds (as a donor of one hydrogen bond and an acceptor of two) (Quiocho et al., 1989). Stacking or hydrophobic interactions of one or both faces of the sugar ring can be formed with aromatic residues at the binding site of GHs enzymes (Vyas et al., 2002). Relating thermodynamic parameters to structural and biochemical data allows a better understanding of substrate binding and its contribution to catalysis. The analysis of the binding of carbohydrates to proteins or enzymes is a special challenge because of the multiple interactions and forces involved (Zolotnitsky et al., 2004).

Generally noncatalytic amino acids are routinely assumed to be only important for facilitating substrate binding (Chuenchor et al., 2011). But some previous studies have shown that the contributions of enzyme active-site residues that are not directly involved in bond making and breaking could be as essential to the enzyme's function as catalytic residues (Dopitova et al., 2008; Lucas and Siegel, 2015; Pengthaisong et al., 2012; Zhang et al., 2015). This enables enzymes to conduct catalysis through a variety of strategies, such as introducing strain to destabilize the substrate (Desjardins et al., 2017) and preorganizing active-site dipoles for transition-state stabilization (De Giuseppe et al., 2014; Shanmugam et al., 2018). Previous GHs enzyme work indicated that noncovalent interactions mediated by aromatic rings are pivotal to sugar molecule binding (Laughrey et al., 2008; Spiwok, 2017), which opens up another new research area.

The enzyme used in this work *Thermoanaerobacterium xylanolyticum* GH116

(*Tx*GH116) is a thermostable β-glucosidase, which consists of an N-terminal domain formed by a two-sheet β-sandwich, and the slot like active site is located in a C-terminal $(\alpha/\alpha)_6$ solenoid domain surface, near the surface of the C-terminal domain close the interface of two domains (Charoenwattanasatien et al., 2016). High-resolution crystal structures of *Tx*GH116–glucose and inhibitor complexes provide invaluable insight into the different elements involved in substrate binding and catalysis (PDB ID: 5BX2, 5BX3, 5BX4, 5BX5 and 5FJS) (Artola et al., 2017; Charoenwattanasatien et al., 2016). Our previous work verified that the enzyme has a retaining mechanism by identifying the initial hydrolysis product as β-D-glucose, which retains the anomeric stereochemistry of the substrate, by monitoring the reaction time course by NMR spectroscopy (Charoenwattanasatien et al., 2016). E441 acts as the catalytic nucleophile and D593 as the catalytic acid/base, which was confirmed by mutation of these residues to alanine, followed by chemical rescue using small nucleophiles.

Catalytic residue mutations has proved to be applicable to modify *Tx*GH116 for synthesis of glycosides and oligosaccharides (Gorantla et al., 2019; Pengthaisong et al., 2021). The amino acid residues forming hydrogen bonding interactions between the *Tx*GH116 active site and the substrate glucosyl moiety include E441 at OH2; H507 and D452 at OH3; D452, T591, and R792 at OH4; and E777 and R786 at the OH6. In the last case, OH6 appears to mediate a salt bridge between E777 and R786. Y523 and Q727 interact with the catalytic acid/base D593; while T450 and Y790 interact with the nucleophile, E441. Y445 and W525 located in subsite +1, +2 for aromatic stacking roles. The highly tractable *Tx*GH116 model, in which all pathogenic GBA2 mutation sites are strictly conserved with human GBA2, and helped

to elucidate the molecular bases of human GBA2 mutations leading to ataxias and paraplegias (Sultana et al., 2020; Woeste and Wachten, 2017).

Although residues previously proposed to be directly involved in the reaction chemistry have been analyzed (Charoenwattanasatien et al., 2016); however, the importance and role of other residues within the active site pocket of *Tx*GH116 and other GH116 family enzymes remain to be clarified. In this work, we systematically mutated amino acids that form the remainder of the active-site residues interacting with the glucose ligand to quantify the contribution of each side chain toward the catalytic efficiency of the *Tx*GH116 enzyme. These mechanistic details improve our understanding of how this enzyme achieves function by teasing out the relative importance of active site residues in substrate and inhibitor binding, which may allow the modification of *Tx*GH116 β-glucosidase for improved application. They also allow improved elucidation of related human disorders from GBA2 mutants which play important role in balancing sphingolipid level in cell (Jatooratthawichot et al., 2020).

To gain further insights into factors influencing the hydrophobic interactions, a bioinformatic analysis of the GH116 family as well as biochemical determinations were carried out. Our aim is to understand the influence of hydrogen bonding interactions between enzyme and substrate, catalytic residues related interactions in hydrolysis activity influence and carbohydrate/aromatic stacking on the pyranose reactivity to eventually elucidate the role of amino acids in the GH116 family active site. The catalytic efficiency and the mode of action of the enzyme also changed dramatically after subsite site +1, +2 aromatic residues systematically mutated. It is hoped that the key findings in this study will have wide applications in guiding the rational design of other more divergent GH116 enzymes.

One group of GH of industrial interest is β-glucosidases, which play an important role in breaking down glucose oligomers and cellobiose during lignocellulosic biomass conversion, producing glucose and alleviating the inhibition of cellobiohydrolases and endoglucanases by these oligosaccharides (Decker et al., 2000; Yun et al., 2001). This is the important part in the context of second generation bioethanol production. On the other hand, β-glucosidases are themselves inhibited by an increasing glucose concentration (Decker et al., 2000; Korotkova et al., 2009; Yun et al., 2001), which makes the search for highly glucose-tolerant enzymes crucial for an effective hydrolysis of plant biomasses (Guo et al., 2016; Madhavan et al., 2017; Singhania et al., 2017).

TxGH116 thermostable β-glucosidase may be useful for biomass conversion applications, since thermophilic β-glucosidases have been shown to be important for this industrial application (Guo et al., 2016). Although TxGH116 β-glucosidase is stable at 60-70°C, which is appropriate for biomass conversion, its use in biomass conversion is likely to be limited by its glucose inhibition (K_i 4 mM at 60°C) which supposed to getting product inhibition feedback. Therefore, I assessed the roles of the glucose-binding residues in substrate binding and hydrolysis and glucose inhibition (competitive K_i) by mutagenesis, followed by kinetic characterization. Variant enzymes that have lower affinity for glucose but rapid hydrolysis of oligosaccharides at medium to high concentrations may be appropriate for use in plant biomass conversion, in collaboration with endoglucanases (e.g. cellulases) and cellobiosidases. Residues interact with catalytic residues and subsite +1, +2 site show their potential roles in maintain catalytic function and species dependent substrate recognition. This will help us in understanding GH116 family enzyme function performance and structural based high efficiency inhibitor designment.

1.2 Research objectives

1.2.1 To elucidate the essential amino residues for glucose binding and substrate hydrolysis of *Tx*GH116 by evaluating the effects of mutations on substrate hydrolysis and glucose-binding.

1.2.2 To check the glucose tolerance of mutants with high hydrolysis activity, to learn whether they have improved characteristics for biomass conversion.

CHAPTER II

LITERATURE REVIEW

2.1 Carbohydrate Active Enzymes

Carbohydrates in nature normally have similar chemical composition but have a variety of biochemical traits. Carbohydrates can be linked to other, non-carbohydrate molecules to generate a wide range of interesting glycoconjugates (Chen, 2017; Dwek, 1996). Complex carbohydrates are widely distributed in nature and play important roles in vital metabolic pathways. They mediate a multitude of biological functions, serving as carbon reserves, structural molecules, and mediators of intra- and intercellular recognition within one organism and even between different organisms. The diversity of carbohydrates is controlled by a panel of enzymes involved in their assembly (glycosyltransferases) and their breakdown (glycoside hydrolases, polysaccharide lyases, carbohydrate esterases), all of which are collectively designated as Carbohydrate-Active enzymes (CAZymes) (Lombard et al., 2014). How these enzymes achieve selective recognition of target substrates that display only subtle stereochemical differences is key to prediction of substrate specificity and potential functional modification for application.

Carbohydrates differ from the other classes of biomolecules in that their constitutive moieties (monosaccharides) may be connected to one another by a great variety of linkage types. In addition, they can be highly branched from their multiple side chain groups, thus allowing oligosaccharides to provide an almost infinite array

of structural variations (Poveda, 1998). The decoding process of the existing information in oligosaccharide structures involves their recognition by interactions with other biomolecules. Thus, they are most often specifically recognized by proteins and these interactions usually mediate a particular biological response, such as host–parasite interactions, fertilization, autoimmune disorders and cellular differentiation (Dwek, 1996), which were important in biological metabolism.

2.2   Glycoside hydrolase enzymes

Glycoside hydrolases (GHs) are ubiquitous in nature and catalyze the removal of carbohydrates from a range of biomolecules to perform their biological functions (McCarter and Withers, 1994; Zechel and Withers, 2000). This includes essential roles in pathogen infection, antibacterial defense, and many other essential cellular processes for digesting carbohydrates. GHs are also critical for degrading plant biomass and now days are widely used for industrial application (Guo et al., 2016; Liu et al., 2019; Santos et al., 2019; Singhania et al., 2017; Varki, 2017). Enzymes which catalyze the hydrolysis of glycosidic linkages are widely distributed in nature that include varieties like amylases, invertases, maltases, α-galactosidases, β-galactosidases, etc (Dwek, 1996). Different glycosidases display completely different ranges of bond and substrate specificity and form different products.

In higher organisms, many glycoside hydrolases are found within the endoplasmic reticulum and Golgi apparatus, where they are involved in processing of N-linked and O-linked glycoproteins, and in the lysosome as enzymes involved in the degradation of carbohydrate linked biomolecules. Deficiency in specific lysosomal

glycoside hydrolases in humans can lead to lysosomal storage disorders that result in developmental problems or death (Mistry et al., 2014; Sultana et al., 2020).

2.3   Catalytic mechanism of glycoside hydrolases

The enzymatic hydrolysis of the glycosidic bond takes place via general acid/base catalysis that requires two critical residues: a proton donor and a nucleophile/base (G. Davies and B. Henrissat, 1995). Hydrolysis of the glycosidic bond can occur with one of two basic outcomes: inversion or retention of the anomeric configuration (Figure 2.1) based on the anomeric configuration changing from (β to α) or being retainedas (β to β). Generally, all enzymes within a sequence-related family appear to catalyze reactions with the same stereochemical outcome (Gebler et al., 1992). The two mechanisms differ in that inverting glycosidases operate via a direct displacement of the leaving group by water, whereas retaining glycosidases utilize a double displacement mechanism involving a glycosyl-enzyme intermediate (McCarter and Withers, 1994). The two carboxyl groups in inverting glycosidases serve as general acid and general base catalysts and are suitably placed, some 10.5 Å apart on average (McCarter and Withers, 1994), to allow the substrate and a water molecule to bind between them (Figure 2.2). Reaction occurs via a single-displacement mechanism involving an oxocarbenium ion-like transition state.

By contrast, the acid/base and nucleophie typically carboxyl groups in retaining glycosidases are usually only 5-5.5 Å apart, consistent with a double-displacement mechanism involving a covalent glycosyl enzyme intermediate (

Figure 2.1) (G. Davies and B. Henrissat, 1995; McCarter and Withers, 1994). In the first step, glycosylation called one of the carboxyl groups functions as a general acid catalyst, protonating the glycosidic oxygen concomitantly with bond

cleavage. The other acts as a nucleophile, forming a covalent glycosyl enzyme intermediate. In the second step, deglycosylation called the side-chain carboxylate deprotonates the incoming water molecule, which attacks at the anomeric center and displaces the sugar (Zechel and Withers, 2000). Both steps occur via transition states with oxocarbenium ion character. In some cases, one of the carboxyl groups is replaced by an activated Tyr ohenolic hydroxyl or other ionizable group. Although there are other more rarely described GH mechanisms, most GH are thought to use one of these two classic mechanism.
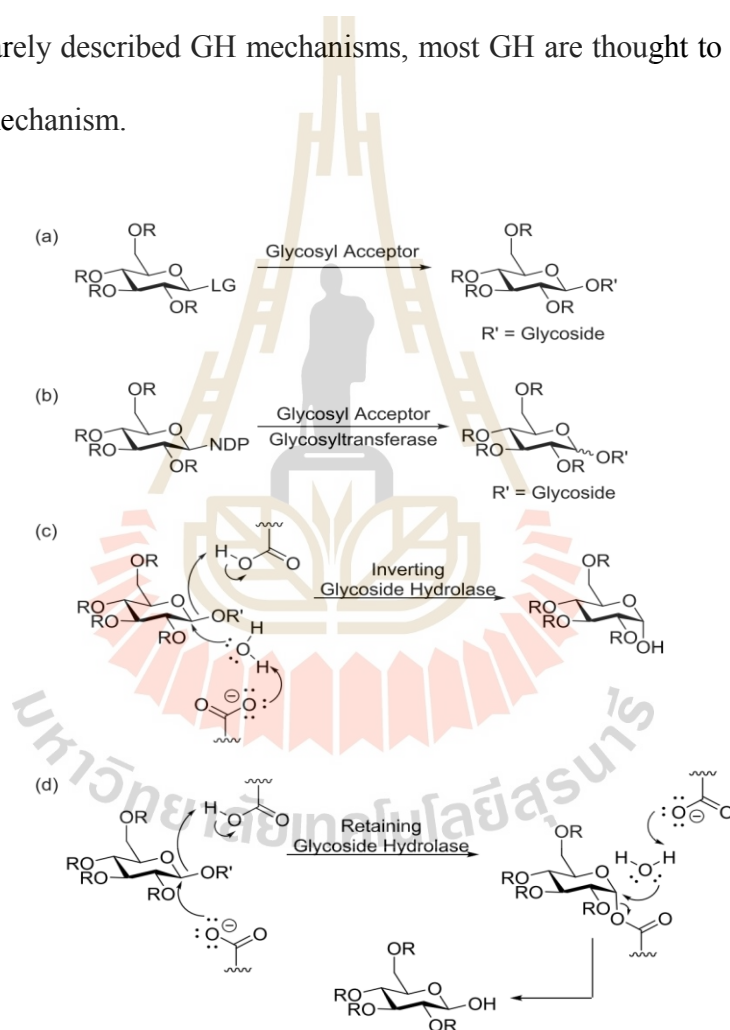
Figure 2.1 Mechanisms of different types of glycosidases. (a) General Glycosylation Reaction; (b) Glycosyltransferase Reaction (Retaining or Inverting); (c) Mechanism of Inverting Glycoside Hydrolases; (d) Mechanism of Retaining Glycoside Hydrolases. (McCarter and Withers, 1994; Quirke and Crich, 2020).
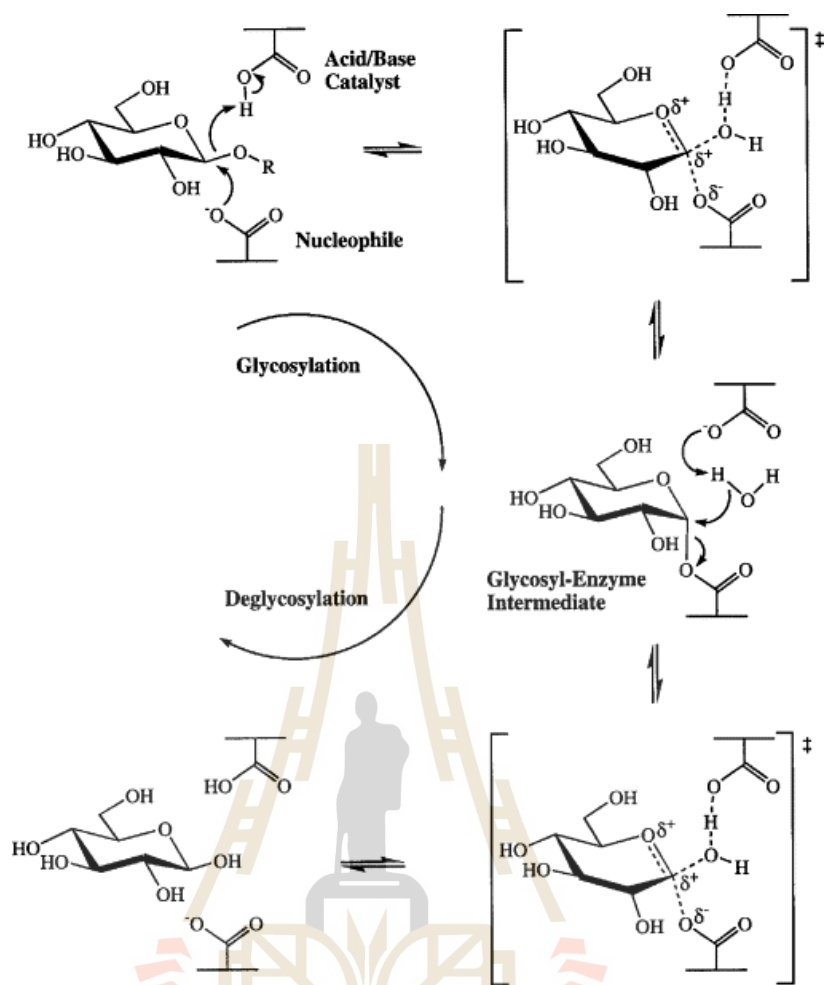
Figure 2.2 Double-displacement mechanism of a retaining *β*-glycosidase.
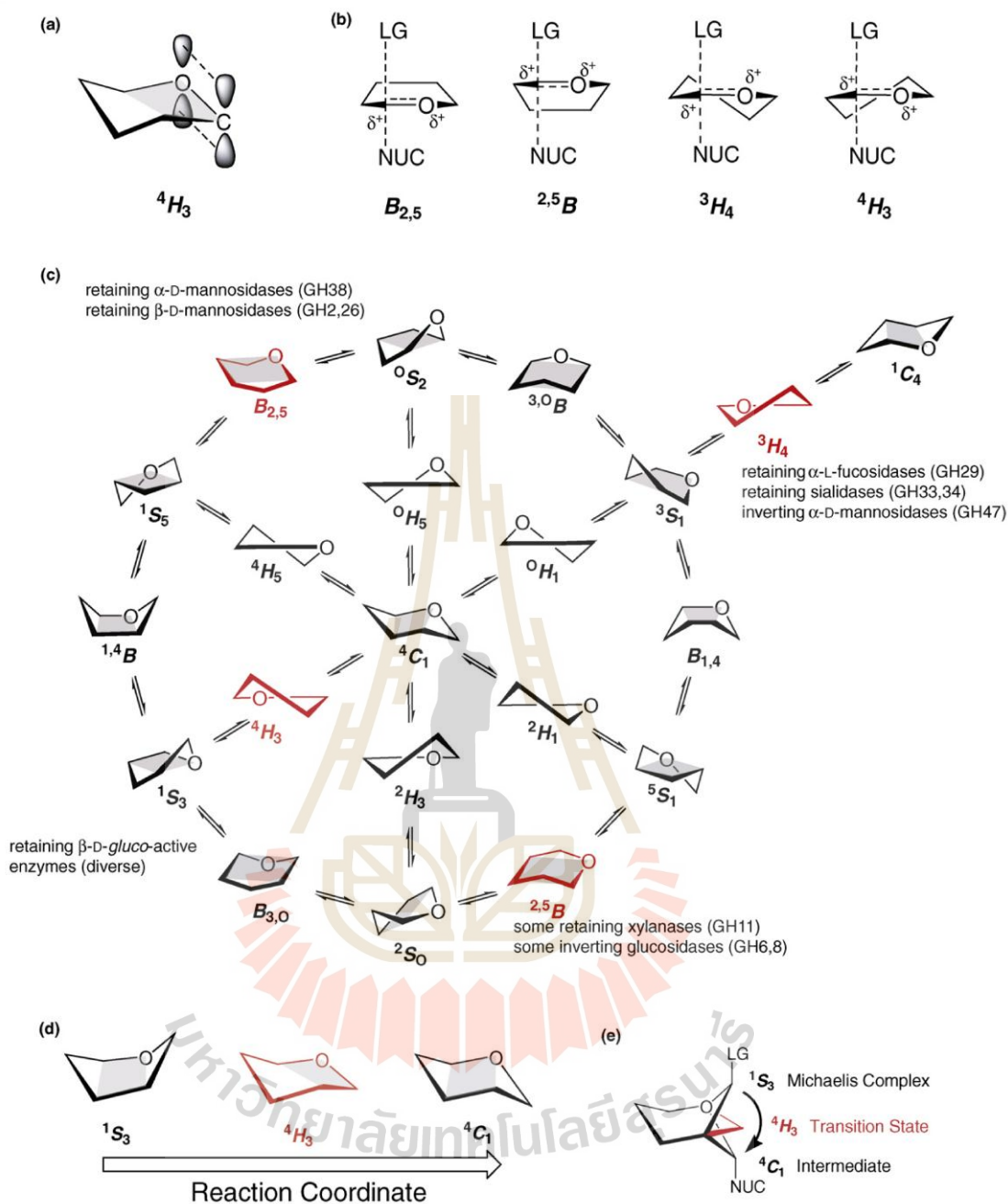
Figure 2.3 The reaction coordinates of different glycoside hydrolases follow different conformational itineraries that proceed through specific transition state structures. (a) Stabilization of the developing positive charge at the anomeric center by the p-like lone pair of the endocylic oxygen favors the oxocarbenium ion-like transition state adopting certain conformations. (b) Possible conformations of oxocarbenium ion-

like transition states indicating the relative positions of nucleophile and leaving group. (c) Stoddard's pseudorotational itinerary showing the conformational itineraries followed by various families of GHs. Predicted transition state structures are shown in red. (d) The conformational itineraries followed by nearly all GHs reflect electrophilic migration of the anomeric center. (e) Simplified schematic showing the electrophilic migration of the anomeric center with the predicted transition state shown in red (Vocadlo and Davies, 2008).

The flexibility from ligands main with the forms: for cyclohexane, a typical six-membered ring structure with six equivalent carbon atoms, five distinct conformers are defined: chair (C), boat (B), twist/skew-boat (S), half-chair (H), and envelope (E) conformations (Figure 2.3, a, b). For pyranoses, the existence of the ring oxygen and C6 carbon makes all ring carbons distinguishable. Because of this asymmetry, 38 types of distinct conformers in total are defined for the pyranose ring: 2C, 6B, 6S, 12H, and 12E conformations (Figure 2.3, c) (Satoh and Manabe, 2013; Vocadlo and Davies, 2008).

Carbohydrate side chain, meaning the exocyclic carbon atom and its associated atoms conformation confers a significant influence on reactivity during glycosylation and anomeric bond hydrolysis, due to stabilization of the oxocarbenium-like transition state. Determined from examination of crystal structure databases and NMR data, the side chains of free hexopyranose sugars populate three staggered conformations called here gauche,gauche (gg), gauche,trans (gt), and trans,gauche (tg), where the first and second descriptors refer to the stereochemical

relationship of the C6–O6 bond in the side chain with the C5–O5 bond and the C5–C4 bond in the pyranose ring, respectively (Figure 2.4). Most glucosidases and β-mannosidases preferentially bind their substrates in the most reactive gg conformation, thereby maximizing stabilization of the corresponding oxocarbenium ion-like transition state during hydrolysis (Figure 2.3). α-Galactoside hydrolases mostly show a preference for the second most activating gt conformation to avoid the energy penalty that would arise from imposing the gg conformation on galacto-configured ligands (Quirke and Crich, 2020). CAZymes constitute a special case, as not only the enzymes themselves but also the substrates they bind are particularly flexible (Ardèvol and Rovira, 2015).
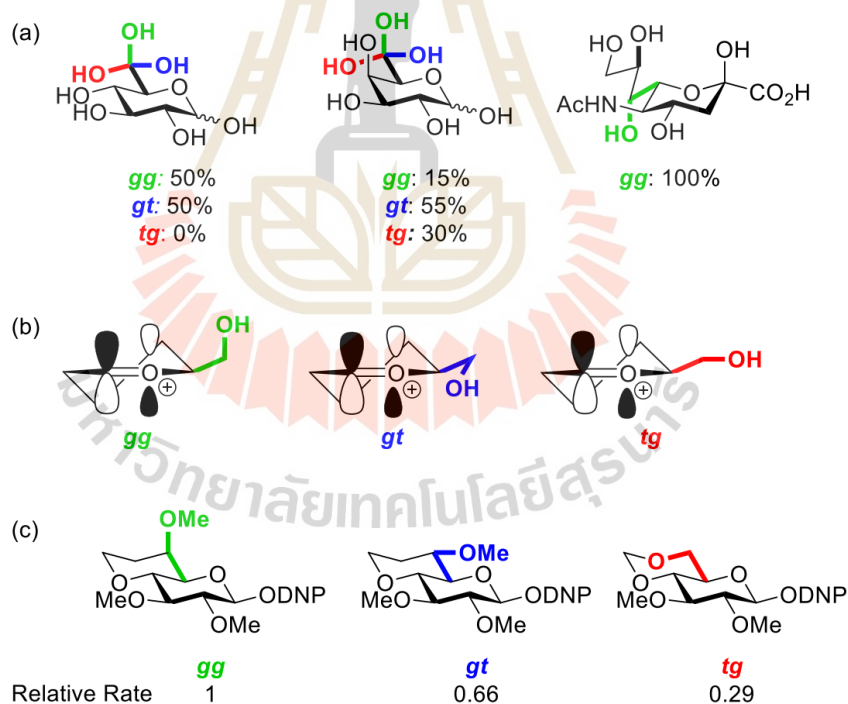
Figure 2.4 The unrestricted distribution of side chain conformations between the *gauche,gauche* (*gg*), *gauche,trans* (*gt*), and *trans,gauche* (*tg*) conformations. (a) Three staggered side chain conformations and their approximate populations in free solution for gluco- and galactopyranoses and N-acetyl

neuraminic acid. (b) Spatial relationships of side chain hydroxyl groups with the oxocarbenium π^* orbital. (c) Relative rates of spontaneous hydrolysis of conformationally locked dinitrophenyl glycopyranosides (Quirke and Crich, 2020).

2.4   β-Glucosidases

β-Glucosidase (EC 3.2.1.21) is one of the earlier discovered and most widely studied enzymes, due to its universal distribution and well defined activity, wide variety of substrates and simple nature of enzyme assay. They have been used for the liberation of glucose from cellooligosaccharides and glucosides linked by a β-glycosidic bond (Ketudat Cairns and Esen, 2010; Shewale, 1982). In cellulolytic microorganisms, β-glucosidase is involved in cellulase induction (due to its transglycosylation activities) and cellulose hydrolysis (Kuntothom and Cairns, 2020). Fungal β-glucosidases are a part of the cellulose degrading enzymes that divide cellobiose into two glucose molecules to facilitate the glucan-breakdown work of endoglucanases and cellobiohydrolases. The action of β-glucosidases can protect these enzymes from the product inhibition effect of cellobiose, which improves the biomass degradation efficiency (Nakazawa et al., 2012). β-Glucosidases are involved in mammals glycolipids and dietary glucosides metabolism and may have signaling functions. Previous work showed that a human β-glucosidase (lactase phlorizin hydrolase) is involved in flavonoid metabolism (Day et al., 2000) and plays an essential role in the metabolism of dietary lactose due to its β-galactosidase activity (Naim, 2001). Mammalian cytoplasmic β-glucosidase has also been implicated in breakdown of exogenous flavonoid glycosides, along with a possible function in

glycolipid metabolism, which is a known role of lysosomal acid β-glucosidase and bile acid β-glucosidase (Ketudat Cairns and Esen, 2010). β-Glucosidases are also extensively utilized in biotechnological procedures in the pharmaceutical and food industries (production of bioactive agents with antioxidant and antimicrobial properties for use in cosmetics and aroma enhancement in winemaking), as well as in biofuels production (Crespim et al., 2016; Madhavan et al., 2017; Sindhu et al., 2017; Zanphorlin et al., 2016). In the lastcase, the application of β-glucosidases in the conversion of high-cellulose-content biomass to fermentable sugars for the production of fuel ethanol is an intensively studied area.

2.5   Carbohydrate-protein interactions

The study of how carbohydrates are recognized by the binding sites of enzymes is a topic of major interest, which helps us to elucidate their working mechanism. It is evident that knowledge of the three dimensional structure of these biomolecules (proteins and carbohydrates) support stereochemical information in the design of new carbohydrate-based therapeutic agents (Kamiya et al., 2011). Protein-carbohydrate interactions play vital roles in several biological processes in living organisms (Dwek, 1996). The comparative analysis of binding site protein amino acid residues along with stabilizing residues in protein-carbohydrate complexes provides ample insights to understand the structure, function and recognition mechanism (Shanmugam et al., 2018). Current technical facilities and biophysical techniques which are applicable for characterizing protein-carbohydrate interactions, mainly X-ray crystallography and NMR, have allowed access to detailed information on the

three dimensional structure of protein-carbohydrate complexes (G. Davies and B. Henrissat, 1995; Poveda, 1998; Weis and Drickamer, 1996).

Table 2.1 Amino acid residue catalytic roles grouped into generic classes (Holliday et al., 2009)

Generic role	Residue roles in catalysis
Stabilisation	Electrostatic stabilisation; electrostatic destabilisation; radical stabilisation
Steric	Steric role
Activation	Activator (includes chemical and thermal activation); increase electrophilicity; increase nucleophilicity; increase basicity; increase acidity; promote homolysis; promote heterolysis
Proton shuttling	Proton donor; proton acceptor; proton relay.
Hydrogen shuttling	Hydrogen acceptor; hydrogen donor; hydrogen relay.
Electron shuttling	Single electron donor; single electron acceptor; electron pair donor, electron pair acceptor; electron relay
Covalent catalysis	Nucleophile; nucleofuge; electrophile; electrofuge; radical combinant; leaving group

On the basis of crystal structures of a variety of complexes, Poveda (1998) show the amino acids most commonly involved in hydrogen bond formation with carbohydrates are known to be: Asp, Asn > Glu > Arg, His, Trp, Lys > Tyr, Gln > Ser, Thr. On the other hand, those most usually observed in van der Waals interactions are Trp, Phe, Tyr, Leu, Val and Ala; that is, those with aromatic or aliphatic side chains (Poveda, 1998). In enzyme active sites, a relatively small number of amino acid residues are involved in substrate molecule (and/or cofactor) binding and an even smaller subset of these are actively involved in the catalytic process. These catalytic residues provide various reactive groups that promote catalysis through the provision of a variety of functions (Table 2.1) (Holliday et al., 2009). Apart from the direct interactions between atoms of both the protein and carbohydrate, other molecules are

often involved in the formation of the complex. For example the common water molecules, either located in the binding site or at the surface of the protein have also been shown to provide additional interactions which also help to stabilize the complex and to achieve multiple reactivity or higher selectivity (Weis and Drickamer, 1996).

2.6   Carbohydrate: aromatic interactions in protein

The recognition and interactions between saccharides and proteins has far reaching in biological, pharmacological, and technological applications (Bertozzi and Kiessling, 2001; Seeberger and Werz, 2007; Vacas et al., 2010). In addition to hydrogen bonding, a common feature of carbohydrate-binding proteins is the interaction of the R-face of the carbohydrate with the face of an aromatic side chain (Laughrey et al., 2008). Carbohydrate-π interactions are important for the recognition of carbohydrates and that these interactions are dependent on the electronic nature of the aromatic group. Noncovalent interactions involving aromatic rings such as π-stacking, cation/π, and anion/π interactions are central to many areas of modern chemistry (Figure 2.5).

Conventional hydrogen-bonding interactions are usually involved in carbohydrate recognition which determined the enzyme substrate specificity (Holliday et al., 2009) as aboved reviewed less expectedly, researchers observed that despite the highly hydrophilic character of most sugars, aromatic rings of the receptor often play an important role in carbohydrate recognition (Andrews et al., 2000; Del Carmen Fernandez-Alonso et al., 2005; Hamre et al., 2015; Tagami et al., 2013; Uchiyama et al., 2001). Some examples have been reported in which tyrosine or phenylalanine residues seem to assist the cleavage/formation of glycosidic linkages by presenting favorable interactions with the

carbohydrate transition states in glycoside hydrolases (GHs) (Nerinckx et al., 2003). It is commonly accepted that noncovalent interactions mediated by aromatic rings are pivotal to sugar binding (Figure 2.6) (Laughrey et al., 2008; Spiwok, 2017). For example, aromatic residues often stack against the faces of sugar pyranose rings in complexes between proteins and carbohydrates (Hsu et al., 2016). Such contacts typically involve two or three CH groups of the pyranoses and the π electron density of the aromatic ring (called CH/π bonds), and these interactions can exhibit a variety of geometries, with either parallel or nonparallel arrangements of the aromatic and sugar rings (Figure 2.7) (Juan et al., 2013). Thus, the hydrophobic platform comprising a phenylalanine residue which is highly conserved in the active center of many glycosyl hydrolases, has been proposed as a mechanistically relevant transition state stabilizing factor (Nerinckx et al., 2003). A notable feature in processive GHs show one special feature is the ubiquity of aromatic and polar residues lining the enzyme tunnels and clefts (Andrews et al., 2000; Hefferon et al., 2019). The conserved tryptophan residues govern the turnover rate of processive cellulase (Rojel et al., 2020). These residues were thought to mainly be responsible for substrate chain acquisition, binding, and hydrolysis in the catalytic tunnel/cleft via carbohydrate-π stacking, hydrogen bonding and salt bridge interactions (Figure 2.7) (Juan et al., 2013). Research work found that these non-bonded interactions are usually assumed to facilitate the processive mechanism whereby the enzyme must maintain attachment to the substrate while still allowing forward processive motion (Hamre et al., 2015; Payne et al., 2011; Zakariassen et al., 2009).
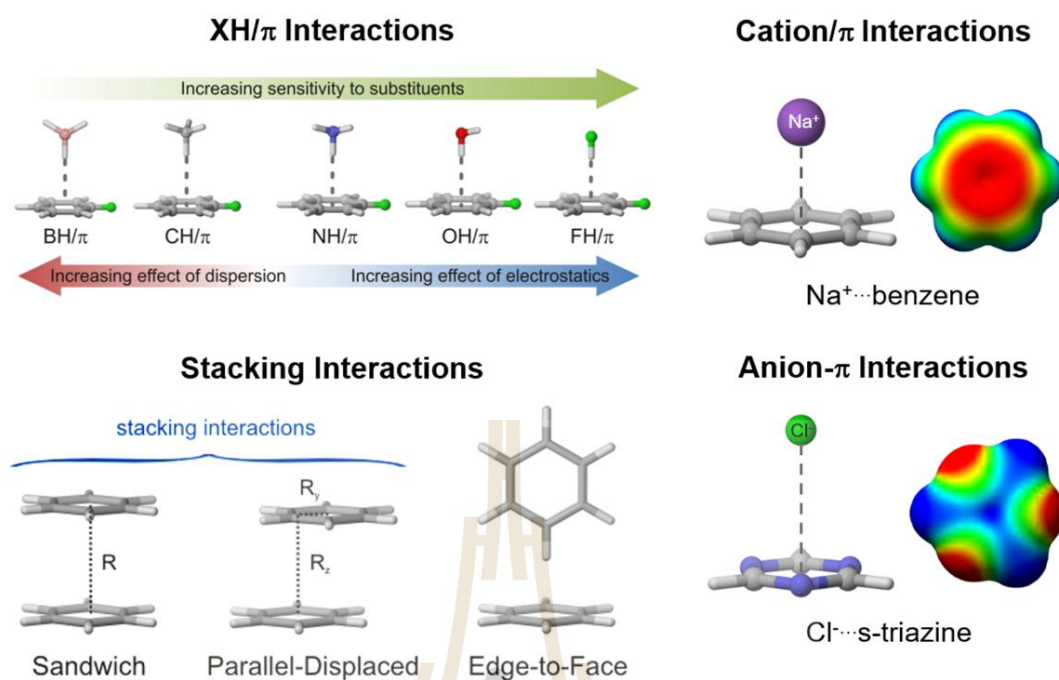
Figure 2.5 Non-covalent interactions between XH, cation, anion and stacking interactions with aromatic molecule. (Wheeler, 2013)
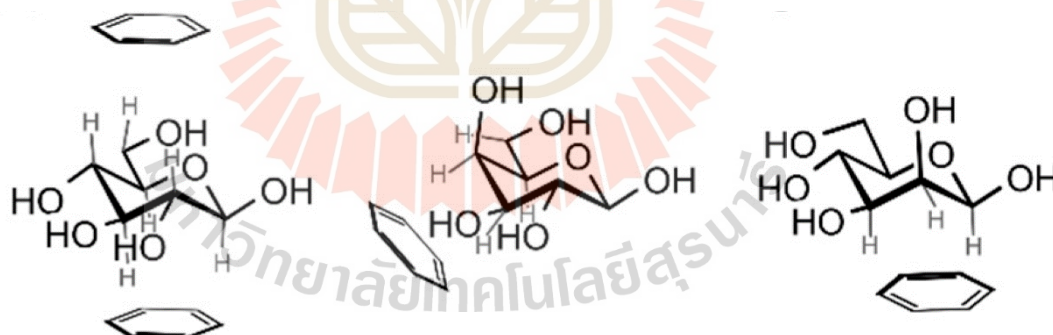
Figure 2.6 Prevalent geometries of carbohydrate–aromatic CH/π interactions: (Left) β-D-glucopyranose can interact via both faces (only the A face is available for α-D-glucopyranose); (Middle) α- or β-D-galactose or D-fucopyranose interacts via hydrogens on atoms C3, C4, C5 and C6 (a mirror assembly is typical for L-fucopyranose); (Right) β-D-mannopyranose interacts via hydrogen atoms on C1, C2, C3 and C5 (Spiwok, 2017).

A novel mechanism with kinetic, structural and theoretical considerations, has made by making use of a tyrosine residue, instead of a carboxylate, to stabilize the oxocarbenium-like transition state generated during the frontface attack of the glycosidic acceptor to the acetalic center (Iglesias-Fernandez et al., 2017). Interestingly, this tyrosine residue is originally involved in a parallel stacking with the α-face of the glycosidic donor and therefore seems to play a dual role, participating both in substrate recognition and catalysis (Montalvillo-Jimenez et al., 2019). These examples illustrate that under appropriate conditions, aromatic platforms can be employed to stabilize glycoside transition states (Rojel et al., 2020).
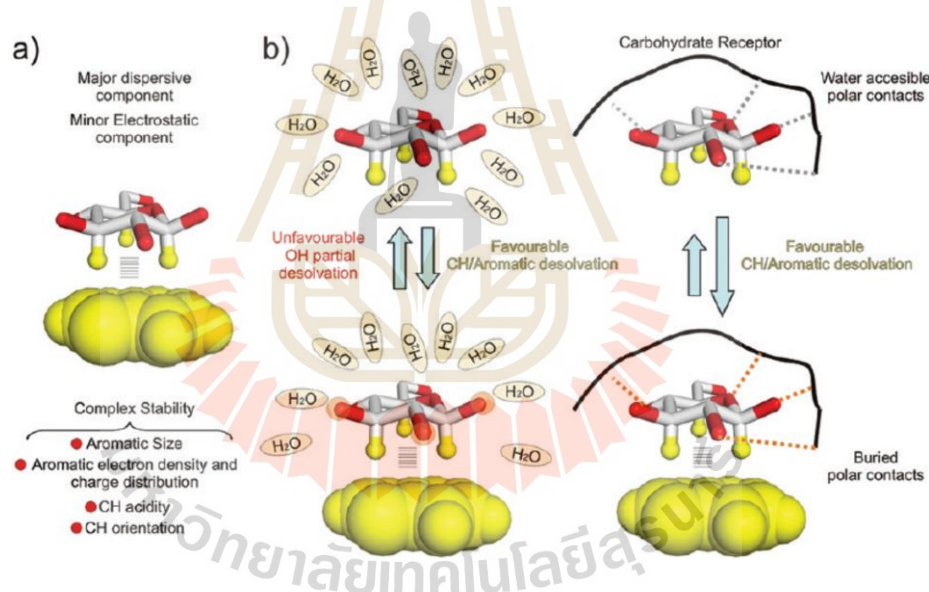
Figure 2.7   Carbohydrate Aromatic Interactions. (a) Carbohydrate-aromatic interactions have a major dispersive component and the complex stability depends on several factors. The orientation of the hydroxyl groups and hydrogens of the sugar has a tremendous impact in the interaction. (b) The solvation/desolvation process plays a major role in complex formation. The architecture and chemical nature of the binding site is also very important. (Juan et al., 2013).

2.7   Glucose tolerance of β-glucosidases

Glucose tolerant β-glucosidases are of particular interest. These β-glucosidases often show peculiar kinetics, including inhibitory effects of substrates and activating effects of inhibitors, such as glucose or xylose (Kuusk and Valjamae, 2017). The mechanisms behind the activating/inhibiting effects are still poorly understood. Knowledge on the mechanisms of the activating/inhibiting effects of inhibitors helps the rational engineering and selection of β-glucosidases for biotechnological applications. Product inhibition of β-glucosidases by glucose is considered to be a limiting step in enzymatic hydrolysis of biomass, especially when higher sugar yields are required. The liberated cellobiose is acted by β-glucosidases, which convert it to glucose. Cellobiose accumulation inhibits the exoglucanase activity by the product inhibition and reduces the overall saccharification rate; therefore, its conversion into glucose via β-glucosidases is necessary. Moreover accumulated glucose reversibly inhibits the β-glucosidases activity. Thus, β-glucosidase is the rate-limiting enzyme in cellulose bioconversion (Singhania et al., 2013) and determines the rate of forward reactions. Glucose tolerance of β-glucosidases is a significant desirable property, as it is capable of decreasing the cellulolytic enzyme loads needed to attain acceptable yields during lignocellulose hydrolysis, thus reducing the costs of bioethanol production (Singhania et al., 2013). Some β-glucosidases belonging to the GH family 1 (GH1) exhibit properties of being tolerant to or even stimulated by high glucose concentrations. Glucose tolerance in β-glucosidases is sometimes coupled to a stimulatory effect of glucose on substrate degradation in a given range of glucose concentrations (up to 1 M). This property seems to be exclusive to some GH1 enzymes in which glucose binds to a secondary site beside the active site and further

stimulates the β-glucosidase hydrolysis activity (Cantarel et al., 2009; Henrissat, 1991; Matsuzawa et al., 2016). Furthermore, family GH3 β-glucosidases are not known to be glucose tolerant and are inhibited by glucose. However, there are reports where GH3 β-glucosidases have shown glucose tolerance up to 400 mM glucose (Ramani et al., 2015).

The limitation of glucose inhibition can also be tackled by removing glucose when it is produced, for which technologies such as membrane filters should be used to separate monomers of sugars, so as to avoid feedback inhibition. Simultaneous saccharification and fermentation could also be an answer to this limitation; where glucose is converted to ethanol simultaneously hence glucose remains always at lower concentration (Singhania et al., 2017). Nonetheless, the feedback inhibition of β-glucosidases is a major challenge to circumvent, so we need glucose tolerant β-glucosidases.

2.8 Glycoside hydrolase GH116 family

The story of the GH116 family began when one cDNA for human bile acid β-glucosidase was cloned, and it was noted that the protein belonged to a family of unknown proteins from animals, plants and cyanobacteria with 34-80% amino acid sequence identity (Matern et al., 2001). The GH116 family was named upon characterization of a β-glycosidase from the hyperthermophilic archaeon Sulfolobus solfataricus that is a divergent member of this group (Cobucci-Ponzano et al., 2010), which demonstrated the family contains members from all three domains of living organisms. The mammalian bile acid β-glucosidase was identified as the non-lysosomal glucocerebrosidase, GBA2 (also known as glucosylceramidase EC

3.2.1.45, and β-glucosidase, EC 3.2.1.21) and is currently the best characterized member of this family. Human non-lysosomal bile acid β-glucosidase, GBA2, is involved in the catabolism of glucosylceramide, which may then converted to sphingomyelin (Scheme 2.1) (Boot et al., 2007; Jatooratthawichot et al., 2020). GBA2 is localized to the cytoplasmic surface of the membranes of the ER and Golgi body, where it breaks down glucosylceramide (Körschen et al., 2013).The β-glycosidase (SSO1353) from S. solfataricus is a β-xylosidase (EC 3.2.1.37) and β-glucosidase specific for the β-glucosides and β-xylosides with hydrophobic aglycones (Cobucci-Ponzano et al., 2010). It was found to hydrolyze these following a retaining reaction mechanism. A GH116 β-N-acetylglucosaminidase from S. solfataricus (SSO3039) (Ferrara et al., 2014) was shown to act as a bifunctional β-glucosidase/β-N-acetylglucosaminidase, which increased our knowledge about the diversity in this family.

Scheme 2.1 Map of sphingolipid metabolism showing hexosylceramide hydrolysis to release ceramides and their subsequent conversion to other species, which is affected by GBA2 overexpression (Jatooratthawichot et al., 2020).

2.9   Thermoanaerobacterium xylanolyticum *Tx*GH116 β-glucosidase

Thermoanaerobacterium xylanolyticum GH116 (TxGH116) was the first bacterial member of GH116 to be characterized (Charoenwattanasatien et al., 2016; Sansenya et al., 2015). The structure of TxGH116 consists of an N-terminal domain, primarily formed by a two-β-sheet sandwich, and a C-terminal (α/α)6 solenoid domain (Figure 2.8 A). The C-terminal domain contains the residues previously proposed as the catalytic nucleophile and general acid/base in the archaeal β-glycosidase (SSO1353) from Sulfolobus solfataricus and human GBA2 (Glucosylceramidase 2) (Cobucci-Ponzano et al., 2010; Kallemeijn et al., 2014). Through functional analysis of TxGH116, Charoenwattanasatien et al. (2016) provided detailed insight into the glucose-binding residues of TxGH116. The catalytic nucleophile, E441, lies near the end of a long loop between the first (H5) and second (H6) α-helices of the C-terminal domain, while the catalytic acid/base, D593, lies in a long loop between the fifth (H9) and sixth (H10) helices of the solenoid (Figure 2.8 A). This loop also contains the binding site for a structural Ca^{2+} ion (Figure 2.8 A). Recombinant TxGH116 shows high active on 4-nitrophenyl (4NP) β-D-glucoside (4NPGlc) and is also active on other glucosides and glucooligosaccharides. Michaelis-Menten kinetics determined at 60ºC (the temperature of the Frying Pan Hot Springs, from which T. xylanolyticum was isolated) revealed that 4NPGlc can be hydrolyzed 17 times more efficiently, in terms of k_{cat}/K_M, than the corresponding β-D-galactoside (4NPGal). TxGH116 shows high activity at pH 5.5 toward both β-1,3- and β-1,4-linked glucooligosaccharides, with similar k_{cat} and K_M values for both cellobiose and laminaribiose. Since T. xylanolyticum was isolated based on its growth on plant cell wall polysaccharides, these oligosaccharides are likely natural substrates for
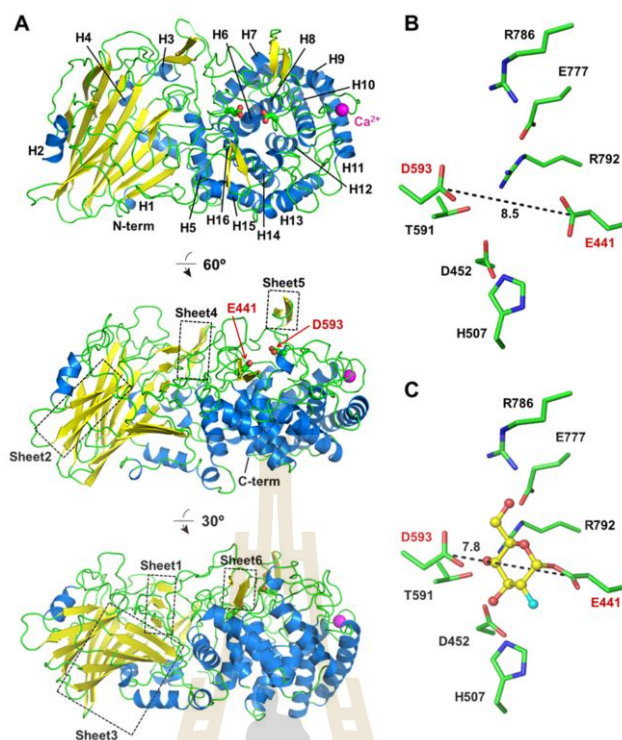
Figure 2.8 Three-dimensional structure of *Tx*GH116 β-glucosidase. (A) Three views of the three-dimensional structure of *Tx*GH116. The N-terminal domain consists of β-sheets (sheets 1-3) surrounded by α-helices (H1-H4). The C-terminal (α/α)$_6$ solenoid domain contains the active site and consists of 12 α-helices (H5-H16) in six outer helix plus inner helix pairs, with H5 situated between the two domains. A long loop with a two-strand β-sheet (sheet 5) lies between helices H7 and H8, near the active site (E441 and D593). α-Helices are shown in blue and β-strands in yellow. (B) Active site residues of the native structure of *Tx*GH116, showing the distance between the catalytic nucleophile (E441) and acid/base (D593) carbons (red labels). (C) Active site of the *Tx*GH116 covalent intermediate complex with 2-deoxy-2-fluoroglucose, showing the distance between the catalytic residues is decreased in this complex. Distances are in Å (Charoenwattanasatien et al., 2016).

the enzyme. Degradation of these substrates makes it valuable for application to second generation biofuel production by breaking down glucose oligomers and cellobiose, producing glucose and alleviating the inhibition of cellobiohydrolases and endoglucanases by these oligosaccharides during enzymatic hydrolysis (Decker et al., 2000; Yun et al., 2001).

2.10 Amino acid residues involved in glucose binding and hydrolysis mechanism in the *Tx*GH116 active site

GH116 family enzymes act via a double-displacement mechanism in which a covalent glycosyl enzyme intermediate is formed and subsequently hydrolyzed via oxocarbenium-ion like transition states. This mechanism relies on two key residues: a general acid/base and a nucleophile. The putative acid/base of TxGH116, D593, is located approximately 8 Å (carboxylate to carboxylate) away from the nucleophile (E441; Figure 2.8) in the P21212 structure and 10 Å away in the P61 structure due to a displacement of the loop containing D593 by 2 Å (Figure 2.8, B, C). This is a longer distance than that was observed in most retaining β-glucosidases, which typically have the acid/base within 4.5-6.5 Å of the nucleophile (Barrett et al., 1995; Davies et al., 1998; Dvir et al., 2003; Varghese et al., 1999). Therefore, Charoenwattanasatien et al. (2016) verified that the enzyme is retaining by identifying the initial hydrolysis product as β-D-glucose, which retains the anomeric stereochemistry of the substrate, by monitoring the reaction time course by NMR spectroscopy. Moreover, they demonstrated that E441 is the catalytic nucleophile and D593 the catalytic acid/base by mutation of these residues to alanine, followed by chemical rescue using small nucleophiles.

The amino acid residues are capable of forming hydrogen bonding interactions between TxGH116 and the substrate glucosyl moiety include E441 at OH2; H507 and D452 at OH3; D452, T591, and R792 at OH4; and E777 and R786 at the OH6. In the last case, OH6 appears to mediate a salt bridge between E777 and R786 (Figure 2.9). The important role of aromatic residues in slot like active site subsites +1 and +2 also indicates their stacking function in substrate and product molecule binding (Figure 2.10). Based on this structural information and the thermal stablility of TxGH116, it is of interest to test the functional importance of each of these residues and test whether the modification of these glucose-binding residues can improve the glucose tolerance, which would be valuable for industrial application to biomass conversion.
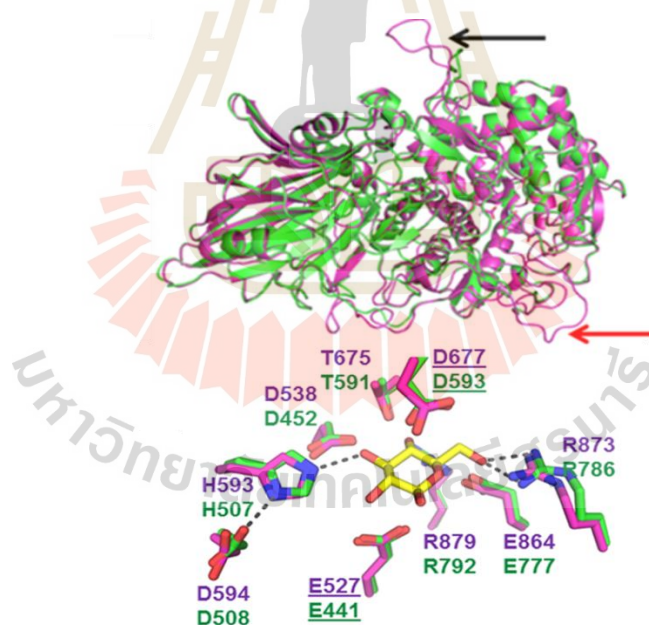
Figure 2.9 Superposition of active site residues of human GBA2 and the *Tx*GH116 complex with glucose. All of the sugar binding residues are conserved between *Tx*GH116 and GBA2. The catalytic nucleophile and acid/base labels are underlined. (Charoenwattanasatien et al., 2016).
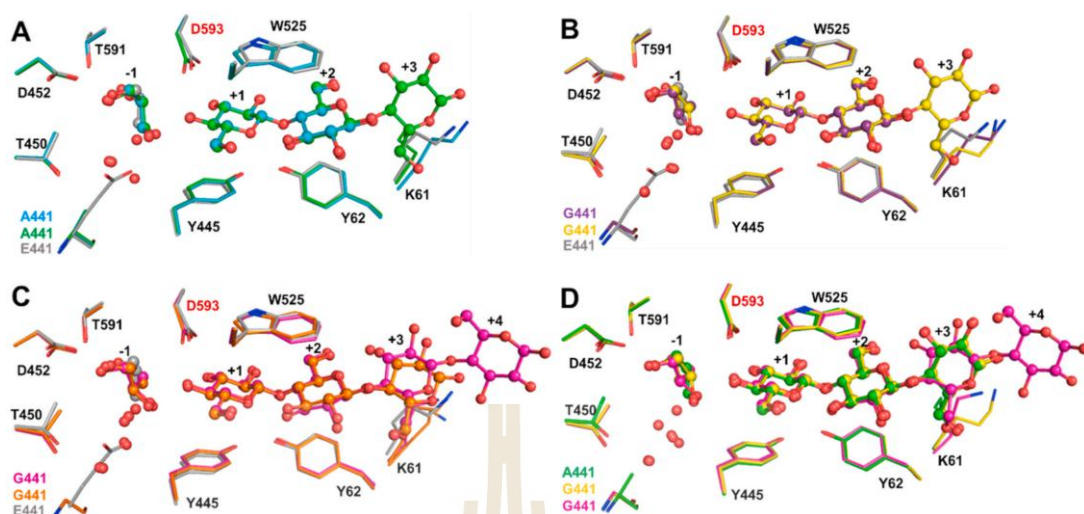
Figure 2.10 Comparison of the active sites of the *Tx*GH116 E441A and E441G

glycosynthases with cellooligosaccharides and free *Tx*GH116 with

subsite+1 and +2 aromatic residues for stacking interaction supporting.

(A) The superimposition of E441A with cellobiose (PDB: 7DKU, cyan

carbons) and cellotriose (PDB: 7DKV, green carbons) and free wild type

enzyme (PDB: 5BVU, grey carbons). (B) The superimposition of E441G

with cellobiose (PDB: 7DKX, violet carbons) and cellotriose (PDB:

7DKY, yellow carbons) and free wild type enzyme (PDB: 5BVU, grey

carbons). (C) The superimposition of E441G with α-GlcF (PDB:

7DKW) molecule A, which exhibits 4 glucosyl residues (pink carbons),

and molecule B, which exhibits 3 glucosyl residues (orange carbons),

from autocondensation products and free wild type enzyme (PDB:

5BVU, grey carbons). An active site water molecule is bound at the

same location as the catalytic nucleophile in the free wild type enzyme

in each glycosynthase structure. (D) The superimposition of E441A and

E441G with cellotriose (PDB: 7DKV green and PDB: 7DKY yellow

carbons, respectively) and E441G with 4 residues of autocondensation product (PDB: 7DKW Mol A, pink carbons). The glycerol and cellooligosaccharides are represented as balls and sticks and the water molecules are represented as balls. Oxygen is shown in red, nitrogen in blue (Pengthaisong et al., 2021).

2.11 Protein engineering techniques: site directed mutagenesis and directed evolution

Site directed mutagenesis (SDM) is an important tool to engineer the structures and functions of proteins. It can cause single base changes, insertions of DNA, or deletions, in a specific manner. Therefore, SMD is able to probe the importance of specific residues, such as their roles in enzyme catalysis (Edelheit et al., 2009; Madhavan et al., 2017). A method based on PCR using a thermostable DNA polymerase and a plasmid vector as the template, and mutagenic oligonucleotide primers containing the desired mutation complementary to both strands of a target sequence is currently popular (Braman et al., 1996). In this method, the primers anneal to the DNA template, which replicates the plasmid DNA with the mutations to produce a product containing a strand break. The mixture mutant and parental DNA plasmids are then treated with *Dpn*I to remove the methylated parental template DNA from the newly synthesized demethylated mutant DNA, and the remaining DNA is transformed into *E. coli* cells where the nick is sealed by host repair enzymes (Ishii et al., 1998).

Improvement of enzyme activity or identification of residues involved in catalysis can also be done by directed evolution, which may modify the mechanism,

specificity or activity (Guo et al., 2016; Kuchner and Arnold, 1997; Madhavan et al., 2017; Singhania et al., 2017). This can be screened for increased thermostability and stability compared to the wild type enzymes (Palackal et al., 2004; Sindhu et al., 2017). A GH36 α-glycosidase from *Thermus thermophilus* was converted to α-transglycosidase by random mutation to enhance the rate and substrate specificity of the natural transglycosylation activity (Teze et al., 2015). In the GH1 family, the site directed mutagenesis helped the provide indications of the native enzymes' catalytic action and determinants of specificity, and the reasons for the changes observed in the mutants' enzymatic activity (Dopitova et al., 2008). The acid/base and nucleophile glycosynthase mutants of *Tx*GH116 have been recently generated by site directed mutagenesis and applied to product interesting products like 1,2,3-triazole-glucoside inhibitors (Gorantla et al., 2019) and pNP-oligosaccharide (Pengthaisong et al., 2021), which make it to be potential industrial application.

CHAPTER III

MATERIALS AND METHODS

3.1   Chemicals and reagents

Chelating sepharose fast flow resin, lysozyme, N,N'-methylene-bisacrylamide, N,N',N'',N'''-tetramethylethylenediamine (TEMED) and Triton X-100 were purchased from GE Healthcare (Uppsala, Sweden). Bromophenol blue, Coomassie Brilliant Blue R-250, sodium acetate, sodium carbonate, sodium chloride (NaCl), sodium dodecyl sulfate (SDS), sodium hydroxide, disodium ethylenediamine tetraacetate (Na2-EDTA), sulfuric acid, Tris base, methanol (MeOH), ethanol, glacial acetic acid, glycine and ethyl acetate (EtOAc) were purchased from Carlo ERBA (Milano, Italy). Acrylamide, imidazole, bovine serum albumin (BSA), dichlorodimethylsilane, polyethylene glycol monomethyl ether (PEG MME) 3000, 2-(N-morpholino)-ethanesulfonic acid (MES) and ethidium bromide were purchased from Fluka (Steiheim, Swizerland). Ammonium sulfate, calcium chloride and thin layer chromatography silica gel 60 aluminum F254 plates for oligosaccharide detection were purchased from Merck (Darmstadt, Germany). Isopropyl thio-β-D-galactoside (IPTG) was purchased from United States Biological (Swampscott, MA, USA). 2,2′-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS), PGO enzymes which contains 500 units of glucose oxidase (Aspergilus niger), 100 purpurogallin units of peroxidase (horseradish), and buffer salts, ammonium persulfate, ampicillin, tetracycline, DNase I, Bis-Tris, 2-mercaptoethanol, pNP-α-L-arabinopyranoside

(*p*NPAra), *p*NP-β-D-galactopyranoside (*p*NPGal), *p*NP-β-D-xylopyranoside (*p*NPXyl), *p*NP-β-D-cellobioside (*p*NPC2), cellobiose and phenylmethylsulfonylfluoride (PMSF) were purchased from Sigma Aldrich (St. Louis, MO, USA), 4-Nitrophenyl β-D-glucopyranoside (4NPGlc) was purchased from Shanghai Laihao Trade Co. Ltd. (Shanghai, China). Cellooligosaccharides of degree of polymerization (DP) 3-5 were purchased from Seikagaku Kogyo Co. (Tokyo, Japan). Kanamycin sulfate was purchased from Ameresco (Solon, OH, USA). Agarose was purchased from Invitrogen (Carlsbad, CA, USA). Deoxynucleoside triphosphates (dNTPs) were purchased from Promega (Madison, WI, USA). Peptone, yeast extract and bacto-agar were purchased from HiMedia (Mumbai, India). Ultrafiltration membranes (Centricon, 30 kDa MW and 10 kDa MW cut off) and Ultrafree MC 0.22 μm filters were purchased from Millipore Corporation (Bedford, MA, USA). High vacuum grease was purchased from Dow Corning (Midland, MI, USA). The QuikChange® Site-Directed Mutagenesis Kit was purchased from Stratagene-Agilent (La Jolla, CA, USA) and QIAprep spin miniprep plasmid extraction kit was purchased from QIAGEN (Hilden, Germany). HPLC-grade distilled water and acetonitrile were purchased from RCI Labscan (Bangkok, Thailand). Other chemicals and molecular reagents used but not listed here were purchased from a variety of suppliers.

3.2 General methods

3.2.1 Bioinformatic identification of essential active-site residues in GH116 family

Amino acid conservation analysis was performed to assess the conserved residues in the GH116 enzyme family. The crystal structure of *Tx*GH116

(PDB ID: 5BVU) was used as the initial input to a Pfam (Finn et al., 2006) search to identify its protein family. The enzyme *Tx*GH116 is assigned to the glycoside hydrolase family 116 with 2263 sequence homologs residing in the clan. To filter out redundant catalytic domain sequences, 34 seed sequences were taken from Pfam noline alignment. After aligning the remaining sequences using MEGA 8.0, they were further curated to remove the proteins that do not possess the residues forming the catalytic residues associated with enzyme function. This series of bioinformatics operations produced a final set of 34 protein sequences for analysis. An active-site sequence logo of the GH116 family was generated by WebLogo (http://weblogo. berkeley.edu/).

3.2.2 Primers designment for mutagenesis

The essential amino acid residues that related with glucose binding and catalysis was underlined from the known structure of *Tx*GH116 and glucose complex. We identified them into four interaction groups: Acid/base mechanism interactions, nucleophile interactions, glycone sugar binding interactions and subsite +1 to +2 interactions.

Site-directed mutagenesis was performed on the previously reported on recombinant pET30a(+)/*Tx*GH116 plasmid (Charoenwattanasatien et al., 2016) by the QuikChange method (Stratagene, Agilent Corp.). The mutagenic primers for the corresponding amino acid residue mutations are as follows tables and was separated into 4 groups by their different interactions with glucose (the mutated nucleotides are in bold type):

Table 3.1 Acid/base mechanism & Interactions.

Primer name	Primer sequence
TxGH116_Y523F_for	5'- GATTAAAATCAACGCTTTCGACTGGCAGAACCCGAATATC -3'
TxGH116_Y523F_rev	5'- TCGGGTTCTGCCAGTCGAAAGCGTTGATTTTAATCCAC -3'
TxGH116_Q727A_for	5'- GTTGACGAATCTGATATTGCGGCTCAAGAAGTGTGGAC -3'
TxGH116_Q727A_rev	5'- GTCCACACTTCTTGAGCCGCAATATCAGATTCGTCAAC -3'
TxGH116_Q727E_for	5'- GTTGACGAATCTGATATTGAAGCTCAAGAAGTGTGGAC -3'
TxGH116_Q727E_rev	5'- GTCCACACTTCTTGAGCTTCAATATCAGATTCGTCAAC -3'

Table 3.2 Nucleophile interactions.

Primer name	Primer sequence
TxGH116_Y790F_for	5'- TTATCGCGCAAGTATGTTCATGCGTCCGCTGAGTATTTGG -3'
TxGH116_Y790F_rev	5'- ATACTCAGCGGACGCATGAACATACTTGCGCGATAATTGC -3'
TxGH116_T450A_for	5'- CGACTACAATTATTACGAAGCACTGGATGTTCGTTTTTATGGC -3'
TxGH116_T450A_rev	5'- CATAAAAACGAACATCCAGTGCTTCGTAATAATTGTAGTCGAAG -3'

All of the mutated genes will be sequenced (Macrogen) to confirm that only the desired mutations were inserted.

3.2.3 Preparation of competent cells of *E. coli* strains DH5α and BL21 (DE3)

Bacteria from glycerol stocks were streaked onto a Lennox broth (LB, 10 g/l tryptone, 5 g/l yeast extract, 5 g/l sodium chloride) agar plates containing 15 µg/ml kanamycin for BL21(DE3) or without antibiotic for DH5α. A single colony was picked into 5 ml LB media containing antibiotics as listed above. The cells were grown at 37°C for 16-18 hr with shaking at 200 rpm. Then, 0.5 ml of starter culture was inoculated into 50 ml LB in a 250 ml flask and incubated at 37°C with shaking until the OD600 reached 0.3-0.4. The culture was cooled on ice for 15 min, then transferred

Table 3.3 Glycone sugar binding interactions related residues and their response mutations.

Primer name	Primer sequence
TxGH116_E777Q_for	5'- CTATTGGTTTCGTACCCCGCAGGCCTGGACGAAAGATG -3'
TxGH116_E777Q_rev	5'- CATCTTTCGTCCAGGCCTGCGGGGTACGAAACCAATAG -3'
TxGH116_E777A_for	5'- GTTTCGTACCCCGGCGGCCTGGACGAAAGATG -3'
TxGH116_E777A_rev	5'- CATCTTTCGTCCAGGCCGCCGGGGTACGAAAC -3'
TxGH116_R786K_for	5'- ACGAAAGATGGCAATTATAAAGCAAGTATGTATATGCGTCCGCTG -3'
TxGH116_R786K_rev	5'- CGCATATACATACTTGCTTTATAATTGCCATCTTTCGTCCAGGC -3'
TxGH116_R786A_for	5'- CGAAAGATGGCAATTATGCGGCAAGTATGTATATGCGTCCGC -3'
TxGH116_R786A_rev	5'- GCATATACATACTTGCCGCATAATTGCCATCTTTCGTCCAGG -3'
TxGH116_W732F_for	5'- CAGGCTCAAGAAGTGTTCACCGGTGTTACGTATGCAC -3'
TxGH116_W732F_rev	5'- CATACGTAACACCGGTGAACACTTCTTGAGCCTGAATATC -3'
TxGH116_E730A_for	5'- CTGATATTCAGGCTCAAGCGGTGTGGACCGGTGTTAC -3'
TxGH116_E730A_rev	5'- GTAACACCGGTCCACACCGCTTGAGCCTGAATATCAG -3'
TxGH116_E730Q_for	5'- CTGATATTCAGGCTCAACAGGTGTGGACCGGTGTTAC -3'
TxGH116_E730Q_rev	5'- GTAACACCGGTCCACACCTGTTGAGCCTGAATATCAG -3'
TxGH116_R792A_for	5'- GCAAGTATGTATATGGCCCCGCTGAGTATTTGGAGTATGG -3'
TxGH116_R792A_rev	5'- CTCCAAATACTCAGCGGGGCCATATACATACTTGCGC -3'
TxGH116_R792K_for	5'- CGCAAGTATGTATATGAAACCGCTGAGTATTTGGAGTATGGAAG -3'
TxGH116_R792K_rev	5'- CCAAATACTCAGCGGTTTCATATACATACTTGCGCGATAATTGC -3'
TxGH116_T591A_for	5'- CATCCCGGACCAGGCGTACGATACGTGGTCAATG -3'
TxGH116_T591A_rev	5'- GACCACGTATCGTACGCCTGGTCCGGGATGCCTTC -3'
TxGH116_D452A_for	5'- ATTACGAAACCCTGGCGGTTCGTTTTTATGGCAGCTTC -3'
TxGH116_D452A_rev	5'- CTGCCATAAAAACGAACCGCCAGGGTTTCGTAATAATTG -3'
TxGH116_D452N_for	5'- ATTACGAAACCCTGAACGTTCGTTTTTATGGCAGCTTCC -3'
TxGH116_D452N_rev	5'- GCTGCCATAAAAACGAACGTTCAGGGTTTCGTAATAATTG -3'
TxGH116_H507A_for	5'- CAGGGTATGATTCCGGCTGATCTGGGCTCATCGTACG -3'
TxGH116_H507A_rev	5'- GATGAGCCCAGATCAGCCGGAATCATACCCTGGAC -3'
TxGH116_H507Q_for	5'- CAGGGTATGATTCCGCAGGATCTGGGCTCATCGTACG -3'
TxGH116_H507Q_rev	5'- GATGAGCCCAGATCCTGCGGAATCATACCCTGGAC -3'
TxGH116_H507E_for	5'- CAGGGTATGATTCCGGAGGATCTGGGCTCATCGTACG -3'
TxGH116_H507E_rev	5'- GATGAGCCCAGATCCTCCGGAATCATACCCTGGAC -3'

Table 3.4 Subsite +1 to +2 interactions.

Primer name	Primer sequence
TxGH116_Y445L_for	5'- CTGCTGGAATGCTTCGACTTGAATTATTACGAAACCCTGG -3'
TxGH116_Y445L_rev	5'- GGTTTCGTAATAATTCAAGTCGAAGCATTCCAGCAGACC -3'
TxGH116_W525F_for	5'- AAATCAACGCTTATGACTTCCAGAACCCGAATATCTGGAAAG -3'
TxGH116_W525F_rev	5'- CAGATATTCGGGTTCTGGAAGTCATAAGCGTTGATTTTAATCC -3'
TxGH116_W525L_for	5'- AAATCAACGCTTATGACCTTCAGAACCCGAATATCTGGAAAG -3'
TxGH116_W525L_rev	5'- CAGATATTCGGGTTCTGAAGGTCATAAGCGTTGATTTTAATCC -3'
TxGH116_D593A2_for	5'- GACCAGACCTACGCGACGTGGTCAATGAAAGGC -3'
TxGH116_D593A2_rev	5'- TTTCATTGACCACGTCGCGTAGGTCTGGTCCG -3'

to pre-cooled centrifuge tubes, and the cells were collected by centrifugation at 3,000 rpm for 10 min at 4°C. Cells were gently and slowly resuspended with 10 ml of cold 0.1 M CaCl2 and placed on ice for 20 min. The cells were centrifuged as before and resuspended with 0.5 ml of cold 0.1 M CaCl2 and placed on ice for 1 hr. Finally, cold glycerol was added to the cells to the final concentration of 15%. Then, they were resuspended and 50 μl aliquots were stored in pre-cooled tubes into liquid nitrogen and then kept at -80°C.

3.2.4 Mutagenesis with the QuikChange® Site-Directed Mutagenesis Kit

The QuikChange® Site-Directed Mutagenesis Kit (Stratagene) was used to construct the mutants of *Tx*GH116 amino acid residues involved in glycone sugar binding, subsites+1 and +2, and interactions with the catalytic nucleophile and acid/base. Site-directed mutagenesis was performed on the previously reported on recombinant pET30a(+)/*Tx*GH116 plasmid (Figure 3.1) (Charoenwattanasatien et al., 2016). The pET30a(+) *Tx*GH116 plasmid was used as a template for amplification of full-length plasmid strands using two overlapping oligonucleotide primers containing the desired mutation (Table 3.1, 3.2, 3.3, 3.4). The mutagenic oligonucleotide primers

were specifically designed according to the criteria of the QuikChange manual to have lengths of 25-45 bases with Tm $\geq 78°C$. The Tm was calculated from following formula: Tm $= 81.5 + 0.41$ (%GC) $- 675/N -$ %mismatch, where N is the primer length in nucleotides, and %GC and %mismatch are whole numbers.

$PfuTurbo$ DNA polymerase (with proofreading activity) was used to polymerize the mutated plasmid DNA during the temperature cycling, which included step 1, 95°C 30 s; step 2, 95°C 30 s; step 3, 55°C 15 min and step 4, 68°C 7.5 min, with steps 2 to 4 repeated for 16 cycles. The PCR products were treated with DpnI endonuclease to eliminate methylated and hemimethylated DNA of the parental DNA template. Repair of the nicked circular dsDNA products was performed by transformation into competent DH5α cells. The transformants were selected on agar plates containing kanamycin and were screened for the target mutation by DNA sequencing.

3.2.5 Commercial plasmid extraction and purification

The GF1 plasmid DNA extraction kit was used for plasmid DNA extraction from $E.$ $coli$ strain DH5a according to the manufacturer's instructions. Each single colony was picked into 5 mL LB broth with 15 μg/mL of kanamycin, to select for the pET30a(+) TxGH116 plasmid, which contains the kanamycin resistance gene. The culture was grown overnight for 16 hours at 37°C with agitation. Cell pellets from 5 mL of bacterial culture were collected by centrifugation at 6,000 x g for 2 min at room temperature, and the medium was completely removed and discarded. The pellet was resuspended in 250 μL of S1 reagent by pipetting. Then, 250 μL of S2 reagent was added into the suspension and the tube was gently mixed by inverting for 4-6 times. To neutralize the lysate, 400 μL of NB buffer was added, then gently mixed

by inverting the tube for 6-10 times until a white precipitate formed. The tube was centrifuged at 13,000 x g for 10 min. The supernatant was transferred into a column assembled in a clean collection tube, then centrifuged at 13,000 x g for 1 min, and the flow through was discarded. Then, 700 µL of wash buffer was added into the column, which was then centrifuged at 13,000 x g for 1 min, and the flow through was discarded. The empty column was centrifuged at 13,000 x g for 1 min to remove residual ethanol, then the column was placed into a sterile 1. 5 mL tube. Then, 50 µL of elution buffer was added directly onto the column membrane and the column was left to stand for 1 min. Finally, the column was centrifuged at 10,000 x g for 1 min to elute the DNA. The DNA sample was stored at -40°C for further use.
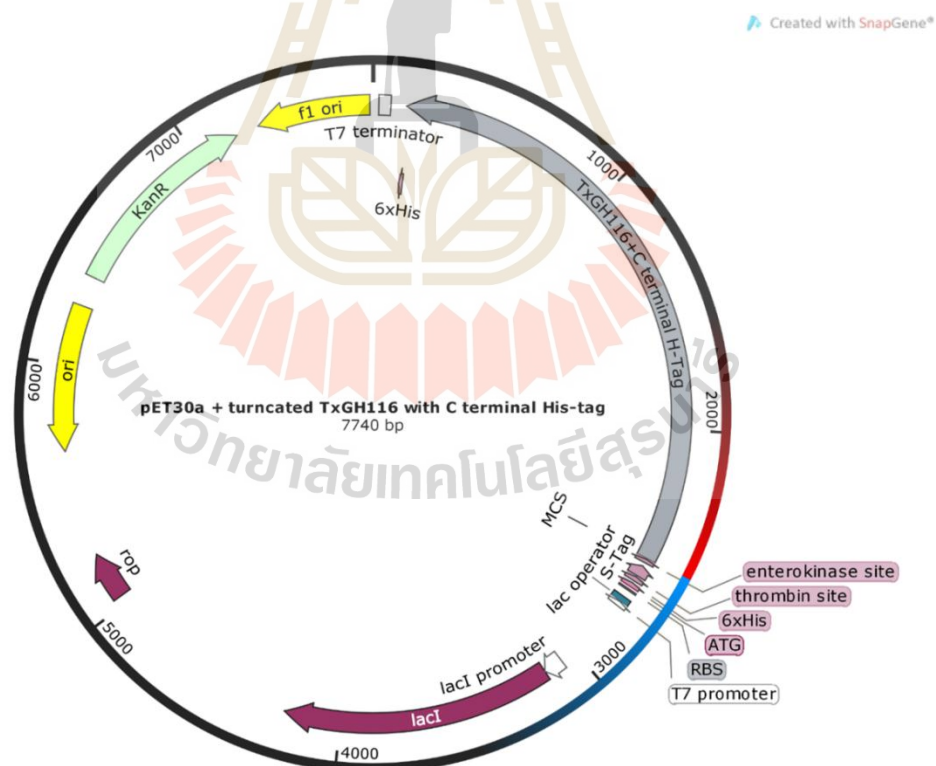
Figure 3.1 Construct of the protein-coding sequence of recombinant pET30a(+) with the turncated *Tx*GH116 inserted after the enterokinase cleavage site, with both N and C-terminal His-tags.

3.2.6 Transformation of plasmid DNA into competent cells

Super coiled plasmid DNA (50-100 ng) was added into the 50 µl of competent cells, mixed by tapping and immediately place on ice for 30 min. The cells were heat shocked at 42°C for exactly 60 s and then immediately put on ice for 3 min. Then, 0.2 ml of LB medium was added into the cell mixture and the tube was incubated in a 37°C shaker in 200 rpm for 1 hr. Then, 100-200 µl of the cells were spread on an LB plate containing the appropriate antibiotics and incubated at 37°C overnight.

3.2.7 SDS-PAGE electrophoresis

Polyacrylamide gels were prepared from 5 mL of 10% separating gel consisting of 2 mL of distilled water (DW), 1.25 mL of 1.5 M Tris-HCl, pH 8.8, 1.65 mL of 30% acrylamide, 50 µL of 10% SDS, 25 µL of 10% ammonium persulfate (APS) and 5 µL of TEMED. The 2.5 mL of 4% stacking gel consisted of 1.5 mL of DW, 625 µL of 0.5 M Tris-HCl pH 6.8, 332.5 µL of 30% acrylamide, 25 µL of 10% SDS, 12.5 µL of 10% APS and 5 µL of TEMED. After adding the APS and TEMED, the separating gel was poured into the gel plates, covered with butanol and left to harden for 20-30 min for setting. Similarly, the stacking gel was poured on top of the rinsed stacking gel, the comb added, and the gel left to harden. The two gel polymerizations were done in a Hoeffer gel electrophoresis cassette (GE Healthcare). Fifteen microliters of protein sample was mixed with 5 µl of 5x denaturing sample buffer (0.05 M Tris-HCl, pH 6.8, 10% (w/v) SDS, 50% (v/v) glycerol, 0.5 mg/ml bromophenol blue, 20% (v/v) 2-mercaptoethanol) and boiled at 100°C for 5 min. The sample was spun down and loaded onto the gel under running buffer (25 mM Tris, 192 mM glycine, 0.1% (w/v) SDS) and placed in an electric field of 140 volts for 1 hr

and 20 min. The protein bands were detected with Coomassie brillant blue staining solution (0.1% (w/v) Coomassie brilliant blue R-250, 40% (v/v) MeOH, 10% (v/v) glacial acetic acid) and the blue background was washed out with destaining solution (40% (v/v) MeOH, 10% (v/v) glacial acetic acid). The migration distance of the protein band was compared to the low molecular weight calibration electrophoresis standards (GE Healthcare); phosphorylase b (96 kDa), bovine serum albumin (66 kDa), ovalbumin (45 kDa), bovine carbonic anhydrase (30 kDa), soy bean trypsin inhibitor (20.1 kDa), and bovine milk α-lactalbumin (14.4 kDa).

3.2.8 Protein expression and purification of *Tx*GH116 mutations

The recombinant pET30a(+)/*Tx*GH116 plasmid and its mutated variants were transformed into BL21(DE3) competent cells and a single colony that was grown overnight was picked and inoculated into LB media containing the appropriate antibiotics to make a starter culture. To express recombinant *Tx*GH116, 1% final concentration of starter culture was added into the same type of media and cultured at 37ºC with rotary shaking at 200 rpm until the optical density at 600 nm of the culture reaches 0.4.

The optimum expression conditions was performed with the final concentration of IPTG at 0.3 mM, and the temperature at 20ºC as previous described (Charoenwattanasatien et al., 2016). The cell pellets was collected by centrifugation at 4000 rpm at 4°C for 20 min. The cell pellets was kept at -80ºC to allow freeze-thaw breakage before use. The IPTG-induced bacterial cell pellets were thawed on ice and then resuspended in freshly prepared extraction buffer (20 mM Na_2HCO_3, pH 7.4, 150 mM NaCl, 200 µg/ml lysozyme, 1% Triton-X 100, 1 mM phenylmethylsulfonylfluoride (PMSF), 25 µg/ml DNase I, 1 mM 6-aminohexanoic acid and 1 mM benzamidine

hydrochloride) in a ratio of 5 ml extraction buffer per gram fresh weight of cell pellets. The resuspended cells were incubated at room temperature for 60 min. Then, the insoluble debris was removed by centrifugation at 13,000 rpm for 30 mins.

Considering about the thermostability of *Tx*GH116 protein, the supernatant fraction was heated by gradient temperature from minimum 30 to maximum 90ºC at 5 or 10ºC intervals and then centrifuged to remove the unstable protein, and the supernatant was loaded onto a pre-equilibrated immobilized metal ion affinity

chromatography (IMAC) column (GE Healthcare, Buckinghamshire, United Kingdom) charged with Co^{2+} (from $CoCl_2$). After loading the cell extract supernatant, the IMAC column was washed twice with 10 column volumes (CV) of equilibration/wash buffer (20 mM Na_2HCO_3, pH 7.4, 150 mM NaCl) to remove unbound protein and washed again with 5 CV each of equilibration/wash buffer containing 10 and 20 mM of imidazole, respectively. Bound protein was eluted with 5 CV of elution buffer (20 mM Na_2HCO_3, pH 7.4, 150 mM NaCl containing 500 mM imidazole). The eluted protein was checked by assaying activity with *p*NPGlc and the presence and purity of protein of appropriate size was evaluated by SDS-PAGE. The fractions that contained target protein were pooled and imidazole removed by exchanging the buffer with 20 mM Tris-HCl, pH 8.0, 150 mM NaCl in a 30 kDa Amicon® Ultra-15 (Merck-Millipore, Billerica, MA, USA).

The N-terminal thioredoxin and histidine tags was removed by digestion with 0.1 µl of enterokinase per 1 mg of total protein, according to supplier's protocol. The released fusion tag was removed by adsorption to a second IMAC column. The unbound and low-stringency wash fractions from IMAC was checked for β-

glucosidase activity and protein was evaluated by SDS-PAGE. The fractions that contained β-glucosidase were pooled and the imidazole removed by exchanging the buffer in a 30 kDa MWCO Amicon® Ultra-15, as before.

The purified protein concentration was determined by measuring its absorbance at 280 nm. Extinction coefficient at 280 nm was calculated as ε280 of 113320 $M^{-1}cm^{-1}$ for most of the *Tx*GH116 proteins and ε280 of 112040 $M^{-1}cm^{-1}$ for mutants without N-terminal tag, by the method of Gill and von Hippel (1989), and used to calculate the enzyme concentrations. *Tx*GH116 contains 29 tryptophan (Trp) residues, 55 tyrosine (Tyr) residues and 6 cystine residues. The equation of Gill and von Hippel is:

$$\varepsilon protein\ (M^{-1}cm^{-1}) = a\ \varepsilon trp + b\ \varepsilon tyr + c\ \varepsilon cystine \text{ , where } \varepsilon trp = 5690\ M^{-1}cm^{-1}, \varepsilon tyr = 1280\ M^{-1}cm^{-1}, \varepsilon cys = 120\ M^{-1}cm^{-1}\ ;$$

a = amount of tryptophan residues

b = amount of tyrosine residues

c = amount of cystine residues

3.3   Experimental procedures

3.3.1   Analysis of substrate specificity and kinetics for hydrolysis of synthetic and natural substrates by each *Tx*GH116 variant containing a mutation

The purified TxGH116 variant proteins were used for enzymatic characterization. The hydrolytic activity of purified enzymes was determined as release of p-nitrophenol from pNPGlc and other 4-nitrophenyl glycosides. 4-

Nitrophenol appears colorless below pH 5.4 and yellow above pH 7.5. The yellow color is due to a maximum of absorbance at 405 nm from the p-nitrophenolate ion in the solution. The activity of the enzyme was assayed against these synthetic substrates in 50 mM sodium acetate buffer, pH 5.5, at 60 ºC in 140 μL reactions. The enzyme and incubation time amount for WT and each mutant: WT 0.05 μg, 15 mins; H507A 1 μg, 20 mins; H507E 25 μg, 30 mins; H507Q 0.5 μg, 20 mins; D452A 25 μg, 30 mins; D452N 100 μg, 30 mins; R792A 50 μg, 30 mins; R792K 25 μg, 15 mins; T591A 0.5 μg, 30 mins; W732F 0.05 μg, 20 mins; E730A 0.1 μg, 20 mins; E730Q 0.05 μg, 15 mins; R786A 0.05 μg, 15 mins; R786K 0.05 μg, 15 mins; E777A 25 μg, 30 mins; E777Q 50 μg, 30 mins; T450A 0.5 μg, 15 mins; Y790F 1 μg, 30 mins; Y523F 0.5 μg, 15 mins; Q727A 0.05 μg, 15 mins; Q727E 0.05 μg, 15 mins; Y445L 0.5 μg, 15 mins; W525F 0.05 μg, 15 mins and W525L 0.05 μg, 15 mins. Reactions were stopped by alkalinization with 70 μL 2 M Na2CO3, and p-nitrophenol was quantified by the 405 nm absorbance of its phenolate ion at 405 nm, in comparison to a p-nitrophenolate standard curve in the same solution. The kinetic parameters, including kcat, KM, and kcat/KM, of purified TxGH116 mutants from E. coli were calculated by nonlinear regression of Michaelis-Menten plots with the Grafit 5.0 computer program (Erithacus Software, Horley, UK) and checked with linear (Hanes-Wolff [S]/v0 vs. [S]) plots (Scheme 3.1, Figure 3.2). For standard activity assays, 1 mM 4NPGlc was incubated with enzyme at 60 °C for 15 min. (Charoenwattanasatien et al., 2016).

3.3.2 PGO assay of oligosaccharide substrates

The activity of enzyme with cellobiose or cellotriose was assayed in 50 mM sodium acetate buffer, pH 5.5, at 60ºC in 50 μl reactions. The incubation time for each mutant was based on their relative activity which gave a detectable signal value.

Reactions were bioled for 5 mins to stop the reactions after incubation, followed by centrifugation at 12,000 rpm 4℃ for 30 mins to remove enzyme Forty microliters of supernats were pipetted in a microplate well that contain 10 µl reaction buffer, with 50 µl 1 mg/ml ABTS (2,2′-Azinobis-(3-ethylbenzthiazolin-6-sulfonic acid), Sigma) and 100 µl 0.01 g/ml PGO enzyme (Sigma) was added and the plate incubated at 37 ℃ for 30 mins, the reactions absorbance at 405 nm was measured and compared to that of a glucose standard curve.
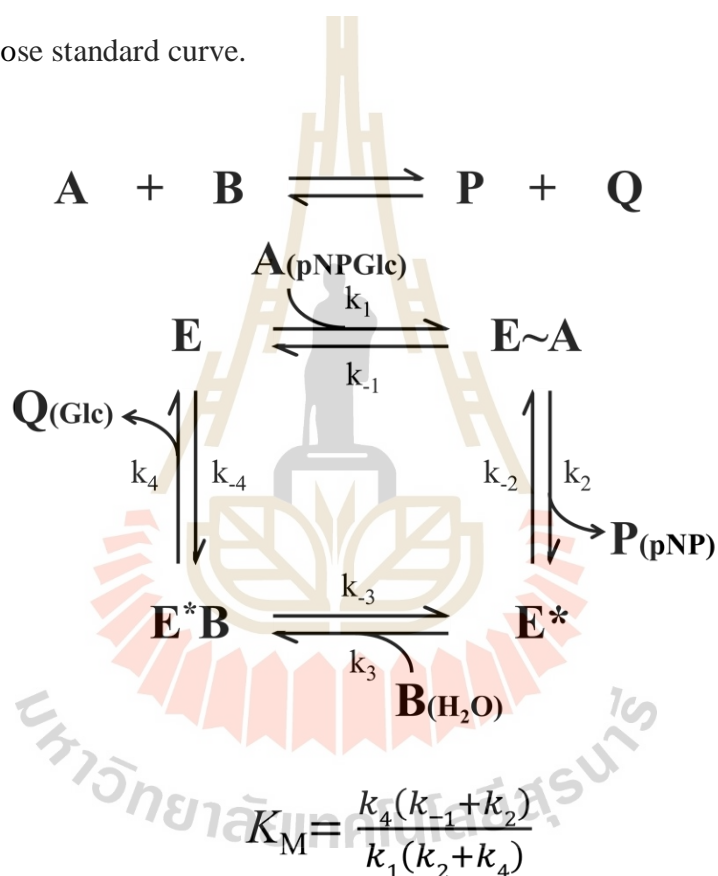
Scheme 3.1 Schematic view of Ping-Pong Bi Bi reaction of *Tx*GH116 with pNPGlc substrate.

$$\frac{[S]}{v} = \frac{1}{V_{max}}[S] + \frac{K_M}{V_{max}}$$
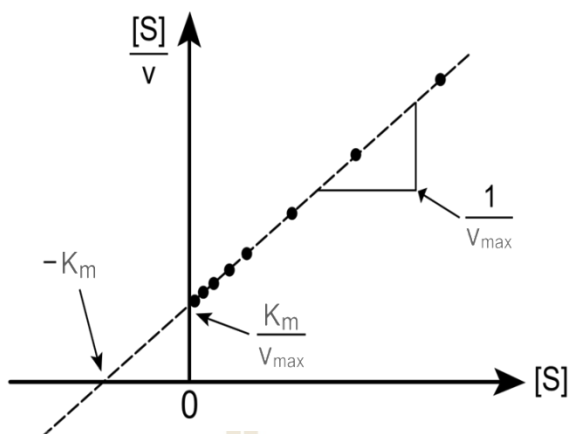
Figure 3.2 Hanes-Woolf plot representation of enzyme kinetics. Perfect data will yield a straight line of slope $1/V$max, a y-intercept of K_M /Vmax and an x-intercept of $-K_M$.

3.3.3 pH and temperature dependence

The pH dependence of the enzymatic activity was determined over the pH range of 2.5-8.0 at 0.5 pH unit intervals, at 60°C for 30 min in 100 mM McIlvaine universal (citrate-phosphate) buffer (McIlvaine, 1921). The optimal temperature was evaluated at the optimal pH in assays at temperatures over the range of 10-90°C at 5°C intervals. The optimum reaction condition of *Tx*GH116 mutants for hydrolysis of 4NPGlc was tested by incubating specific amounts of enzyme and specific incubation times which were setting from their kinetics assay, with 1 mM 4NPGlc in 100 mM McIlvaine universal (citrate-phosphate) buffer at pH 5.5.

3.3.4 Glucose and cellobiose inhibition kinetics of different mutants

Activity against 4-NPGlc was assayed in a range of glucose and cellobiose concentrations (Table 3.5) in triplicate reactions at five 4NPGlc substrate

concentrations bracketing the apparent KM value (Figure 3.3). Glucose was preincubated with 0.05 μg of enzyme at 37°C for 10 min before the addition of 4NPGlc and incubation at 60°C for 15 min. Because of the hydrolysis reaction with cellobiose, no preincubation was done for cellobiose inhibition kinetics. The Ki values were calculated from the x intercept of the derivative plot of the slopes (KMapp/Vmaxapp) of Lineweaver-Burk reciprocal plots (1/v vs 1/[4NPGlc]) vs the inhibitor concentration (Dixon, 1953).

Table 3.5 Glucose treatment for *Tx*GH116 WT and variants.

Protein	Glucose concentrations (mM)					
WT	0	2.5	5	7.5	10	12.5	15
T450A	0	10	20	30	40	50	60
W525F	0	2.5	5	7.5	10	12.5	15
W525L	0	2.5	5	7.5	10	12.5	15
E730A	0	5	10	15	20	25	30
E730Q	0	5	10	15	20	25	30
W732F	0	2.5	5	7.5	10	12.5	15
Q727A	0	5	10	15	20	25	30
Q727E	0	2.5	5	7.5	10	12.5	15
R786A	0	2.5	5	7.5	10	12.5	15
R786K	0	15	20	25	30	40	60
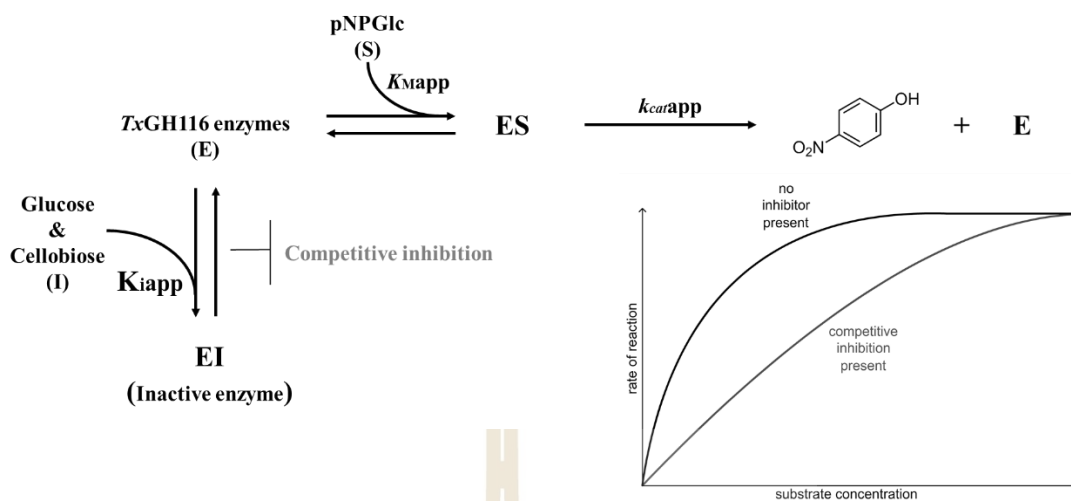
Figure 3.3 Mechanism of glucose and cellobiose molecular competitive inhibition to *Tx*GH116 enzyme.

3.3.5 Thin-layer chromatography (TLC)

The products of the reaction mixtures which were expected to have transglycosylation products were monitored by TLC (silica gel 60 F254, Merck, Germany) using ethyl acetate-acetic acid-water (2:1:1) as solvent. 4-nitrophenyl glycosides and *p*-nitrophenol were observed under UV. Carbohydrates were visualized by painting with 10% sulfuric acid in ethanol followed by charring.

3.3.6 Protein crystallization

Activity against 4-NPGlc was assayed in a range of glucose and cellobiose concentrations (Table 3.5) in triplicate reactions at five 4NPGlc substrate concentrations bracketing the apparent KM value (Figure 3.3). Glucose was preincubated with 0.05 µg of enzyme at 37°C for 10 min before the addition of 4NPGlc and incubation at 60°C for 15 min. Because of the hydrolysis reaction with cellobiose, no preincubation was done for cellobiose inhibition kinetics. The Ki values were calculated from the x intercept of the derivative plot of the slopes

(KMapp/Vmaxapp) of Lineweaver-Burk reciprocal plots (1/v vs 1/[4NPGlc]) vs the inhibitor concentration (Dixon, 1953).

Table 3.6 Crystallization precipitant factor composition created from Hampton research online survice.

0.1 M MES	0.1 M MES	0.1 M MES	0.1 M MES	0.1 M MES	0.1 M MES
19 % PEG3000	20 % PEG3000	21 % PEG3000	22 % PEG3000	23 % PEG3000	24 % PEG3000
0.15 M Ammo	0.15 M Ammo	0.15 M Ammo	0.15 M Ammo	0.15 M Ammo	0.15 M Ammo
0.1 M MES	0.1 M MES	0.1 M MES	0.1 M MES	0.1 M MES	0.1 M MES
19 % PEG3000	20 % PEG3000	21 % PEG3000	22 % PEG3000	23 % PEG3000	24 % PEG3000
0.2 M Ammo	0.2 M Ammo	0.2 M Ammo	0.2 M Ammo	0.2 M Ammo	0.2 M Ammo
0.1 M MES	0.1 M MES	0.1 M MES	0.1 M MES	0.1 M MES	0.1 M MES
19 % PEG3000	20 % PEG3000	21 % PEG3000	22 % PEG3000	23 % PEG3000	24 % PEG3000
0.25 M Ammo	0.25 M Ammo	0.25 M Ammo	0.25 M Ammo	0.25 M Ammo	0.25 M Ammo
0.1 M MES	0.1 M MES	0.1 M MES	0.1 M MES	0.1 M MES	0.1 M MES
19 % PEG3000	20 % PEG3000	21 % PEG3000	22 % PEG3000	23 % PEG3000	24 % PEG3000
0.3 M Ammo	0.3 M Ammo	0.3 M Ammo	0.3 M Ammo	0.3 M Ammo	0.3 M Ammo

3.3.7 Synchrotron X-ray diffraction and structure solution

Crystals of *Tx*GH116 mutants were soaked briefly in cryo solution in which the concentration of each component from precipitant solution was increased 10% containing 20% glycerol. For glucose and cellobiose complexes, the soaking solution contained 100 mM D-glucose and 100 mM cellobiose. For G2F complexes, the soaking solution contained 10 mM G2F. The crystals were flash vitrified in liquid nitrogen and shipped to the National Synchrotron Radiation Research Center (NSRRC) for data collection. Diffraction data was collected on beamline TPS05A1

and beamline BL15A1 on a Rayonix MX300HE CCD detector, beamline BL13B1 on an ADSC Quantum-210r CCD detector with 1.0 Å wavelength X-rays, maintaining the crystals at 100 K in a nitrogen cold stream in each case. Data was processed and scaled with the HKL-2000 package (Otwinowski and Minor, 1997). The initial phases for the mutants and their complexes were determined by molecular replacement with the MOLREP program (Vagin and Teplyakov, 2010) in the CCP4 suite (Winn et al., 2011) using the wild type structure as template. The structure was built with the COOT graphic program (Emsley and Cowtan, 2004) in alternation with refinement with REFMAC5 (Murshudov et al., 2011). The final models were analyzed with PROCHECK (Laskowski et al., 1993) and MolProbity (Chen et al., 2010) and validated on the PDB Website. The mutants were visualized and compared with wild type to identity their structural based mechanism for glucose binding and substrate hydrolysis by Pymol (Schrödinger LLC, Portland, OR, USA) and Discovery Studio 3.1 (BioVia, San Diego, CA, USA) graphics programs.

CHAPTER IV

RESULTS AND DISCUSSION

4.1 Systematic Functional Analysis of Glycone Sugar Binding Residues in GH116 family from *Thermoanaerobacterium xylanolyticum*

4.1.1 Introduction on active site glycone sugar binding related residues

An enzyme's catalytic reaction normally occurs in a buried pocket-like active site where multiple functional groups concert to comprise unique environments that enable the catalysis of reactions with unparalleled efficiency and specificity. In the active site and even outside surrounding this site, a relatively small number of specific amino acid residues are involved in substrate binding and an even smaller subset of these are directly participant in the catalytic reaction. These catalytic residues provide various reactive groups that promote catalysis or improve substrate binding (Holliday et al., 2005). Functions performed by amino acid residues during catalysis have been combined into seven basic classes: stabilization roles, steric roles, activation roles, proton shuttling residues, hydrogen shuttling residues, electron shuttling residues and residues acting through covalent catalysis (Holliday et al., 2009).

Beta-glucosidases (β-D-glucopyranoside glucohydrolases, E.C. 3.2.1.21) are functional enzymes which remove the nonreducing terminal β-D-glucosyl residue from glucoconjugates, including glucosides, 1-O-glucosyl esters, and

oligosaccharides (Ketudat Cairns et al., 2015). β-Glucosidases have been categorized into the protein sequence-based glycoside hydrolase families GH1, GH2, GH3, GH5, GH9, GH30 and GH116 (http://www.cazy.org) (Lombard et al., 2014). Different family show their own specific traits in terms of hydrolase or transglycosylase activities. The specificity of β-glucosidases toward different substrates varies mostly depending on the enzyme source and function. Previous theoretical and experimental developments have revealed that hydrolysis of the glycosidic bond can occur with one of two possible stereochemical outcomes: inversion or retention of anomeric configuration (McCarter and Withers, 1994; Zechel and Withers, 2000). Many β-glucosidase studies elucidated the roles of the catalytic residues that perform the chemistry, including the general acid/base and nucleophile (McCarthy et al., 2004). Covalent (Burmeister et al., 1997; Kurakata et al., 2008), and electrostatic (Geronimo et al., 2018) contributions to catalysis and even convertion between glycosidase with transglycosidase activities (Frutuoso and Marana, 2013; Lundemo et al., 2017; Teze et al., 2015) have been studied. Generally noncatalytic amino acids are routinely assumed to be only important for facilitating substrate binding (Chuenchor et al., 2011). However some previous studies have shown that the contributions of enzyme active-site residues not directly involved in bond making and breaking could be as essential to the enzyme's function as catalytic residues (Dopitova et al., 2008; Lucas and Siegel, 2015; Pengthaisong et al., 2012; Zhang et al., 2015). This enables enzymes to conduct catalysis through a variety of strategies, such as introducing strain to destabilize the substrate (Desjardins et al., 2017) and preorganizing active-site dipoles for transition-state stabilization (De Giuseppe et al., 2014; Shanmugam et al., 2018).

The enzyme used in this work *Thermoanaerobacterium xylanolyticum* GH116 (*Tx*GH116) is a thermostable β-glucosidase which consists of an N-terminal domain formed by a two-sheet β-sandwich, and a C-terminal (α/α)$_6$ solenoid domain caontaing one slot like active site (Charoenwattanasatien et al., 2016). High-resolution crystal structures of *Tx*GH116–glucose and inhibitor complexes provide invaluable insight into the different elements involved in substrate binding and catalysis (PDB ID: 5FJS, 5BVU, 5BX2, 5BX3, 5BX4 and 5BX5) (Artola et al., 2017; Charoenwattanasatien et al., 2016). Our previous work verified that the enzyme has an retaining mechanism by identifying the initial hydrolysis product as β-D-glucose, which retains the anomeric stereochemistry of the substrate, by monitoring the reaction time course by NMR spectroscopy. E441 is the catalytic nucleophile and D593 is the catalytic acid/base, as was confirmed by mutation of these residues to alanine, followed by chemical rescue using small nucleophiles. The *Tx*GH116 model, in which all pathogenic GBA2 mutation sites are strictly conserved, and helped to elucidate the molecular bases of human GBA2 mutations leading to ataxias and paraplegias (Sultana et al., 2020; Woeste and Wachten, 2017).

Although residues previously proposed to be directly involved in the reaction chemistry have been analyzed (Charoenwattanasatien et al., 2016), the importance and role of other residues within the active-site pocket of *Tx*GH116 and even GH116 family enzymes remain unclear. In this work, we systematically mutated amino acids that form the reminder of the active-site residues interacting with glucose ligand to quantify the contribution of each side chain toward the catalytic efficiency of *Tx*GH116. These mechanistic details improve our understanding of how this enzyme achieves function by the relative importance of active site residues in substrate and inhibitor binding, which may allow the modification of *Tx*GH116 β-glucosidase for

improved application. It may also allow development of treatment of human disorders related to GBA2, which plays important role in balancing sphingolipid levels in cells (Jatooratthawichot et al., 2020).

4.1.2 Results
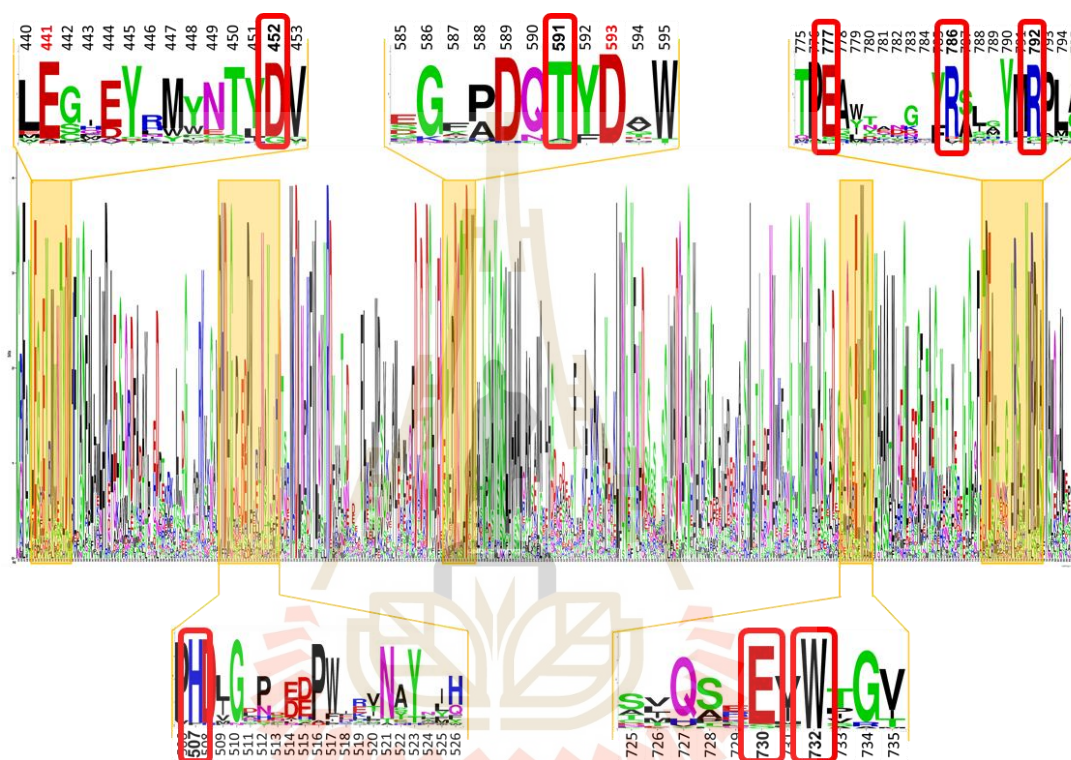
Figure 4.1.1 Sequence logo of 34 diverse sequences of GH116 family enzymes active site glycone sugar binding related residues from different species. The conserved regions around the active site are highlighted, and glycone sugar binding related residues' positions in the sequence are shown in bold with red boxes and catalytic residues are shown in red bold. Amino acid conservation analysis was performed to assess the conserved residues in GH116 enzyme family catalytic C-terminal domainin. Each logo consists of stacks of symbols, one stack for each position in the sequence. The overall height of the stack indicates the

sequence conservation at that position, while the height of Y-axis symbols within the stack indicates the relative frequency of each amino or nucleic acid at that position.
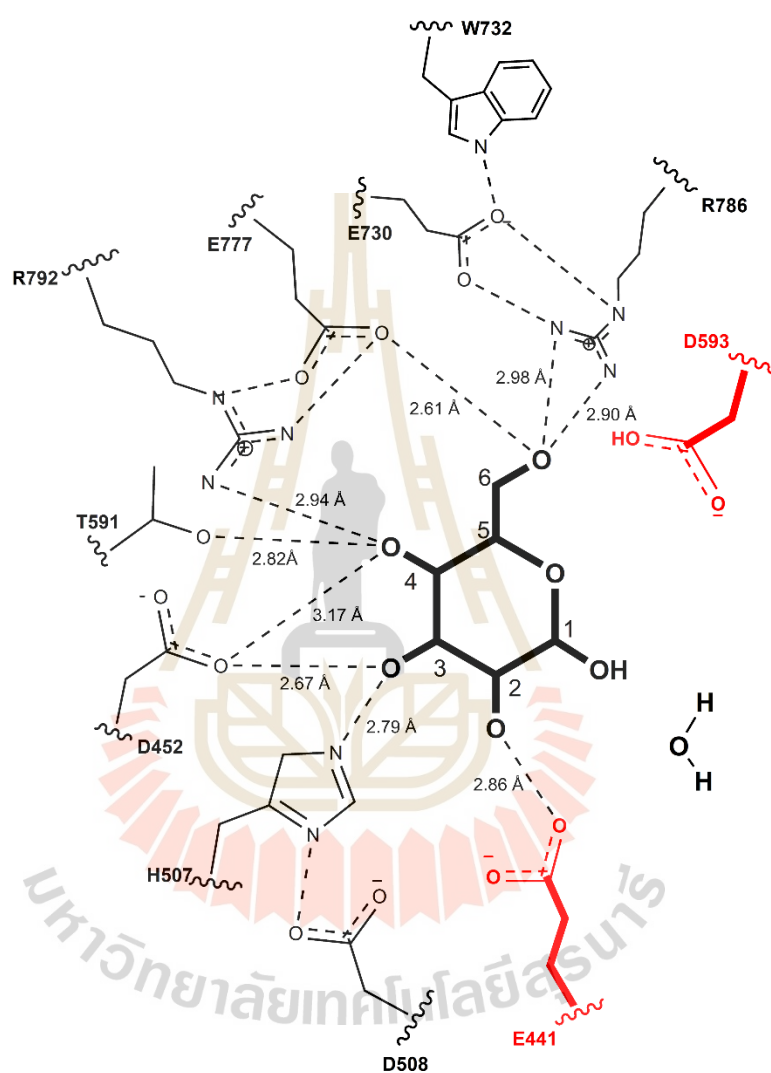
Figure 4.1.2 Binding interactions of *Tx*GH116 WT with glucose. The ligands and protein side chains from wild type *Tx*GH116 bound to glucose (PDB: 5BX5) (Charoenwattanasatien et al., 2016) are shown in line representation. Hydrogen bonds are shown as black dotted lines. Water molecules located in the active site are shown as 2D line.

Glycone-binding residues were identified and shown highly conservation in GH116 family from different species (Figure 4.1.1). The complex of TxGH116 WT with glucose (PDB: 5BX5) (Charoenwattanasatien et al., 2016), the polar interactions between these residues and glucose include: D452 with the C3OH and C4OH, H507 with C3OH, T591 with C4OH, E777 and R786 with C6OH, and E730 with R786 (Figure 4.1.2). The catalytic nucleophile, E441 also makes a putative hydrogen bond with the glucose C2OH, but this is difficult to separate from its catalytic function and mutation of this residue was previously studied (Charoenwattanasatien et al., 2016; Pengthaisong et al., 2021). We systematically mutated for the glucose-biding residues mentioned above and characterized the kinetics of the mutants to elucidate their relative importance and specific roles.

4.1.2.1 Effects of H507 residue mutations on the activity

Previously, mutations of D508, which interacts with H507, were shown to cause drastic decreases in *Tx*GH116 activity (Charoenwattanasatien et al., 2016). The H507A, H507E and H507Q mutants have similar pH optima at pH 5.0, slightly lower than wild type at pH 5.5 (

Figure **4.1.3** A) and all showed lower temperature optima, suggesting lower stability of the enzyme or enzyme-substrate complex. The H507 mutations increased the K_M value 58 to 196-fold compared with WT. H507A eliminates the imidazole ring, thereby eliminating the interaction with the 3OH group, while in principle H507Q and H507E could still interact, but H507Q may also lose the interaction with the 3OH and instead interact with D508, since it has similar kinetic parameters to H507A. While H507A and H507Q decreased the k_{cat} value less than 10-fold, the H507E mutation decreased the k_{cat} over 4000-fold and k_{cat}/K_M values over 200,000-fold (Table 4.1.1), indicating that the negative charge at this residue is not tolerated.
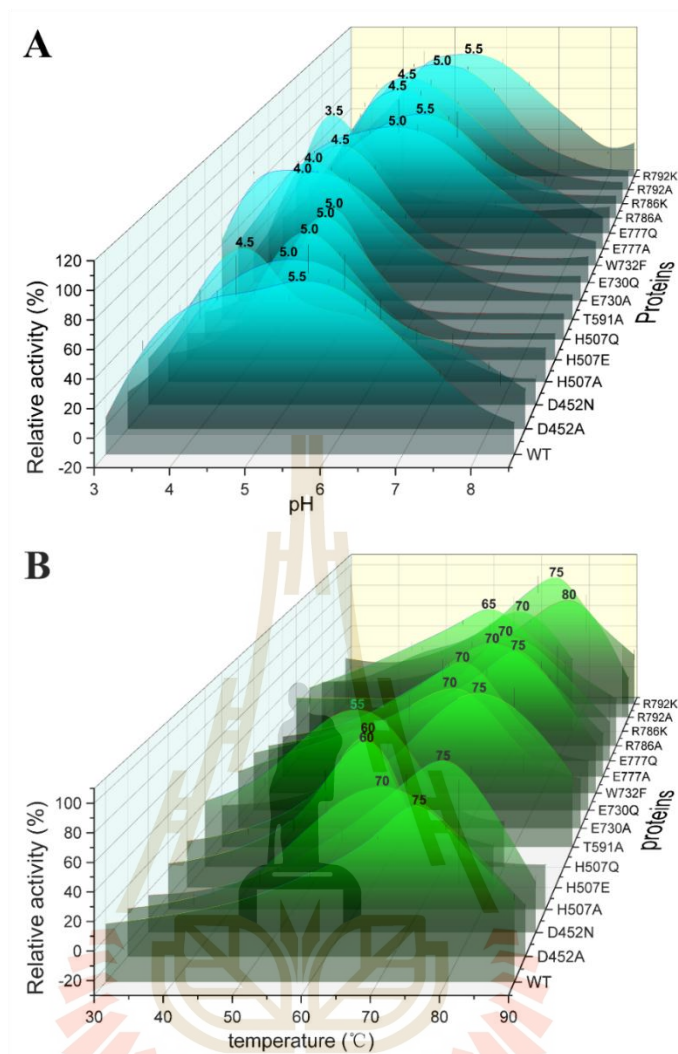
Figure 4.1.3 Optimal pH (A) and temperature (B) analysis of mutants and WT *Tx*GH116. All assays were carried out in triplicate. 3-D stacked graph was made by Origin 8.0.

4.1.2.2 Effects of D452 residue mutation on the activity

The interactions of D452 with the glucose C3OH and C4OH are important for catalysis, since both D452 mutations, D452A and D452N, greatly decreased activity. The D452A and D452N mutations led to 160 and 70-fold increases in the K_M values, respectively, and they decreased the k_{cat} values more than 340-fold

and 5000-fold, which led to a reductions in k_{cat}/K_M of more than 56,000 and 352,000-fold (Table 4.1.1). These results are consistent with the much weaker binding of the sugar of the substrate and transition state expected based on this residue's interaction with the glucose C3OH and C4OH (Figure 4.1.2).

4.1.2.3 Effects of R792 residue mutations on the activity

In the complex structure, R792 appears to have H-bond interactions with C4OH and its neighboring E777 residue (Figure 4.1.2). R792 residue mutations R792A and R792K did not affect the pH dependance much (Figure 4.1.3 A), as they have broad working pH profiles and pH optima comparable with WT. The R792A and R792K mutants show similar KM values that were >25 times higher than wild type and very low kcat values 5040- and 780-fold lower than wild type, respectively (Table 4.1.1).

4.1.2.4 Effects of T591 residue on the activity

T591 is on the flexible loop containing the catalytic acid/base D593 and forms hydrogen bonds with C4OH via its backbone carbonyl and sidechain hydroxyl group, which also interacts with D452 (Figure 4.1.2). The T591A mutation eliminates the interactions, resulting in a 70-fold increase in the KM and 10-fold reduction in the kcat value compared with WT (Table 4.1.1). At its optimum pH of 4.5, T591A (Figure 4.1.3 A) has significantly higher activity than at pH 5.5, the pH optimum of the wild type enzyme.

4.1.2.5 Effects of W732 residue mutation on the activity

W732 appears to form a CH-pi interaction with Glc C6 (PDB: 5BX5). Although it does not form an H-bond to the sugar, its indole NH group can form a H-bond with E730. The W732F variant has a nearly 3-folder higher K_M, but also a 1.5-fold increase in k_{cat} value compared to wild type TxGH116 (Table 4.1.1).

Table 4.1.1 Kinetic constants of the *Tx*GH116 WT enzyme and mutants with *p*NPGlc substrate. The heat map represents the fold changes in k_{cat}, K_M, and k_{cat}/K_M each mutant relative to the WT enzyme. The assay were inducted in 50 mM NaOAc buffer, pH5.5. Red represents different fold increases and blue represents different fold decreases (from light to dark). Values in white boxes are within a factor of 3 of those for WT.

Protein	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1}s^{-1})
WT	0.170 ± 0.014	41.3 ± 0.63	243
D452A	28.3 ± 1.3	0.120 ± 0.0021	0.0043
D452N	11.7 ± 0.70	0.00810 ± 0.00010	0.00069
H507A	23.0 ± 0.71	4.72 ± 0.051	0.21
H507E	9.93 ± 0.62	0.0100 ± 0.00020	0.0010
H507Q	33.3 ± 4.3	12.2 ± 0.63	0.37
T591A	12.6 ± 0.45	3.63 ± 0.037	0.29
E730A	3.79 ± 0.25	80.8 ± 1.4	21.3
E730Q	7.21 ± 0.49	55.6 ± 1.1	7.72
W732F	0.470 ± 0.010	62.3 ± 0.39	133
E777A	29.9 ± 0.63	0.120 ± 0.0010	0.0039
E777Q	6.97 ± 0.22	0.0200 ± 0.00020	0.0028
R786A	7.89 ± 0.57	171 ± 4.0	21.7
R786K	7.30 ± 0.71	88.9 ± 2.7	12.2
R792A	4.53 ± 0.38	0.00820 ± 0.00010	0.0018
R792K	5.32 ± 0.44	0.0530 ± 0.0013	0.0099

← Loss of function						Neutral	Gain of kinetics value →			
>1000 fold	300-1000 fold	100-300 fold	20-100 fold	10-20 fold	3-10 fold	Neutral	3-20 fold	20-50 fold	50-100 fold	100-300 fold

4.1.2.6 Effects of E777 residue mutations on the activity

E777 forms a hydrogen bond to the CO6H and also interacts with R792 (Figure 4.1.2). Both E777A and E777Q show very low activity compared with WT (Table 4.1.1). E777A exhibits a 170-fold increased KM, while E777Q shows only a 40-fold increase compared with WT. Both mutants show very low efficiency

with a reduction around 80,000-fold in kcat/KM value. E777Q shifts the optimum pH to pH 3.5 (Figure 4.1.3 A). This suggests that the charge on E777 influences the pH optimum and is important to catalysis, along with its hydrogen bonding to the C6OH group, which can be maintained in E777Q.

4.1.2.7 Effects of R786 residue mutations on activity and substrate recognition

R786 forms two H-bonds to the glucose C6OH in the -1 subsite and one to the phenolic OH of Y445, in addition to E730 noted above (Figure 4.1.2). The R786A mutation eliminates the H-bonds, while R786K diminishes the number of H-bonds interactions possible. Both of two mutants displayed lower temperature optima (Figure 4.1.3 B), suggesting lower thermostability, especially R786A. These mutants have broad pH profiles like wild type, but their pH optimum is shifted to pH 4.5 (Figure 4.1.3 A). Their KM values are 43-46-fold higher than wild type, but R786A has a 4-fold and R786K a 2-fold higher kcat value compared with wild type (Table 4.1.1). At its pH optimum of 4.5, R786K also has a lower kcat (144 s-1) compared with R786A (171 s-1) (Table 4.1.1). In order to understand the effects of the R786 mutations, we determined the structures of the R786 mutants, the active sites of which are shown in In order to understand the effects of the R786 mutations, we determined the structures of the R786 mutants, the active sites of which are shown in Figure 4.1.5 (A, B, C, D). Instead of making the expected interaction with the Glc 6OH, E730 or Y445, the lysine side chain in R786K was found to turn away from the sugar toward the outside of the active site (Figure 4.1.5 D and E). The flexible loop in the entrance to the active site also shows a different rotation in the R786K mutant (Figure 4.1.5 G and H) which appears to make room for the K786 side chain.

R786K had lower activity with cellobiose compared with R786A and WT TxGH116 (Table 4.1.2). R786K and R786A had similar KM values for hydrolysis of cellobiose, which were about 5-fold higher than WT, but the kcat of R786K is significantly lower than those of the R786A mutant and wild type TxGH116. As such, the specificity constant (kcat/KM) of R786K was nearly 2-fold lower than R786A and over 10-fold lower than WT enzyme. Thus, the expected milder mutation R786K, had a greater defect than R786A, likely due to the unexpected displacement of the K786 side chain (Figure 4.1.5 C and D).

4.1.2.8 Effects of E730 residue mutations on the activity

E730 interacts with both the W732 and R786 residues, which in turn make direct interactions with the substrate molecule (Figure 4.1.2). Our mutants E730A and E730Q show less thermostability and pH optima at pH 4.0 (Figure 4.1.3 A). The E730A mutant should eliminate interactions with W732 and R786, while E730Q mutant should maintain hydrogen bonding, but lose the coulombic component of those interactions. Remarkably, E730Q also appears slightly less stable than E730A (Figure 4.1.3 B). Interestingly, the KM value for pNPGlc increased 20-fold and its kcat value increased 2-fold in E730A compared to wild type, while in E730Q the KM value increased 40-fold and the kcat value increased 15 fold (Table 4.1.1).

4.1.2.9 Correlation analysis of glucose or cellobiose with enzyme specificity

Both glucose and cellobiose display competitive inhibition of TxGH116 hydrolysis of pNPGlc (Figure 4.1.4), with WT and R786K appearing more sensitive than TxGH116 R786A. Based on this inhibition mechanism, the KM and KM/Vmax of pNPGlc hydrolysis should follow linear curves with increasing inhibitor

concentration with an intercept of the x-axis representing the Ki of glucose or KM of the competing substrate cellobiose (Figure 4.1.4). Protein X-ray crystallography results indicates an enlarged mouth to the active site in R786A and R786K (Figure 4.1.5), which may affect substrate and product movement in and out of the active site in which TxGH116 performs a ping-pong bi bi reaction mechanism. Cellobiose gives Ki values of 0.53 mM for wild type, 6.2 mM for R786A and 8.0 mM for R786K, which are a bit higher than their respective KM values for cellobiose in Table 4.1.2.

Table 4.1.2   Kinetic constants of the *Tx*GH116 WT enzyme and mutants with cellobiose substrate. The heat map represents the fold changes in k_{cat}, K_M, and k_{cat} /K_M each mutant relative to the WT enzyme. Red represents different fold increases, and blue represents different fold decreases (from light to dark with increasing fold change).
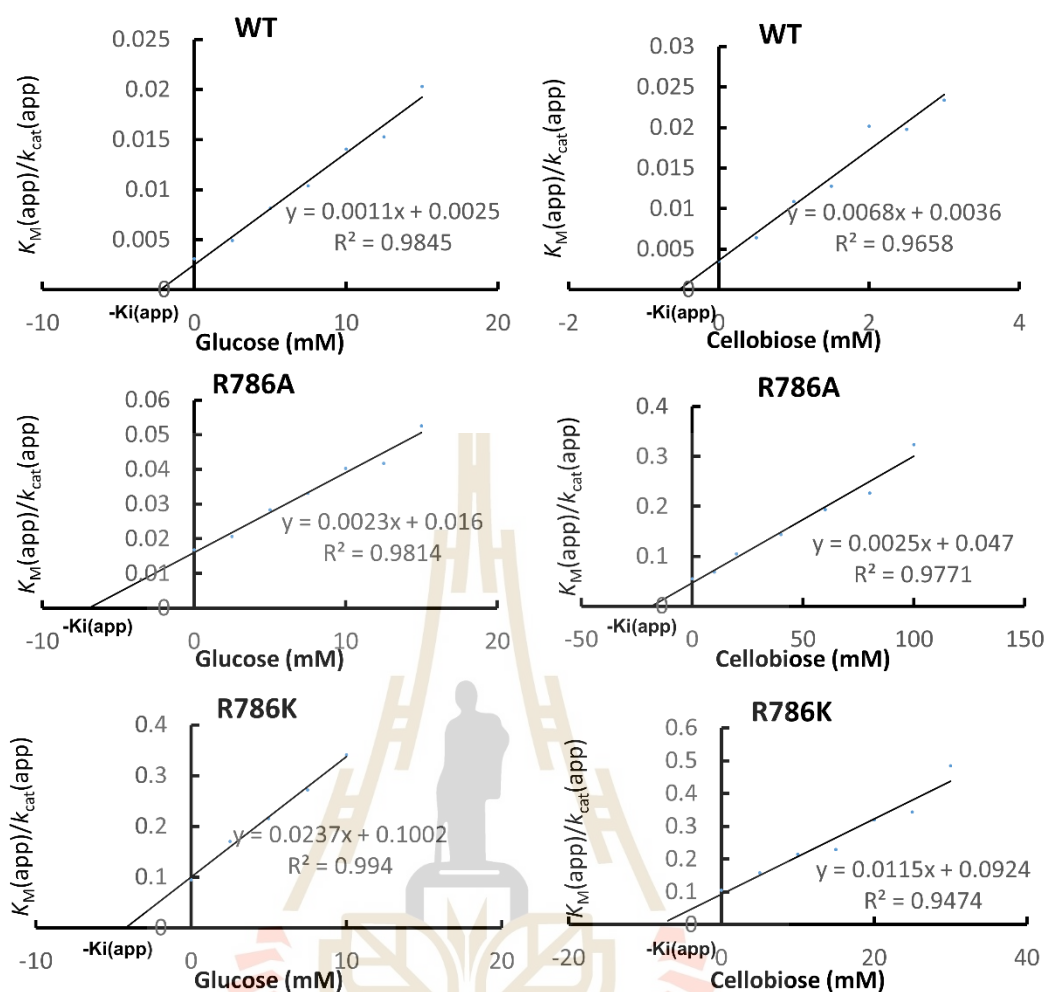
Figure 4.1.4 Correlation of $K_M(\text{app})/k_{cat}(\text{app})$ with glucose or cellobiose concentrations for WT and R786 mutants. WT and R786A, R786K mutants show linear curves in response to different glucose and cellobiose because of their competitive inhibition mechanism. The slope value from linear curve represent for enzyme sensitivity to glucose or cellobiose.

4.1.2.10 Effect of R786 and E730 mutations on sugar specificity

Since R786 interacts with the glucose C6OH group, we tested whether loss of this interaction would affect sugar specificity, especially for β-

D-glucopyranoside versus β-D-xylopyranoside, which lacks this hydroxymethyl group. We compared the activity of C6OH related mutants to wild type on pNPGlc, pNP-β-D-xylopyranoside, pNP-β-D-galactopyranoside and pNP-β-cellobioside. All mutants have reduced activity on pNPGlc compared with WT, but with pNP-xylopyranoside, R786A and R786K show high hydrolysis activity compared with WT (Figure 4.1.6). All four mutants had very low activity on pNP-β-D-galactoside, which differs from glucose in the chirality of the C4OH group, consistent with R786 not interacting with this group.
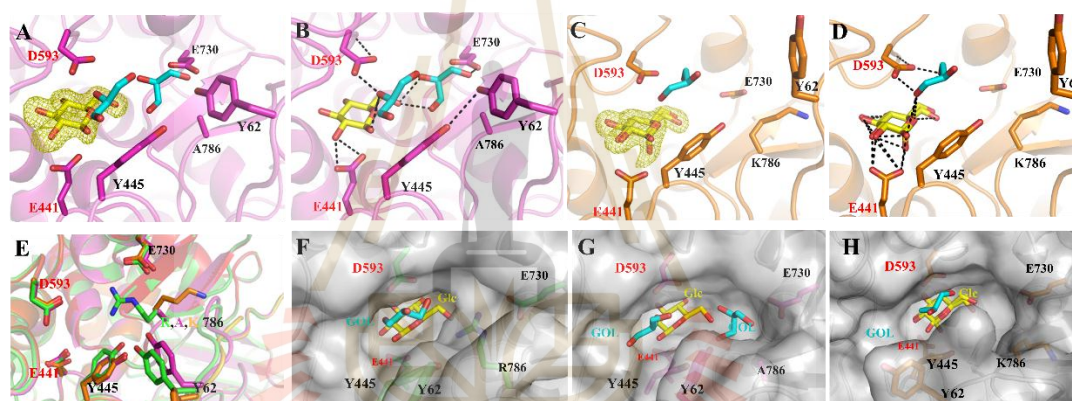
Figure 4.1.5 Active site comparison between *Tx*GH116 R786A and R786K and WT, with glucose ligand. (A), R786A (C), R786K complex with glucose ligand electron density omit map, contoured at 3σ, in the active sites of the complexes. Active site interactions with glucose and glycerol molecules. (B), In R786A mutant the C6OH group and E730 residue have lost their interactions with the original arginine. (D), R786K mutant also lost the interactions with C6OH group and E730 residue with two conformations glucose (alpha 80% and beta 20%) ligand. (E), Superposition of active site residues of WT (PDB: 5BX5) (Charoenwattanasatien et al., 2016) green), R786A (magenta) and R786K (orange). Catalytic residues and E730 are

conserved, Y445 and Y62 residues shown in different rotation. (F, G, H), Slot like active site entrance comparison between WT (F), R786A (G) and R786K (H). R786A and R786K have an enlarged entrance space, with an extra glycerol molecule in the place of the original arginine side chain. The potential hydrogen bonds between the glucose ligands (yellow sticks) and surrounding residues are displayed by black dashed line. Glycerol molecules are shown as cyan sticks, and catalytic residues are labeled in red.
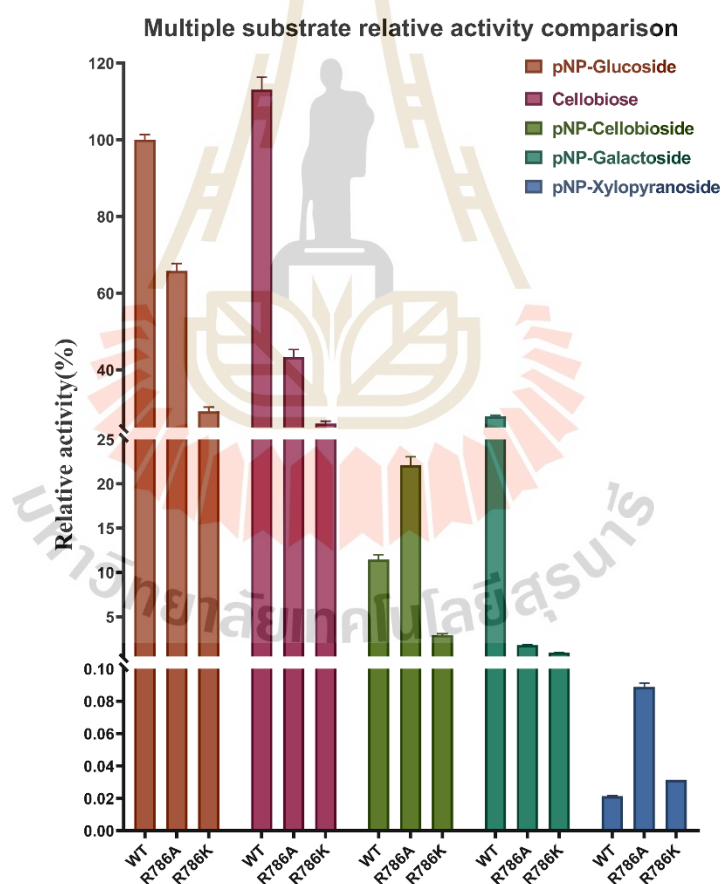
Figure 4.1.6 Relative activity on different substrates. pNP-β-D-glucoside, cellobiose, pNP-β-cellobioside, pNP-β-D-galactoside, pNP-β-D-xylopyranoside. The assay contained 1 mM substrate in 50 mM sodium acetate buffer, pH

5.5, at 60°C. Relative activity is based on 4NP released from 4NP β-glycosides. The values for cellobiose hydrolysis were determined by dividing the amount of glucose released by 2, since two glucose molecules are released per glycosidic bond hydrolyzed. Data are expressed as means of three independent reactions ± SD.
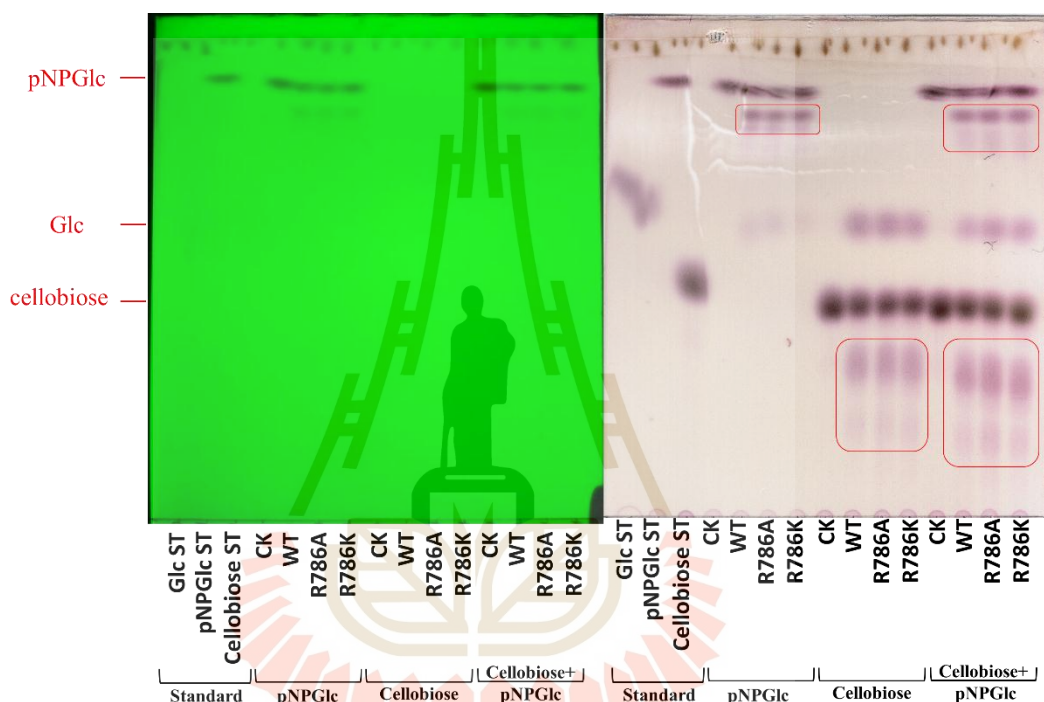
Figure 4.1.7 Tranglycosylation analysis between WT and R786 mutation variants on hydrolysis of pNPGlc and cellobiose substrate. The assay reaction contained 10 mM substrate in 50 mM NaOAc buffer, pH 5.5, at 60°C incubated for 18 hours. The TLC was developed in EtOAc:Acetic acid:H_2O ratio of 11:5:4 mobile phase for two rounds.
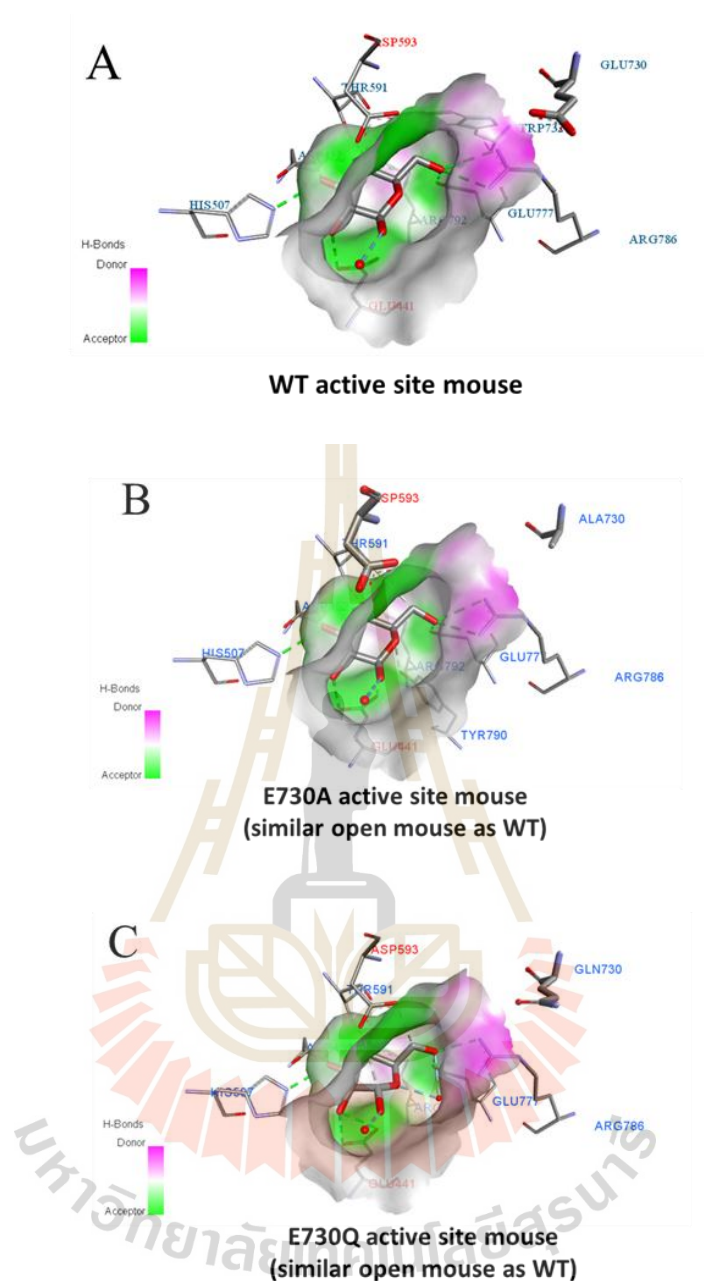
Figure 4.1.8 Hydrogen bond formate ability from E730 mutations and comparison with WT. Catalytic residues are shown in red, glycone sugar binding related residues in blue. Surface colors represent the hydrogen bond donor/acceptor. Hydrogen bond donors are colored in green and hydrogen bond acceptors in purple.

4.1.3 General discussion

In this work, we provide the systemic functional analysis of TxGH116 active site glycone sugar binding related residues in substrate binding and catalysis. From enzyme kinetics and structural analysis, we clarified the roles of each amino acid residue. In general, simple binding interactions with the glycone of the substrate cannot fully explain the activity changes upon mutation of the glucose binding molecule and their charge-distribution and stabilization of the transition state appear to be more critical, based on larger effects on kcat than KM.

Residues in the glycone sugar binding site show their importance in hydrolysis activity performance from their site directed mutagenesis. R792 forms H-bonds to the glucose C4OH and hydrogen bond interaction with E777 (Figure 4.1.2). The mutants R792A and R792K show very low hydrolysis activity indicating the essential role on R792 in substrate and transition state binding.

The reduced activity of mutants of T591 confirmed its interaction with the substrate C4OH group (Figure 4.1.2) is important in substrate binding. The 10-fold drop in kcat in T591A suggests this interaction may also help stabilize the flexible loop containing the catalytic acid/base, D593, and/or the transition state conformation. This mainly because the unrestricted distribution of side chain conformations between the gauche,gauche (gg), gauche,trans (gt), and trans,gauche (tg) (Quirke and Crich, 2020) conformations depends mainly on the configuration at C4 in the pyranose ring of the glucose sugar (Zhao et al., 2007). Beta-glucosidases have one common phenomenon that the binding interactions at the glucose C3 and C4 positions individually contribute 3-10 kJ mol-1 to each transition state, whereas contributions in the ground state are much weaker (≤3kJ mol-1) (Namchuk and Withers, 1995). Changes in the O2-C2-C3-O3 torsional interaction play a significant

role in the determination of anomeric stereochemistry in 4,6-O-benzylidene directed β-mannosylation (Crich and Li, 2007). An inhibitor lacking the 3-hyxy group shows lower inhibition of sweet almond β-glucosidase indicated the importance of the substrate C3OH-enzyme hydrogen bond in cleavage (Ouairy et al., 2013). D452 appears to form H-bonds to the glucose C3OH and C4OH and H507 forms an H-bond to the C3OH (Figure 4.1.2). D452 and H507 residue mutants show a larger reduction in the kcat/KM value than the KM value, and had very low activity (Table 4.1.1). This indicates that maintenance of the C3 hydroxyl group interaction with these two residues also plays an essential role in TxGH116 substrate binding and glycoside bond hydrolysis. Mutants of the C4OH interaction residues D452 and R792 have lower kcat/KM values, mostly from the largely reduced kcat value, compared to T591A (Table 4.1.1). This is consistent with previous theory that the hydrogen bonding at the 4 position appears to accelerate glycosylation but decelerate the rate-limiting deglycosylation (Namchuk and Withers, 1995). T591A has less effect on hydrolysis compared with mutations of D452, which has both C3 and C4 hydroxyl group interactions (Figure 4.1.2). In the meantime, the large scale activity decrease in R792A mutant indicates the importance of the guanidino group participating in hydrogen bond binding interactions with the substrate glucosyl hydroxyl side chain (Romaniouk and vijay, 1997). The R792K mutant has a similar KM value with the alanine mutation, but a 6-fold high kcat value (Table 4.1.1) indicating that the residue 792 side chain charge-charge interaction with its neighbor residue E777 is likely also important (Figure 4.1.2).

The E777 mutants have low kcat values, likely due to the eliminated interaction with R792, which binds to the C4OH. The strong negative charge of the E777 residue indicates that the electrical stabilization between the E777 and R792

residues should be considered in their stabilization cooperation (Borders Jr et al., 1994; Shi et al., 2003). In the H507 mutants, H507E shows a lower kcat value compared with H507A and H507Q (Table 4.1.1). This may also because of glutamic acid residue has negative charge that disturbs the electron distribution in C4OH and its neighbor C3OH or even residue D508 (Charoenwattanasatien et al., 2016) that leads to low activity.

The R786A mutation eliminates H-bonds with C6OH and gave higher KM and kcat values for pNPGlc hydrolysis than WT. The R786K mutant displays a similar KM value to R786A with pNPGlc substrate (Table 4.1.1 and Table 4.1.2). This is consistent with our previous work in which TxGH116 with the R786H mutation (corresponding to a GBA2 R873H mutation causing spastie paraplegia) (Sultana et al., 2015) showed a 90-fold increase in the KM value and a reduction in kcat/KM of more than 20-fold (Charoenwattanasatien et al., 2016). This result confirmed that the mutation effected binding of the sugar expected based on this residue's interaction with the glucose C6OH. Glucose competitively inhibited R786A and R786K with Ki values similar to the WT. Both glucose and cellobiose molecule work as competitive inhibitors in hydrolysis of pNPGlc substrate. WT and R786A responded differently to the competition of glucose and cellobiose, which can be indicated from their linear curve slope values (equal to 1/Ki) with glucose and cellobiose treatment. WT has a higher slope value with cellobiose treatment than glucose indicating stronger inhibition by glucose, but R786A showed similar slopes with both molecules (Figure 4.1.6). This may reflect their differences in product releasing mechanism. As pNPGlc has both glycone and aglycone moieties, greater mutant enzyme activity might indicate that an amino acid substitution had occurred in a residue responsible for aglycone recognition or binding. Such substitutions might have a wholly different

effect on cellobiose that has different steric and electrostatic characteristics of the aglycone moiety (McCarthy et al., 2004). The x-ray crystal structure of R786K has a similar active-site structure as WT, but the nucleophile residue E441is closer to the substrate glycoside bond (Figure 4.1.5 C and D), which the rate-limiting deglycosylation step (Charoenwattanasatien et al., 2016) (for which the reaction is the same) is affected. Since the interactions with beta-glycosidases at the C2OH position contribute substantially (18-22 kJ mol-1) to stabilization of the glycosylation and deglycosylation transition states (Namchuk and Withers, 1995), the nucleophile residue in the R786K mutant may have stronger stabilizing effects on the oxocarbenium ion transition state due to this change (Figure 4.1.5 C and D). This may facilitate flattening of the sugar ring through a half-chair conformation, which appears to be favored by a strong hydrogen bond between the sugar 2-hydroxyl and the nucleophile, and electrostatic interactions with the sugar endocyclic oxygen (Zechel and Withers, 2000). The structure of R786A has a wider entrance to the active site (Figure 4.1.5 G and H). The original arginine residue position was taken by a glycerol molecule in R786A, which changes interactions between the nitrophenyl moiety and subsites +1 and +2. This steric change may facilitate release of the product molecule from the active site. Furthermore, the widened entrance of the R786A active site provides more freedom for water molecules to diffuse in and out, which may benefit the hydrolysis process. The transglycosylation function of the TxGH116 enzyme was comparable level in the two R786 mutants indicate the mutation mainly affects their hydrolysis activity. WT and R786 mutants show similar levels of transglycosylation products when reactions have single substrates, but R786 mutants produce more product (red boxes) when have both pNPGlc and cellobiose substrate in the reaction (Figure 4.1.7). These free waters will initiate the catalytic residues protonation and

hydrolysis of the covalent intermediate. This R786A mutant also has high hydrolysis activity with pNP-xylopyranoside substrate, which lacks the C6OH group, reflecting the role of R786 in establishing specificity for glucopyranoside substrates (Figure 4.1.6).

The E777 residue mutations, which eliminated its hydrogen bond interaction with C6OH, also show large scale activity reduction, which further confirmed the important role of binding the C6OH group. The improved pNP-Xyloside substrate hydrolysis ability (Figure 4.1.6) benefit from binding the glucosyl side chain of their substrates in a conformation that stabilizes the transition state within the active site for substrate hydrolysis (Kancharla and Crich, 2013; Quirke and Crich, 2020). This stereochemical related enzyme and substrate specificity-rate tradeoffs indicates the role of specific residues in fulfilling functional enzyme selectivity and catalytic efficiency (Tawfik, 2014).

E730 residue mutants show significantly high activity at relative low pH (Figure 4.1.3 A), likely indicating a change in the environment of the catalytic acid/base, D593. E730A showed a small increase in active site space compared with WT and E730Q but less than R786A (Figure 4.1.8). With a relatively higher Ki value from glucose inhibition treatment of the E730A mutant, its structure shows relatively little difference compared with WT and E730Q, which led to loss of deprotonation of the amino acid side chain (Figure 4.1.8). Glucose tolerance of E730A is 4-fold higher than E730Q and 7-fold higher than WT, in terms of Ki. The E730A mutant lost the hydrogen bond interactions with both the W732 and R786 residues and also led to a steric change. The E730Q mutant has a similar stereochemical configuration as WT but lost the coulombic (charge-charge) interaction with Arg786 (Figure 4.1.8). Reduced hydrolysis activity of E730 residue mutants further confirmed the

importance of the R786 residue in substrate hydrolysis. E730 forms hydrogen bonds/salt bridges with W732 indole groups and Arg786 guanidino, R786 would donate a proton in hydrogen bond to the glucose C6 hydroxyl oxygen (Figure 4.1.2).

Mutation of W732 residue did not change turnover rate but led to an increased KM value compare with WT (Table 4.1.1 and Table 4.1.2), which suggest that the W732 CH-pi interaction with C6 makes a small contribution to substrate binding.

This work elucidates the roles of glycone sugar binding related residues in the TxGH116 active site in substrate binding and catalysis. When the substrate binds, the enzyme active site glycone sugar binding residues may restrict the -1 subsite substrate conformation (Quirke and Crich, 2020) and stretch or distort the glycoside bond then weaken it so that less activation energy is needed to break the bond at the start of the reaction (Tawfik, 2014). These would help in optimizing the same enzyme either for rate, or for both rate and accuracy. It will also provide useful insights to develop potent and selective inhibitors (Artola et al., 2017; Kuusk and Valjamae, 2017) for undesired GH116 family enzymes, and also support more information for potential applications (Pengthaisong et al., 2021). Design of chemical glucosyl donor substrates also needs this kind of information, because not only the enzymes themselves but also the substrates they bind are particularly flexible (Satoh and Manabe, 2013).

Table 4.1.3 Data collection and refinement statistics of glycone sugar binding related mutants. Values in parentheses are for the outer resolution shell.

Dataset	E730A-glucose complex	E730Q-glucose complex	R786A-glucose complex	R786K-glucose complex
PDB code				
Beamline	BL13B1	BL13B1	BL13B1	BL13B1
Wavelength (Å)	1.00	1.00	1.00	1.00
Data collection				
Space group	$P2_12_12$	$P2_12_12_1$	$P2_12_12$	$P2_12_12$
Cell dimensions				
a, b, c (Å)	a = 177.364	a = 174.263	a = 177.444	a = 177.285
	b = 54.708	b = 80.377	b = 54.225	b = 54.463
	c = 83.216	c = 124.458	c = 83.112	c = 83.074
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution range (Å)	30.00 – 1.85	50.00 – 2.10	30.00 – 1.80	50.00 – 1.79
Resolution outer shell (Å)	1.92 – 1.85	2.18 – 2.10	1.86 – 1.80	1.85 – 1.79
No. Unique reflections	70181	92905	73832	75984
No. Observed reflections	473312	568582	413994	546499
Completeness (%)	99.8 (99.5)	99.0 (95.2)	98.3 (90.4)	98.5 (86.7)
Average redundancy per shell	6.7 (5.6)	6.1 (3.7)	5.6 (2.6)	7.2 (5.8)
I/σ (I)	16.5 (1.88)	10.91 (1.78)	17.44 (2.05)	16.34 (1.89)
$R_{(merge)}$ (%)	0.113 (0.722)	0.184 (0.688)	0.101 (0.485)	0.121 (0.810)
CC1/2	0.765	0.766	0.635	0.708
Refinement				
R_{factor} (%)	15.53	15.95	15.08	14.81
R_{free} (%)	19.16	20.45	18.40	18.59
No. of residues in proteins	768	771/769	768	768
No. Protein atoms	6233	6223/6182	6217	6193
No. Ligand atoms	12	12/12	12	12/12
No. Other hetero atoms	57	38	55	59
No. Waters	462	692	433	469
Mean B-factor				
Protein	22.319	29.126/38.354	23.403	21.762
Ligand	17.569	27.671/35.598	23.548	22.235/9.164
Other hetero atoms	42.656	51.111	41.859	39.053
Waters	34.032	38.743	32.442	32.329
r.m.s. bond deviations (length)	0.0100	0.0085	0.0121	0.0112
r.m.s angle deviations (degrees)	1.5783	1.5260	1.7172	1.6809
Ramachandran plot				
Ramachandran favored (%)	95.16	95.24	95.29	96.07
Ramachandran outliers (%)	0.00	0.13	0.00	0.00

4.2 Importance of residues that interact with the catalytic acid/base and nucleophile residues

4.2.1 Introduction

The large and diverse family of glucoside hydrolases normally utilize two carboxylate groups in catalysis (Gideon Davies and Bernard Henrissat, 1995;

McCarter and Withers, 1994). Abundant evidence has indicated that the general acid/base catalyst initially protonates the glycosidic bond, catalysing bond fission (even through sometimes the bond breaks before protonation for active substrates) ; then the nucleophilic glutamate of the retaining enzymes form a covalent enzyme substrate intermediate (Charoenwattanasatien et al., 2016; White and Rose, 1997; Zechel and Withers, 2000). The pK_a value of glutamic acid and aspartic acid side-chains in solution are 4.5 and 3.9 respectively (Hass and Mulder, 2015). Carboxylic acid groups from aspartate and glutamate easily dissociate into a carboxylate anion and a positively charged hydrogen ion (proton) (Stewart and Yates, 1960). The negative charge that is left after deprotonation of the carboxyl group is delocalized between the two electronegative oxygen atoms in a resonance stable structure (Siggel et al., 1988). This delocalization of the electron cloud from both of the oxygen atoms are less strongly negatively charged; the leaving positive proton is therefore less strongly attracted back to the carboxylate group once it has left; hence, the carboxylate ion is more stable and less basic as a result of resonance stabilization of the negative charge (Wiberg et al., 1991). For the catalytic enzyme to be active, the proton donor acid/base must be protonated prior to catalysis, and the nucleophilic residue must be unprotonated (Bender, 1960). As the proton donor glutamic acid has to be protonated in the resting state (White and Rose, 1997), most of these types of enzyme are more active in acidic environments indicates the potential pattern that supporting the catalytic residues.

Functional enzymes from same family normally have conserved catalytic residue and relatively conserved active site residues that do not directly participant in catalysis progress (Desjardins et al., 2017). In the serine protease family, the interactions among the residues form the classical Asp-His–Ser catalytic

triad, in the special environment of the enzyme–substrate complex, activate the nucleophilic potential of the seryl Oγ (Dodson and Wlodawer, 1998; Polgár, 2005). This classic triad has been observed in a number of functional enzymes with a varity of structures, and is the first well characterized example of convergent evolution at the molecular level. Numerous of variations of the classic catalytic triad have been discovered that catalyse the hydrolysis of many classes of substrate (Madison et al., 1993). An atypical α-amylase (BmaN1) show its three-dimensional structure modeling and amino acid sequence confirmed one aberrant catalytic triad which consisting of aspartic acid, glutamic acid and histidine (Sarian et al., 2017). Phylogenetic analysis showed that this group of α-amylases comprises a new subfamily of the GH13 family. This indicates the important role of adjacent amino acid residues that mostly related with catalytic residues of functonal enzyme are worth to be identified.

The glycoside hydrolase GH116 family was verified by characterising a β-glycosidase from the hyperthermophilic archaeon *Saccharolobus solfataricus* (Cobucci-Ponzano et al., 2010) and contains mammalian GBA2 glucosylceramidase. It contain β-glucosidase (EC 3.2.1.21), β-xylosidase (EC 3.2.1.37) and N-acetyl glucosaminidase (EC 3.2.1.50) activities. *Tx*GH116 is a thermostable β-glucosidase which was the first GH116 enzyme with a high-resolution crystal structures. The acid/base of *Tx*GH116, D593, is located approximately 8 Å (carboxylate to carboxylate) away from the nucleophile E441. Its glucoside hydrolysis activity has the optimum pH at 5.5 at 60°C (Charoenwattanasatien et al., 2016). Our work here present two independent potential "proton-relay" triads that with the catalytic acid/base and nucleophile respectively. Systematic mutation of these residues shows that the "Thr-nucleophile–Tyr" pattern of catalytic residues is generally conserved, which theronine play role in both proton

dornor and acceptor function based on the demand of nucleophile residue to be deprotonated. The glutamine residue in "Gln-acid/base-Tyr" can vary to histine, valine and tyrosine which not show that highly conservation. Which tyrosine plays an essential role in balancing acid/base residue protonation. The variations in sequence and organization illustrate the adaptability shown by *Tx*GH116 enzymes in generating catalytic stereochemistry on different of GH116 family.

4.2.2 Results and discussion
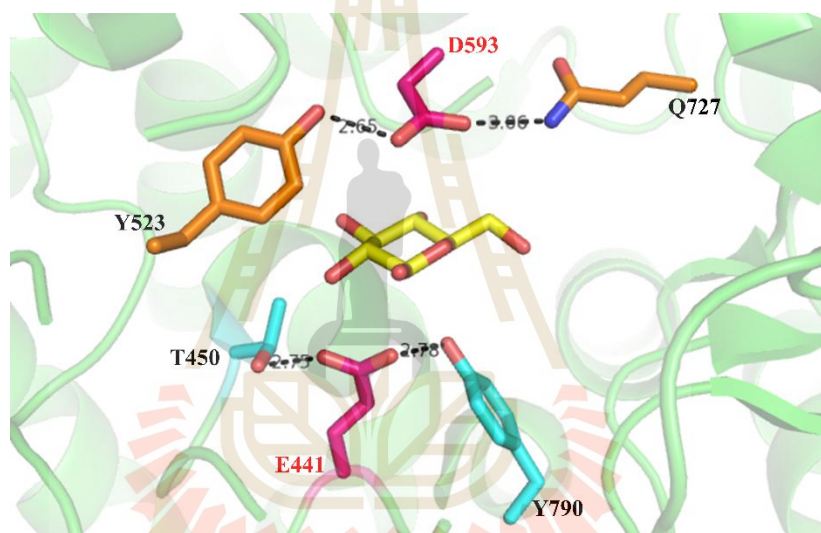
Figure 4.2.1 Residues interacting with the catalytic nucleophile and acid/base residues in the TxGH116 WT active site glucose complex.

Adjacent T450 and Y790 interact with the catalytic nucleophile residue, E441, with hydrogen bond lengths of 2.75 and 2.78 Å, respectively (Figure 4.2.1). Threonine commonly is found as a hydrogen bond proton donor capacity which induced some small negative pKa shifts observed for the interactions with adjacent acid residues (Cargile et al., 2008). In the meantime, for both D and E residues the effect on the pKa value of interactions with hydrophobic residues (A, F)

is minimal. Threonine has a high pKa value of 16, which indicates its commonly in protonated state in enzyme reaction condition, which simutaneously forms a hydrogen bond with E441 to maintain its carboxylate anion nucleophilic state to be the transition state stabilizer.

Y523 and Q727 interact with the catalytic acid/base residue, D593, with hydrogen bond lengths of 2.65 and 3.06 Å, respectively, in which Q727 Nε play role as proton donor (Figure 4.2.1). Aspartic acid residue normally have the charge-carrying capacity and hydrogen-bond potential with glutamine residue (Cargile et al., 2008). Y523 is supposed to have both potential proton donor and acceptor roles to balance the acid/base residue protonated state for glycosylation step activation. The mutations of T450 and Y790 were intended to break the hydrogen bond interaction with the nucleophile residue. The Y523 mutation also eliminated hydrogen bond to acid/base residue, but the Q727 mutations were made not only to break the hydrogen bond to D593 (Q727A) but also try to change the pKa value of the acid/base residue (Q727E).

T450 and Y790 residues which form H-bond interactions with the nucleophile residue are highly conserved in GH116 family (Figure 4.2.2). Threonine 450 was conservatived replaced with similarly nucleophilic serine and cysteine residues in some species. The residue 790 position was nearly always tyrosine, but have one species appeared to be lysine, which can also hydrogen bond with the nucleophile residue. This may indicate the importance of residue 450 and 790 position H-bond interactions with the nucleophile residue. Q727 and Y523 interact with the acid/base residue and tyrosine 523 is highly conserved but there is less conservation of glutamine at the 727 residue position. This glutamine residue can be substituted by

histidine, valine and tyrosine (Figure 4.2.2) of which histidine and tyrosine may play similar role as glutamine.
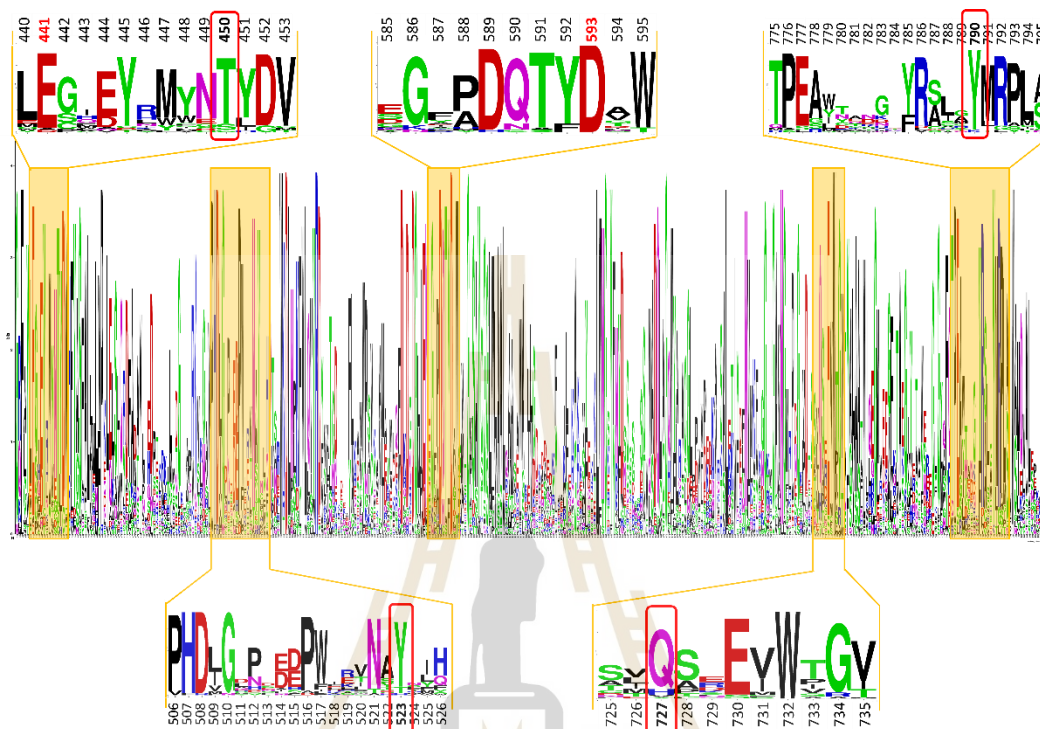
Figure 4.2.2 Sequence logo of 34 diverse sequences from GH116 family enzymes C-terminal catalytic dormain in difference species. The conserved regions around the active site are highlighted, and acid/base and nucleophile related residues' positions in the sequence are shown in bold in red box, catalytic residues are shown in red bold.

4.2.2.1 pH and temperature optima of mutants
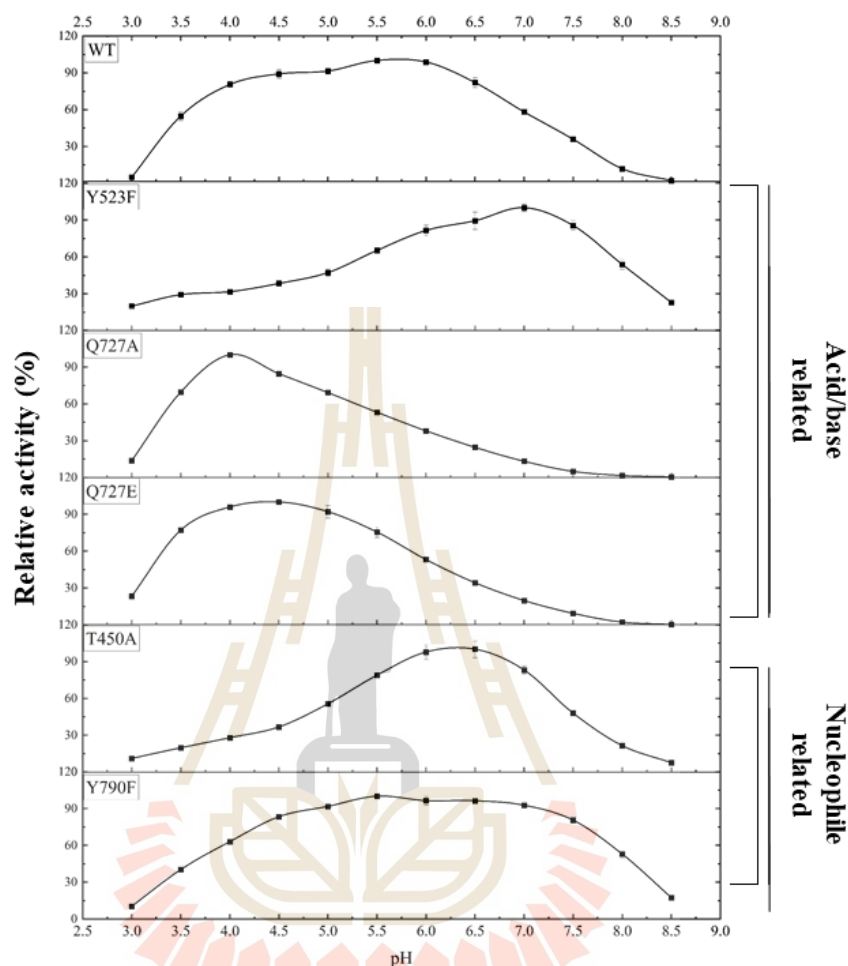
Figure 4.2.3 pH optima kinetics comparison between WT and acid/base & nucleophile related residue mutants. Data are expressed as means of three independent reactions ± SD.

Y790F mutant show broad pH profile. It got highest activity at pH 5.5, but its catalytic activity not show much difference between pH 4.5 to pH 7.5 (Figure 4.2.3). Since high pH decreases protonation, it may increase transglycosylation by increasing deprotonation of the acceptor. The broad acidic to alkaline pH optimal range of the Y790F mutant (Figure 4.2.3) indicate the special trait of adjacent residues

with nucleophile: as the threonine T450 and glutamic acid nucleophile residue E441 members are strictly conserved (Figure 4.2.2). It is likely that the nature of the proton donor tyrosine phenoliehydroxyl will attenuate the pKa of the nucleophilic glutamic acid (Shaw et al., 2002). The Y790F mutation mostly affects the nucleophile residue, which plays an essential role in glucosyl transition state stabilization (McCarter and Withers, 1994). However, Y790F still shows catalytic activity at pH 7.5 (Figure 4.2.3) which indicates the deprotonation of nucleophile residue mainly modulated by the hydrogen bond from the T450 residue. Consideration of the broad pH profile of WT, means T450 and Y790 residue are important in balancing the nucleophile residue deprotonation based of their bi-functional proton donor and acceptor ability (Bischoff and Schluter, 2012). Our TLC results of the Y790F mutant with pNPGlc and cellobiose reaction not get improved transglycosylation level compare with WT (Figure 4.2.6). This trait consist with GH3 (Seidle and Huber, 2005) and GH1 (Lundemo et al., 2017) family which indicate that hydrolytic activity decreases with increasing pH, while transglycosidic activity is pH independent. This three invariant amino acid residues in the GH116 family is supposed to have pivotal role in the hydrolysis process.

T450A mutant kinetics further confirm the importance of Y790 since it shows 20-fold higher kcat/KM value than Y790F (Table 4.2.1), but even with high hydrolytic activity it still displays high transglycosylation ability, based on its TLC of its products (Figure 4.2.6). This may because of hydrophobic influence from Y790F mutant that reduce water molecule enter into active site position where it acts in deglycosylation (the rate limiting step). The T450A mutation breaks the hydrogen bond interaction between threonine and nucleophile residue, which appears to lead to an unstable nucleophile rotation (Figure 4.2.10 A). T450A has a narrow working pH

profile compared with WT and a pH optimum at pH 6.5. This is consistent with the Y790F mutant result that T450 forms the interaction with E441 by donating a hydrogen bond to the nucleophile The T450 mutation eliminated this interaction so the nucleophile needs a lower proton concentration condition to maintain its deprotonated nucleophilic state.

Interestingly, the acid/base residue related mutants Y523F, Q727A and Q727E mutants all show narrow pH optimum profiles (Figure 4.2.3). The pH activity profiles Y523 and Q727 residue mutants shift two different ways. Y523F got high activity at neutral condition, it may from the hydrophobic effect of mutation reduce accessibility of substrate or water molecule and destabilizes the negative charge of the deprotonated state, which decreases the ability of the carboxylic acid of D593 to be deprotonated in solvent (Shaw et al., 2002). This also indicates the proton acceptor role of Y523 in WT, because Y523F mutant has high activity at high pH when acid/base residue getting less proton donation from the solvent at neutral condition. This means Y523 have the ability to accept the D593 proton on makes its environment more hydrophilic. Q727A, Q727E has a relatively low and narrow pH activity profile (Figure 4.2.3). Q727A and Q727E teach had a single high activity pH point around pH 4.0 and pH 4.5 which indicates that these two mutations change the electrostatic environment around acid/base residue that leads to a decreased pKa. When deprotonated, glutamate becomes highly hydrophilic; thus its amphiphilic nature invests glutamic acid with a unique ability to support proton to acid/base residue and then get more high activity compare with WT (Andrew Karplus, 1997). That is why Q727A and Q727E need a higher proton concentration solvent (pH 4.0 and pH 4.5) to perform their highest activity. This also confirmed the role of Q727 hydrogen bond donor to D593.
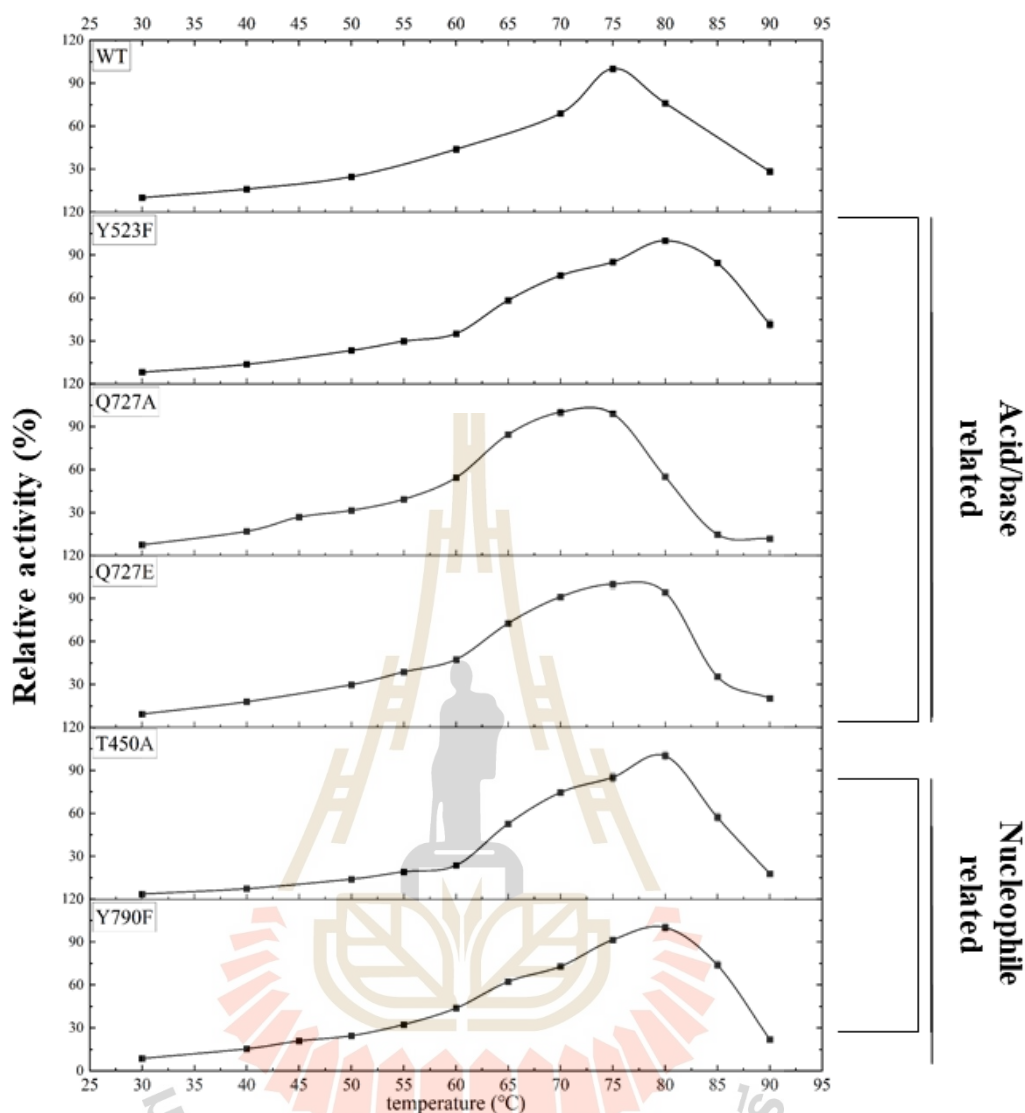
Figure 4.2.4 Temperature optima comparison between WT and acid/base & nucleophile related residue mutants. Data are expressed as means of three independent reactions ± SD.

Nucleophile related residue mutants T450A and Y790F show increased optimum temperature (80°C, Figure 4.2.4) may also come from the inert nucleophile residue without hydrogen bond interaction with T450 and Y790, which may get long distance to glycoside bond and more energy from outside may improve the movement

of the nucleophile into position for catalysis. Y523F mutant show similar thermostability as T450A and Y790F, which indicate the aromatic residue without hydrogen bond interaction to acid/base residue improve active site stability.

4.2.2.2 pNPGlc substrate kinetics with mutants of residues interacting with the catalytic amino acid residues

Table 4.2.1 Characterization of kinetic constants of the WT enzyme and mutants of residues that interact with kinetics values with pNPGlc substrate. The heat map represents the fold changes in k_{cat}, K_M, and k_{cat}/K_M each mutant relative to the WT enzyme. Red represents an improvement (i.e., increase in k_{cat}/K_M, k_{cat} and K_M) of different folds, and blue represents a decrease in k_{cat}/K_M, k_{cat} and K_M at different folds.

Protein	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1}s^{-1})
WT	0.170 ± 0.014	41.3 ± 0.63	243
T450A	3.00 ± 0.28	20.5 ± 0.38	6.84
Y790F	9.91 ± 0.91	3.11 ± 0.086	0.31
Y523F pH5.5	0.120 ± 0.0079	5.87 ± 0.067	48.9
Y523F pH6.5	0.240 ± 0.015	8.89 ± 0.12	37.1
Q727A pH5.5	1.00 ± 0.030	52.0 ± 0.60	52.0
Q727A pH4.0	0.590 ± 0.032	73.4 ± 1.3	124
Q727E pH5.5	0.250 ± 0.0059	48.0 ± 0.27	172
Q727E pH4.5	0.210 ± 0.012	66.5 ± 0.75	316

← Lower kinetic constant						Higher kinetics value →			
300-1000 fold	100-300 fold	20-100 fold	10-20 fold	3-10 fold	Neutral	3-10 fold	10-20 fold	20-50 fold	50-100 fold
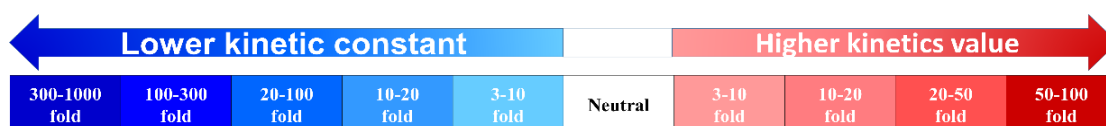
Both the T450A and Y790F mutants have reduced hydrolytic activity compared to WT with increased K_M values and decreased k_{cat} values. T450A caused a 35-fold and Y790F a 780-fold decrease in k_{cat}/K_M (Table 4.2.1). This indicates that Y790 residue has a stronger interaction to nucleophile residue and effects its catalytic ability more. The nearly 60-fold increased K_M value of Y790F shows the importance of tyrosine residue in substrate binding. The lack of hydrogen bond and increased hydrophobic effect from the phenylalanine residue is supposed to be responsible for the decreased binding. The large activity reduction from the T450 and Y790 mutations may indicate that their stabilizing interactions with the catalytic nucleophile residue, which in turn forma a hydrogen bond with the glucosyl C2OH group. In WT *Tx*GH116, the nucleophile's hydrogen bond interaction with glucose C2OH group (Charoenwattanasatien et al., 2016) restrains the orientation of the hydroxyl substituent at C2 play important role in glucose conformation limitation (Ardèvol and Rovira, 2015).

The Y523 residue interacts with the catalytic acid/base and a water molecule in position to act as a serogate acid/base (PDB: 5BX5). Y523F mutant show special kinetic characteristic compared with the Y790 residue mutant which interact with the nucleophile. Y523F has a lower KM value compared with WT but 5-fold reduced kcat/KM value (Table 4.2.1). At its pH optimum of 6.5, its KM value increased 2-fold over that at pH 5.5, but its kcat/KM value remained at a similar level. The relatively more detrimental effect of Y790F compared to Y523F maybe because of the position and interactions differences between Y523 and Y790. Y523 interact with the catalytic acid/base, which determines the glycosylation step, while Y790 interacts with the catalytic nucleophile, which decide the glycosylation rate. Y790 is located more close to subsite+1 site and Y523 is in the inner part of the -1 subsite

beside D593 residue (Figure 4.2.1). Mutation of Y523F further confirmed the interaction of the original 523 tyrosine residue with the catalytic acid/base.

The Q727 residue has the ability to form a hydrogen bond with the D593 catalytic acid/base residue (Figure 4.2.1), in which Nε acts as a proton donor. The Q727A mutant had a 6-fold increased KM value, and a 5-fold decreased kcat/KM value. Q727E displays similar kinetics as WT with a slight increased KM value and even higher kcat value than WT, especially at its optimum pH (Table 4.2.1). Both of Q727A and Q727E have lower KM values at low pH, which indicates that the mutations break the proton-relay bridge between D593 with adjacent residues, especially from the solvent, which facilitate the acid/base aspartic acid residue to be protonated prior to catalysis (White and Rose, 1997). In the case of the Q727A and Q727E mutations, the acid/base needs to have more protons in solution to be protonated and support their high hydrolytic activity, leading to relatively low optimum pH values (Figure 4.2.3). The adjacent glutamic acid E727 residue may also plays an equivalent role to acid/base residue, which stimulates enzyme hydrolysis activity (Shaw et al., 2002). This indicate the proton donor role of Q727 to acid/base residue (Park et al., 1993). The kinetics change after mutations indicated nucleophile residue interactions play more significant role in TxGH116 hydrolysis activity performance.

4.2.2.3 Cellobiose substrate kinetics with mutants that interact with catalytic residues

Table 4.2.2 Characterization of kinetic constants of the WT enzyme and catalytic residue-interacting residue mutants with cellobiose substrate. The heat map represents the fold changes in k_{cat}, K_M, and k_{cat}/K_M of each mutant relative to the WT enzyme. Red represents an improvement (i.e., increase in k_{cat}/K_M, k_{cat} or K_M) of different folds, and blue represents a decrease in k_{cat}/K_M, k_{cat} or K_M at different folds.

Protein	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1}s^{-1})
WT	0.360 ± 0.025	48.0 ± 0.40	133
T450A	1.50 ± 0.076	11.4 ± 0.11	7.57
Q727A pH5.5	0.930 ± 0.053	20.7 ± 0.25	22.2
Q727A pH4.0	1.15 ± 0.11	51.4 ± 1.1	44.7
Q727E pH5.5	0.620 ± 0.021	80.8 ± 0.99	130
Q727E pH4.5	1.54 ± 0.14	211 ± 9.1	137

← Lower kinetic constant						Higher kinetics value →			
300-1000 fold	100-300 fold	20-100 fold	10-20 fold	3-10 fold	Neutral	3-10 fold	10-20 fold	20-50 fold	50-100 fold

Similar as pNPGlc substrate kinetics. T450A show less activity in catalysis of cellobiose hydrolysis (Table 4.2.2). In contrast, both of Q727A and Q727E get quite high activity even higher than WT. At their optimum pH both Q727A and Q727E mutants have increased KM values compared with their reactions in pH5.5 (Figure 4.2.1). Q727A has a 3-fold increased kcat/KM value at pH 4.0 compared with the one at pH 5.5, which is similar to its pNPGlc reaction kinetics (Table 4.2.1). Q727E display similar high kcat/KM values at pH 5.5 and pH 4.5 and

shows relatively high activity compared with WT. Especially at pH 4.0, Q727E had a more than 3-fold increased kcat value. This is consistent with the kinetics of its reaction with pNPGlc substrate (Table 4.2.1). This may come from the alternative transglycosylation reaction of the Q727E mutant when it reacts with oligosaccharides, in which Q727E shows quite high amounts of glycosyltransferase products (Figure 4.2.5, Figure 4.2.7, Figure 4.2.8). Q727E has more products from transglycosylation than T450A (Figure 4.2.5) but similar level compared to WT (Figure 4.2.7).

4.2.2.4 TLC analysis of transglycosylation from mutants related with catalytic residues

T450A shows less hydrolysis and transglycosylation products compared with Q727E mutant, although it was not clear how much oligosaccharide transglycosylation products they produced. However, pNPGlc can be glycosylated to be pNP-cellobioside (Figure 4.2.5), but no longer pNP-linked oligosaccharides were detected. Even in the reaction system that started with only cellotriose and pNPGlc substrate at the same time, Q727E got extra C4 product instead of C5 when reaction have only C3 and C3+pNPGlc, respectively.

Q727A and Q727E show similar transglycosylation levels but with slightly increased pNPC2 product amount from their darker spot, compared with WT, which mainly got pNPC2 products that looks like pNP-laminaribiose too (Figure 4.2.7). The similar levels of glucose product released from their reactions indicated that Q727 mutants show similar hydrolysis levels compared with WT which is consistent with their kinetics (Table 4.2.1).
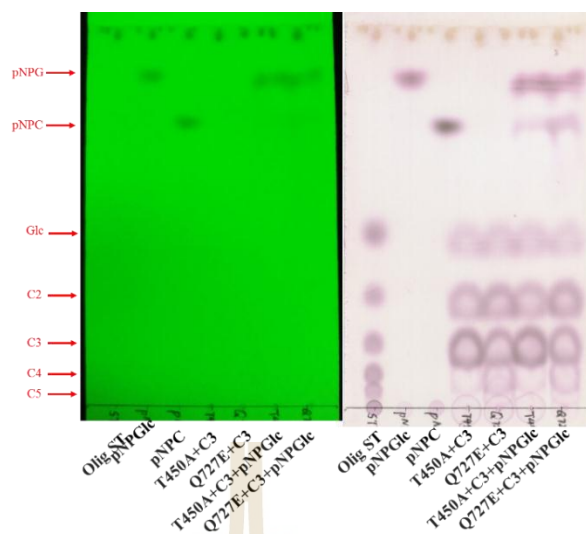
Figure 4.2.5 Transglycosylation analysis of T450A and Q727E mutants by TLC. 0.1 µg enzyme in 70 µl, 60°C 3 hours in MES buffer. The mobile phase was ethylOAc:acetic acid:H_2O in the ratio of 12:5:3 (v/v/v).
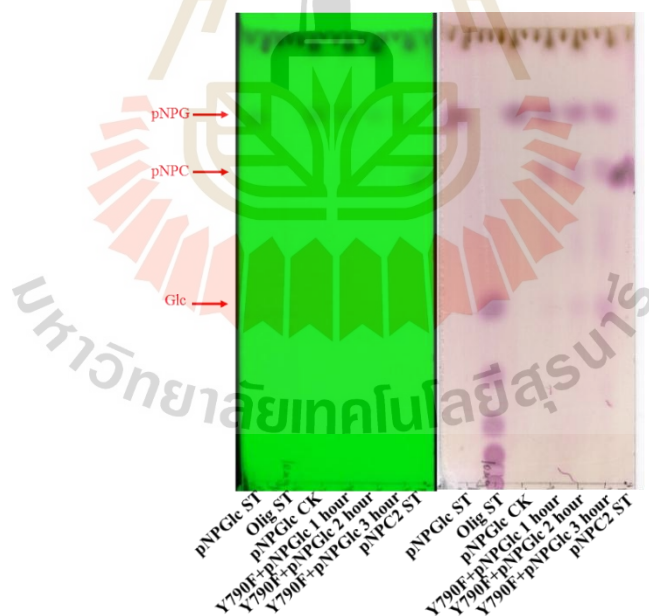
Figure 4.2.6 Transglycosylation analysis of Y790F by TLC. The reaction conditions were: 10 µg enzyme with 10 mM pNPGlc in 70 µl, 60°C 5 hours in MES buffer. The mobile phase was ethylOAc:acetic acid:H_2O in the ratio of 10:5:5 (v/v/v).
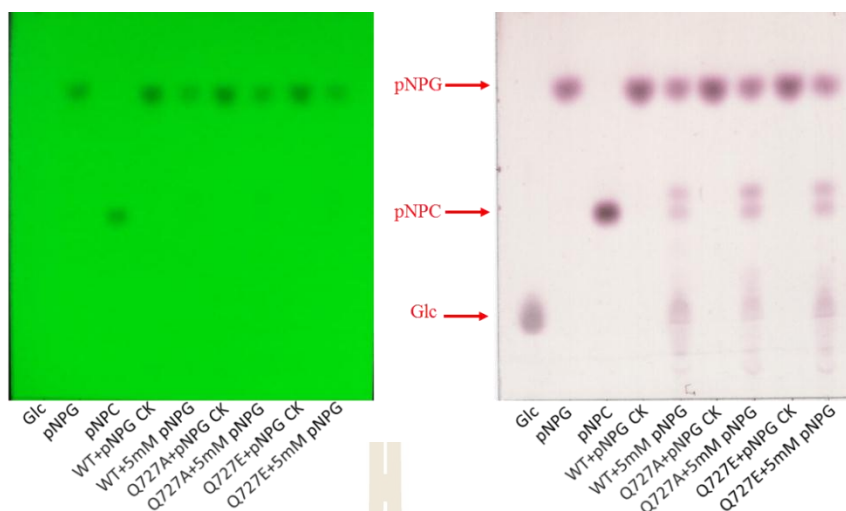
Figure 4.2.7 Transglycosylation analysis of TxGH116 WT, Q727A, E mutants with pNPGlc substrate by TLC. The reaction conditions were: 0.05 μg WT and Q727A, E in 70 μl, 60°C 3 hours in MES buffer. The mobile phase was ethylOAc:acetic acid:H_2O in the ratio of 12:5:3 (v/v/v).
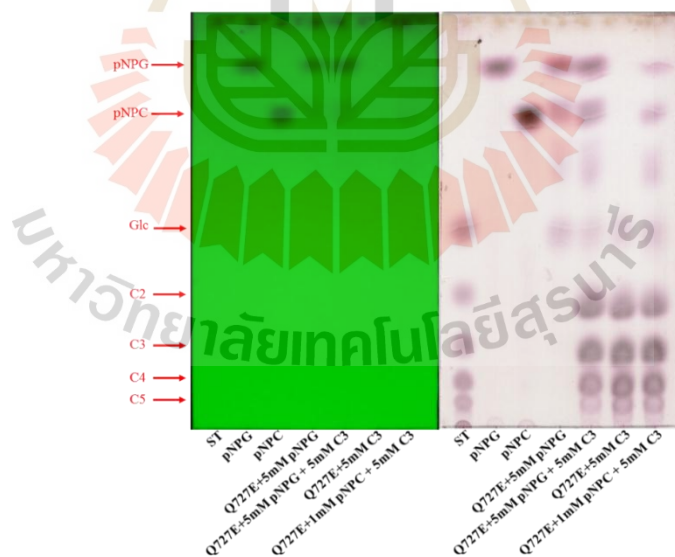
Figure 4.2.8 Transglycosylation analysis of Q727E mutant with multiple substrates by TLC. The reaction conditions were: 0.05 μg Q727E in 70 μl, 60°C 3 hours in MES buffer. The mobile phase was ethylOAc:acetic acid:H_2O in the ratio of 12:5:3 (v/v/v).

TLC results showed that Q727E mutant will produce longer chain oligosaccharides when it reacts with oligosaccharide substrates (Figure 4.2.5). Y790F shows a much high level transglycosylation ability, since with long time incubation of pNPGlc substrate at least two extra products came out below pNPGC2 product (Figure 4.2.6). When the reaction has pNPGlc and cellotriose substrate, it seems to use the oligosaccharide as a transglycosylation acceptor molecule (Figure 4.2.5, Figure 4.2.7, Figure 4.2.8) because Q727E will more frequently hydrolyze the glycoside bond in pNPGlc than that in the oligosaccharide. The products from the reaction have both pNP-β-cellobioside and cellotriose also showed that the transglycosylation products accumulated more than the one with pNPGlc (Figure 4.2.8).

4.2.2.5 Glucose inhibition kinetics of TxGH116 Q727 mutants compared to WT

Table 4.2.3 Glucose inhibition kinetics of Q727 residue mutants.

Protein	K_i(mM)
WT	4.5
Q727A	11.46
Q727E	2.82

We setup glucose inhibition kinetics for Q727A and Q727E, which show relatively high hydrolysis activity. Both of them get competitive inhibition from glucose. Q727A show more glucose tolerance than WT (Table 4.2.3). Theoretically the one that has more transglycosylation ability is supposed to have more glucose tolerance. But, this did not happen in this case, so it may come from the increased size

of the active site in Q727A mutant which has a wide space for diffusion of the free glucose molecule that released from hydrolysis.

4.2.2.6 Hydrogen bond formation ability analysis of acid/base and nucleophile related mutants with glucose ligand
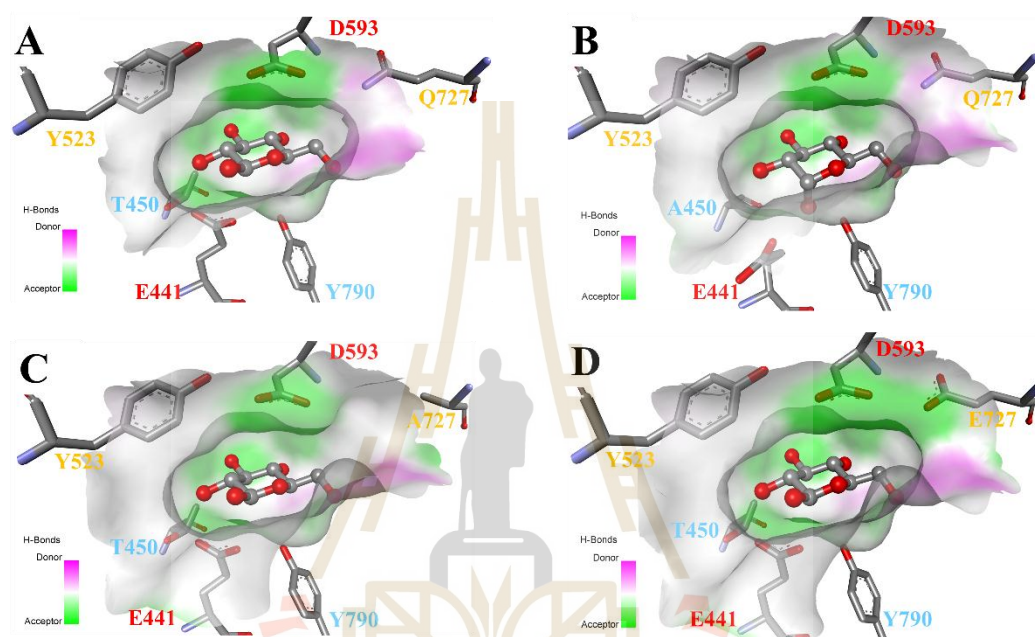
Figure 4.2.9 Catalytic related residues' hydrogen bond formation ability comparison. WT (A), T450A (B), Q727A(C) and Q727E(D) with glucose ligands. Catalytic residues are shown in red, glycone sugar binding related residues in blue. Surface colors represent the hydrogen bond donor/acceptor. Hydrogen bond donors are colored in green and hydrogen bond acceptors in purple.

Structural changes observed in the T450A mutant mainly happened in the ligand conformation that got alpha-glucose in the active site with nucleophile residue shifted far from the WT orientation (Figure 4.2.9 A, B). This change likely to

the large scale hydrolysis activity reduction in T450A mutant. Consider about the hydrogen bond interaction between nucleophile residue with glucose C2OH group, this residue orientation change eliminated the E441 H-bond interaction with the glucose C2OH group, which is critical for flattening of the sugar ring to make it toward the transition state (Ardèvol and Rovira, 2015).

The Q727A mutation that stereochemical space of the active site, causing its mouth to be more open compared with WT (Figure 4.2.9 C), but the Q727E mutation active site mouse is similar to WT (Figure 4.2.9 D). This is corresponding with the kinetics results that TxGH116 Q727A has a high kcat value than WT when reacting with pNPGlc substrate (Table 4.2.1), although less activity with cellobiose substrate (Table 4.2.2), suggesting that the enlarged space may facilitate product molecule releasing. The difference in pNPGlc vs cellobiose may reflect a different rate-limiting step. This is consistent with glycone sugar binding related mutant R786A (unpublished data), which also has a more open active site. Q727E got strong deprotonation ability to increase the negative charge around the acid/base residue, which simultaneously improves enzyme hydrolysis activity (Figure 4.2.9 D). This supposed to be able to increase the glycosylation step which the glycosidic bond was protonated from D593 (Shaw et al., 2002) since more strong proton acceptor function around acid/base residue (Figure 4.2.9 D). Q727E high hydrolysis activity with both pNPGlc and cellobiose substrates (Figure 4.2.5, Figure 4.2.7, Figure 4.2.8) suggests a facilitating role of E727 in acid/base residue deprotonation. Internally sequestered acid/base carboxylate groups are often protonated because of a considerably increased pKa, due to the low internal dielectric constant. If glutamine 727 mutant to glutamate, one would expect the pKa of the acid/base to increase due to charge-charge repulsion in the deprotonated form.

Instead, the overall effect seems to be to decrease the pKa of the solvent exposed proton donor residue D593, by coupling it to the internally sequestered tyrosine 523, via a potential charge relay mechanism (Blow et al., 1969).

4.2.2.7 Active site interaction analysis of acid/base and nucleophile related mutants with glucose and G2F ligand

Figure 4.2.10 Structural analysis of acid/base and nucleophile residues related mutants with ligands. The active site structures shown are: T450A (cyan), Q727A (blue B and purple C) and Q727E (yellow) with glucose ligands (A, C, D) and G2F ligand (B). A, B, C, D show the Fo-Fc electron density omit maps of glucose and G2F ligands in active sites in yellow box. E-L show the interactions with the acid/base and nucleophile and ligands. Catalytic have red label. M, N show the superimposed active sites of WT (green) and mutants. Potential H-bonds are represented by dashed lines.

The structure of T450A socked with glucose has a glycerol molecule in the +1 site with an unproductive nucleophile residue rotation (Figure 4.2.10 A), in which the glycerol molecule also interacts with the D593 and Q727 residues (Figure 4.2.10 E and I). Q727A and Q727E have a MES molecule from the crystal precipitant solution (Figure 4.2.10 C and D). Interestingly, Q727A has a G2F molecule in the +1 site with its alpha face opposite to aromatic Y445 residue (Figure 4.2.10 F and J) in the crystal socked with G2F ligand, in which the conformation of T450 changed (Figure 4.2.10 B). This is consistent with a nucleophile mutation that induced T450 multiple rotation in the E441A mutant with α-GlcF ligand (Pengthaisong et al., 2021). This indicates that the interaction between the T450 hydroxyl and the E441 sidechain carboxyl oxygen in the wild type enzyme stabilizes the position of both T450 and E441. In addition, the +1 site G2F molecule C3OH and C4OH groups interact with D593 (around 3 Å) indicate the +1 site ligand conformation maybe related with the acid/base residue. The +1 site G2F molecule orientation (Figure 4.2.10 J) is flipped compared with the potential glucose moeity binding in oligosaccharides (Pengthaisong et al., 2021). This conformation change in the +1 site ligand molecule may be related with the formation of two spots of transglycosylation products with pNPGlc substrate (Figure 4.2.7) in which the glycoside bond may form in either 1,3 or 1,4 linkage. Curiously, the MES molecule in the +1 site of Q727A (Figure 4.2.10 G and K) and Q727E (Figure 4.2.10 H and L) mainly is stabilized by the Y62 and Y447 residues which form hydrogen bonds with and without water molecules to the MES molecule morpholino oxygen group and ethanesulfonic acid group (Figure 4.2.11). This special affinity change with the subsite+1 site molecule may be related to glycosyltransferase function (Figure 4.2.7 and Figure 4.2.8). The glucose ligand in T450A shows a different conformation from that in WT, but Q727A and Q727E is

highly consistent with WT glucose ligand (Figure 4.2.10 A, M and N). Amino acid residues also show highly consistent position in the superposition of mutants with WT, except the T450 residue in the Q727A G2F complex (Figure 4.2.10 B, M and N). The specific binding characteristic in the Q727 mutants' +1 subsite indicates some potential extra function that needs to be elucidated in the future.
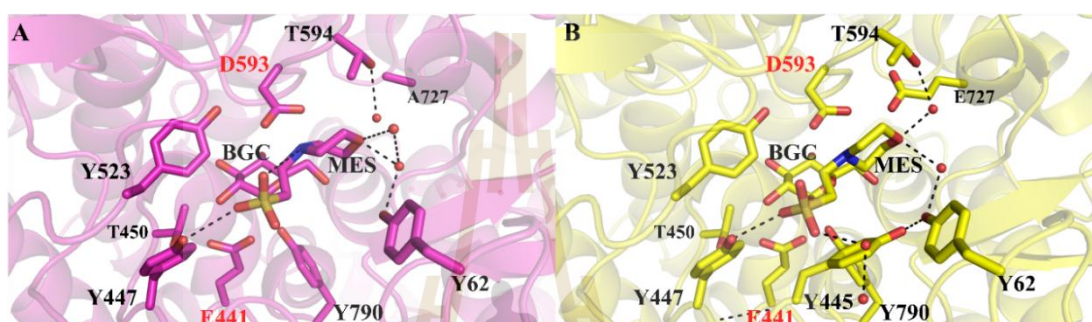
Figure 4.2.11 Subsite +1 MES molecule interactions in the Q727 mutants. The Q727A glucose complex (A) and Q727E glucose complex (B). The MES molecule morpholino oxygen group interacts with two water molecules (red balls) that bridge it to Y62, its ethanesulfonic acid group interacts with the Y447 residue by a hydrogen bond in the Q727A mutant (A). In the Q727E mutant, T594 and Y62 form water-mediated hydrogen bond interactions to MES morpholino oxygen group. The Y445 main chain interacts with the MES ethanesulfonic acid group via a hydrogen bond network with two water molecules. The MES ethanesulfonic acid group interacts with the Y447 residue by a direct hydrogen bond.

Table 4.2.4 Data collection and refinement statistics of acid/base and nucleophile interacting residue mutants. Values in parentheses are for the outer resolution shell.

Dataset	T450A-glucose complex	Q727A-G2F complex	Q727A--glucose complex	Q727E-glucose complex
PDB code				
Beamline	BL13B1	BL13B1	BL13B1	BL13B1
Wavelength (Å)	1.00	1.00	1.00	1.00
Data collection				
Space group	$P2_12_12$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12$
Cell dimensions				
a, b, c (Å)	a = 177.119 b = 54.795 c = 83.159	a = 172.216 b = 80.374 c = 126.315	a = 174.080 b = 80.424 c = 124.988	a = 177.644 b = 54.749 c = 83.138
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution range (Å)	50.00 – 1.80	50.00 – 1.95	50.00 – 2.20	50.00 – 1.85
Resolution outer shell (Å)	1.86 – 1.80	2.02 – 1.95	2.28 – 2.20	1.92 – 1.85
No. Unique reflections	76591	133716	91771	70610
No. Observed reflections	503160	919284	604621	467266
Completeness (%)	99.7 (99.2)	99.9 (99.9)	99.9 (100.0)	99.9 (99.5)
Average redundancy per shell	6.4 (4.4)	6.9 (5.4)	6.6 (5.5)	6.6 (4.2)
I/σ (I)	17.13 (3.40)	12.43 (2.07)	13.77 (3.26)	14.61 (1.92)
$R_{(merge)}$ (%)	0.165 (0.775)	0.191 (1.199)	0.217 (0.950)	0.133 (0.789)
CC1/2	0.759	0.782	0.736	0.707
Refinement				
R_{factor} (%)	14.52	15.35	15.26	14.73
R_{free} (%)	17.75	19.49	20.07	18.56
No. of residues in proteins	769	773/769	771/769	771
No. Protein atoms	6303	6297/6248	6243/6230	6307
No. Ligand atoms	12	23/23	12/12	12/12
No. Other hetero atoms	43	62	32	45
No. Waters	387	651	475	367
Mean B-factor				
Protein	21.019	20.883/25.745	30.667/37.440	22.575
Ligand	18.591	31.777/36.324	27.117/33.577	17.814/42.255
Other hetero atoms	37.588	36.835	47.603	40.327
Waters	29.901	29.122	35.705	31.174
r.m.s. bond deviations (length)	0.0108	0.0143	0.0085	0.0121
r.m.s angle deviations (degrees)	1.6277	1.9077	1.5440	1.7087
Ramachandran plot				
Ramachandran favored (%)	96.47	95.70	95.11	95.45
Ramachandran outliers (%)	0.00	0.26	0.33	0.13

4.3 Impact of Aromatic Stacking in Positive Subsites on *Tx*GH116 Glycoside Hydrolysis and Transglycosylation Activity

4.3.1 Introduction to aromatic stacking in glycoside hydrolysis

The recognition and interactions between saccharides and proteins has far reaching implications in biological, pharmacological, and technological applications (Bertozzi and Kiessling, 2001; Seeberger and Werz, 2007; Vacas et al., 2010). Conventional hydrogen-bonding interactions are usually involved in carbohydrate recognition to determine the enzyme substrate specificity (Holliday et al., 2009). But less expectedly, researchers observed that despite the highly hydrophilic character of most sugars, aromatic rings of the enzyme often play an important role in protein folding (Newberry and Raines, 2019) and carbohydrate recognition (Andrews et al., 2000; Del Carmen Fernandez-Alonso et al., 2005; Hamre et al., 2015; Tagami et al., 2013; Uchiyama et al., 2001). Some examples have been reported in which tyrosine or phenylalanine residues seem to assist the cleavage/formation of glycosidic linkages by presenting favorable interactions with the carbohydrate transition states in glycoside hydrolases (GHs) (Nerinckx et al., 2003). It was commonly accept that noncovalent interactions mediated by aromatic rings are pivotal to sugar binding (Laughrey et al., 2008; Spiwok, 2017). For example, aromatic residues often stack against the faces of sugar pyranose rings in complexes between proteins and carbohydrates (Hsu et al., 2016). Such contacts typically involve two or three CH groups of the pyranoses and the π electron density of the aromatic ring (called CH/π bonds), and these interactions can exhibit a variety of geometries, with either parallel or nonparallel arrangements of the aromatic and sugar units (Juan et al., 2013). Thus, the hydrophobic platform comprising a phenylalanine residue, which is highly

conserved in the active center of many glycoside hydrolases, has been proposed as a mechanistically relevant transition state stabilizing factor (Nerinckx et al., 2003). A notable feature in processive GHs show one special feature is the ubiquity of aromatic and polar residues lining the enzyme tunnels and clefts (Andrews et al., 2000; Hefferon et al., 2019). The conserved tryptophan residues govern the turnover rate of processive cellulase (Rojel et al., 2020). These residues were thought mainly responsible for substrate chain acquisition, binding, and hydrolysis in the catalytic tunnel/cleft via carbohydrate-π stacking, hydrogen bonding and salt bridge interactions (Juan et al., 2013). These non-bonded interactions are usually assumed to facilitate the processive mechanism whereby the enzyme must maintain attachment to the substrate while still allowing forward processive motion (Hamre et al., 2015; Payne et al., 2011; Zakariassen et al., 2009).

A novel mechanism with kinetic, structural and theoretical considerations, has made by making use of a tyrosine residue, instead of a carboxylate, to stabilize the oxocarbenium-like transition state generated during the frontface attack of the glycosidic acceptor to the acetalic center (Iglesias-Fernandez et al., 2017). Interestingly, this tyrosine residue was originally identified to be involved in parallel stacking with the α-face of the glycosidic donor and therefore seems to play a dual role, participating in both in substrate recognition and catalysis (Montalvillo-Jimenez et al., 2019). These examples illustrate that under appropriate conditions, aromatic platforms can be employed to stabilize glycoside transition states (Rojel et al., 2020).

The GH116 glycoside hydrolase family was established upon characterization of a β-glycosidase from the hyperthermophilic archaeon *Saccharolobus solfataricus* (Cobucci-Ponzano et al., 2010) and contains mammalian GBA2 glucosylceramidase, which contain β-glucosidase (EC 3.2.1.21) and β-xylosidase (EC 3.2.1.37) activities. The roles of

aromaticresidues in the *Tx*GH116 β-glucosidase slot like active site remain to be elusidated. To gain further insights into factors influencing their interactions with substrates, the *Tx*GH116 family was selected for research and a bioinformatic analysis, as well as biochemical characterization were carried out. Our aim is to understand the influence of carbohydrate/aromatic stacking on the pyranose reactivity to eventually elucidate the roles of aromatics in the GH116 family. The catalytic efficiency and the mode of action of the enzyme changed dramatically after subsite site +1 and +2 aromatic residues were systematically mutated. It is hoped that the key findings in this study will have wide applications in guiding the rational design of other more divergent GH116 enzymes.

4.3.2 Results and discussion

4.3.2.1 Aromatic residue roles and position in *Tx*GH116 subsite+1,+2 with glucose and G2F ligand
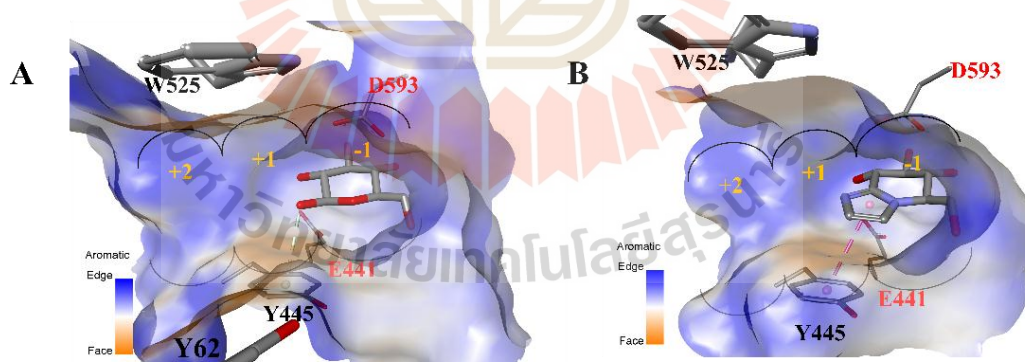
Figure 4.3.1 Aromatic residues in the TxGH116 active site -1 and subsites +1 and +2 in the glucose and glucoimidazole complex. A, Glucose (PDB: 5BX5) and B, GIM (PDB: 5BX4) ligand in WT structure (Charoenwattanasatien et al., 2016). Yellow color shows aromatic faces around the ligand. Catalytic residues are labeled in red. Active

site subsites are indicated by black semicircular curves. Cyan dash

represent OH/pi interaction, pink dash represent pi/pi interaction.

Aromatic residues around the TxGH116 active site mainly include

Y445 and W525 in subsites +1 and +2 (Figure 4.3.1 A). Furthermore, the W732

residue also shows a weak aromatic interaction with C6OH group. The distance

between the center of the Y445 aromatic ring to glucose ligand O5 and C1 is around 5

Å, and glucose analogue inhibitor GIM imidazole ring center to Y445 center around

3.9 Å (Figure 4.3.1 B), which indicates the possibility of the glucose 4H3 half chair

oxycarbonium transition state cation to form a cation-pi interaction (within 6 Å)

(Gallivan and Dougherty, 1999) to be stablized by the Y445 aromatic residue. The

distance between the W732 aromatic ring center to glucose C5 around 5.5 Å and to

C6 is around 5 Å, which indicates the lower possibility to form CH-pi interaction

between C5H and C6H with the W732 aromatic ring. The defined criteria for CH−pi

stacking show a distance between hydrogen and the aromatic ring below 4.5 Å and

deviation of CH hydrogen from aromatic ring axes below 45° (Houser et al., 2020).

4.3.2.2 Aromatic residues conservation analysis in the GH116 family

Amino acid conservation analysis was performed to assess the

conserved residues in GH116 enzyme family (Figure 4.3.2). Like the catalytic

residues, two aromatic residues Y445 and W732 show high conservation in most

species, indicating that they potentially play essential roles in enzyme function. In

contrast, the W525 and Y62 (data not show) residues show very low conservation

between difference species, this maybe because of the specific location of W525 in

subsite +1 and +2 site may determine substrate specificity and is less important for general enzyme function performance.
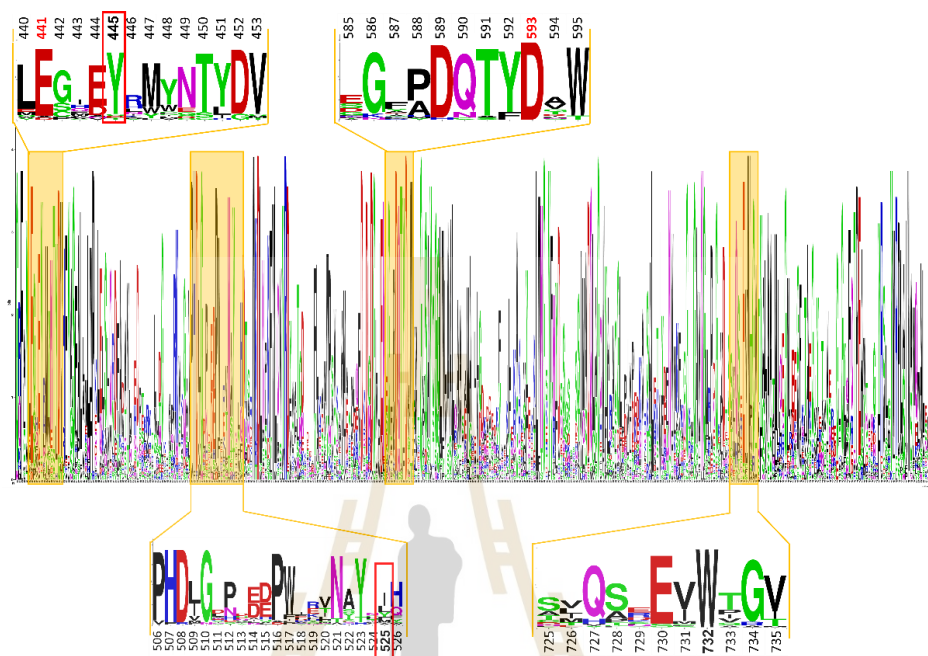
Figure 4.3.2   Sequence logo of 34 diverse sequences that from GH116 family enzymes from difference species. The conserved regions around the active site subsites +1 and +2 are highlighted, and their positions in the sequence are shown in bold, catalytic residues are shown in red bold.

4.3.2.3 pH and temperature optima of mutants

Y445L has highest activity at pH 4.5, which is two-fold higher than that at the commmon reaction condition of pH 5.5 for TxGH116 WT enzyme (Figure 4.3.3). Its narrow working pH profile compared with other wide ones indicate tyrosine to leucine has an effect on the catalysis process. W525 mutants show wide working pH profiles compared with the Y445L mutant but these are similar to WT. W525F had an even wider working pH profile than WT. At pH 7, W525F still have

more than 80% activity compared with its highest hydrolysis activity at pH 6 (Figure 4.3.3). W525L has a similar working pH profile to WT, but its optimum working pH shifted a little bit to pH 5. From this pH optimization result, we can know that the Y445 residue shows a more pivotal role in maintaining the TxGH116 active site environment compared with W525. This may be because Y445 not only acts in CH-pi interactions in subsite +1 but also has a hydrogen bond interaction with R786, which further affect substrates affinity and hydrolysis. Also, the potential cation-pi interaction role of Y445 in stabilizing the transition state was broken by the Y445L mutation (Iglesias-Fernandez et al., 2017; Nerinckx et al., 2003). The W525 residue mostly supports CH-pi interaction at subsites +1 and +2 and may not be essential for catalysis. The stereochemical change of W525F change substrate interactions at subsites +1 and +2.
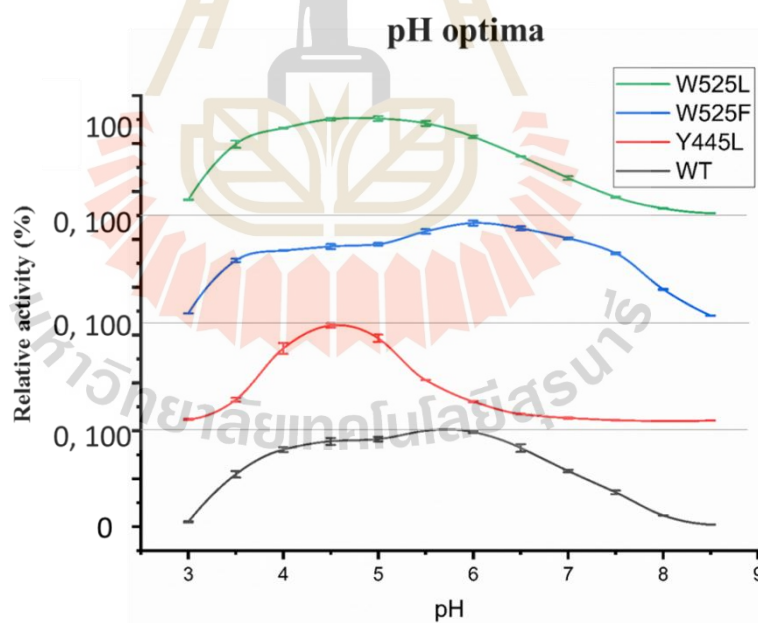
Figure 4.3.3 pH optima of aromatic residues. In order to get pH and temperature adaption trendency. Relative activity were obtained from comparison with their highest activity value in different mutants, respectively. Data are expressed as means of three independent reactions ± SD.
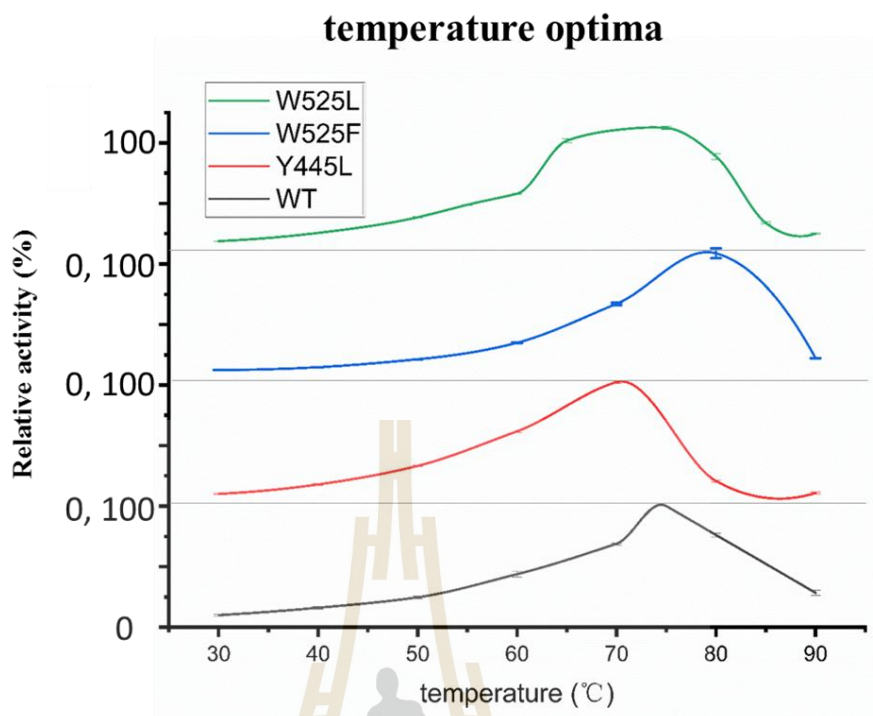
Figure 4.3.4 Temperature optima of aromatic residues. In order to get pH and temperature adaption trendency. Relative activity were obtained from comparison with their highest activity value in different mutants, respectively. Data are expressed as means of three independent reactions ± SD.

The Y445L and W525L mutants show competitive higher hydrolysis activity compared with WT at 70°C, but W525F shows more stability at high temperature like 80°C (Figure 4.3.4). Actually Y445L has less stability at high temperature compared with WT and W525 mutants. This is consistant with pH optima results which show Y445 play more pivotal role in shaping the catalytic environment. Mutation of the W525 residue to phenylalanine improves the subsite +1 site environment to get more strong affinity of substrate binding and also protein~substrate complex stability.

4.3.2.4 Kinetics comparison between TxGH116 WT and subsite +1 and +2 residue mutants with pNPGlc substrate at different pH values and temperatures

Table 4.3.1 Characterization of kinetic constants of the WT enzyme and subsite +1 and +2 mutants with pNPGlc substrate. The heat map represents the fold changes in k_{cat}, K_M, and k_{cat}/K_M of each mutant relative to the WT enzyme. Red represents an improvement (i.e., increase in k_{cat}/K_M, k_{cat} or K_M) of different folds, and blue represents a decrease in k_{cat}/K_M, k_{cat} or K_M at different folds.

Protein	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1}s^{-1})	Reaction temperature (℃)	Reaction pH
WT	0.170±0.014	41.3±0.63	243	60	5.5
Y445L	5.37±0.26	17.0±0.21	3.17	60	4.5
Y445L	3.23±0.25	5.48±0.086	1.70	60	5.5
W525F	0.0740±0.0026	26.5±0.13	359	60	5.5
W525F	0.150±0.0073	59.3±0.53	396	70	5.5
W525F	0.420±0.023	149±2.6	355	80	5.5
W525L	0.410±0.020	120±1.2	293	60	5.5
W732F	0.470±0.010	62.3±0.39	133	60	5.5

Decreasing kinetic constant value ← | **Increasing value** →

| 100-300 fold | 20-100 fold | 10-20 fold | 3-10 fold | Neutral | 3-10 fold | 10-20 fold | 20-50 fold |

From pNPGlc substrate kinetics, Y445L mutant get 8-fold reduced hydrolysis activity and 140-fold enzyme substrate specificity compared with WT

(Table 4.3.1). It confirmed the importance of interactions between Y445 residue and the substrate molecule. W525 residue mutants have similar or higher activity than WT.

Especially, the W525L mutant has a 3-fold k_{cat} value compared with WT at pH 5.5, which means its turnover rate for hydrolysis is 3 times higher than WT. But W525F has a lower K_M value and half turnover rate compared with WT in common reaction condition (pH 5.5, 60°C), which leads to a higher enzyme substrate specificity. This increased specificity constant k_{cat}/K_M value is even higher than the W525L mutant. The K_M value of W525F increases with increasing reaction temperature (from 60-80°C) (Table 4.3.1), and, its k_{cat} value also increases with the similar ratio to K_M value, which leads to similar enzyme substrate specificity at 60, 70 and 80°C reaction conditions. This means its reaction mechanism remains the same in different temperatures. The kinetics of W525F and W525L with high k_{cat}/K_M and k_{cat} values confirmed the important role of pi-pi interaction with phenyl group from the pNPGlc substrate. The eliminated pi-pi interaction in the W525L mutant from its original tryptophan residue may have improved hydrolysed product release from active site leading to the higher turnover rate (Table 4.3.1 and

Figure **4.3.5** A). The W525F mutation leads to a stereochemical change of pi-pi interaction and stacking platform support from both F525 and Y445, which increased its enzyme substrate specificity with pNPGlc substrate and also thermostability. From this we found that Y445 is more essential than W525 in *Tx*GH116 hydrolysis of *p*-nitrophenyl glucoside. This may be because of the specific position of tyrosine 445 parallel with oxocarbenium cation plane, which may play a role in stabilizing the transition state (Montalvillo-Jimenez et al., 2019). This is similar to the processive cellulase Cel7A in which turnover rate is modulated by one conserved tryptophan residue (Rojel et al., 2020). Also, the modification of aromatic amino acid residues in

subsites +1 and +2 that usually perform the classical aromatic-carbohydrate stacking interactions show their application in improving hydrolyzed product release and also protein-substrate thermostability (Baker et al., 2017). The W732F mutant shows comparable activity (Table 4.3.1) to WT, which indicates the weak effect on hydrolysis activity of the tryptophan to phenylalanine mutation.

4.3.2.5 Kinetics comparison between WT and subsite +1 and +2 site mutants with oligosaccharide substrates

Table 4.3.2 Characterization of kinetic constants of WT enzyme and subsite +1 and +2 aromatic residue mutants with cellobiose substrate. The values for hydrolysis were determined by dividing the amount of glucose released by 2, since two glucose molecules are released per glycosidic bond hydrolyzed. The heat map represents the fold changes in k_{cat}, K_M, and k_{cat}/K_M of each mutant relative to the WT enzyme. Red represents an increase in k_{cat}/K_M, k_{cat} or K_M of different folds, and blue represents a decrease in k_{cat}/K_M, k_{cat} or K_M at different folds.

Protein	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1}s^{-1})
WT	0.360±0.025	48.0±0.40	133
Y445L (pH4.5)	11.7±0.88	0.74±0.013	0.13
W525F	0.39±0.022	29.2±0.48	75
W525L	4.45±0.21	73.4±1.3	16.5
W732F	0.700±0.036	61.9±0.60	88.4

← Loss of function					Gain of kinetics value →		
>300 fold	20-100 fold	10-20 fold	3-10 fold	Neutral	3-10 fold	10-20 fold	20-50 fold

Table 4.3.3 Characterization of kinetic constants of WT enzyme and subsite +1 and +2 aromatic residue mutants with cellotriose substrate. For cellotriose, which may release more than one glucose molecule per substrate molecule due to sequential cleavage, the values quoted are for total glucose released, which assumes nonprocessive hydrolysis at initial rate conditions. The heat map represents the fold changes in k_{cat}, K_M, and k_{cat} /K_M of each mutant relative to the WT enzyme. Red represents an increase in k_{cat} /K_M, k_{cat} or K_M of different folds, and blue represents a decrease in k_{cat} /K_M, k_{cat} or K_M at different folds.

protein	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1}s^{-1})
WT	0.170±0.023	57.0±1.6	335
W525F	0.110±0.011	36.1±0.57	328
W525L	0.630±0.041	56.8±1.3	90.2
Y62A	0.270±0.017	81.3±1.3	301

Decreasing kinetic constant value ← → Increasing value

| >300 fold | 20-100 fold | 10-20 fold | 3-10 fold | Neutral | 3-10 fold | 10-20 fold | 20-50 fold |

In order to characterize the kinetic effects with a natural substrate, the cellobiose substrate hydrolysis reaction, in which substrate molecule have a glucose moiety instead of a nitrophenyl group at the reducing end of glycoside bond, was studied. From the oligosaccharide substrate kinetics, Y445 had a 1,500-fold decrease in enzyme substrate specificity constant (kcat/KM, Table 4.3.2), which is more extreme than its effect on pNPGlc substrate hydrolysis kinetics (Table 4.3.1 and Figure 4.3.5 A). With a huge reduction in oligosaccharide hydrolysis (Figure 4.3.5

B), we can confirm that the Y445 residue plays an essential role in the TxGH116 catalytic process. The W525 residue mutants had quite different effects on cellobiose hydrolysis kinetics. The W525F mutation did not affect the KM, but caused a 40% reduction in the turnover rate compared with WT, while W525L increased the KM 12-fold but also increased the kcat by 50% (Table 4.3.2).

The W525L mutation caused a 3.7-fold higher KM value when hydrolysing cellotriose, which around 3-fold less than its affect on the KM with cellobiose substrate (Table 4.3.2 and Figure 4.3.5 B). TxGH116 WT and W525F mutant also got lower KM values and higher kcat/KM values when hydrolyzing cellotriose compared with cellobiose (Figure 4.3.5 B), which means TxGH116 is more specific with cellotriose than cellobiose. The high kcat value from W525L with oligosaccharides may derive from the eliminated CH-pi interaction at the substrate-binding +1 and +2 sites, which accellerates product molecule release (Table 4.3.2). Sequence alignment shows human GBA2, rice GH116 and Arabidopsis AT3G (Figure 4.3.6) have the similar residue leucine and isoleucine in the position of TxGH116 W525 residue (Figure 4.3.2). Since those other enzymes are thought to act on lipid glucosides rather than oligosaccharides, this indicates that subsite +1 and +2 may play an important role in determining substrate specificity.

The W525L mutant has increased KM value with pNPGlc, cellobiose and cellotriose substrates, but W525F mutant shows lower KM value than WT (Table 4.3.1 and Table 4.3.2). The reduction of enzyme substrate specifity of W525F and W525L mutants with oligosaccharide substrate also confirmed that W525 residue essential role in substrate molecule binding. The W525L has high KM values with oligosaccharide substrates confirmed its essential role in substrate binding which mainly related with enzyme substrate specificity (Table 4.3.3). Aromatic stacking

interactions commonly happens with the α-face of the monosaccharide residues of oligosaccharides, which lead to the high fidelity rotation of subsite +1 glucose (Laughrey et al., 2008; Montalvillo-Jimenez et al., 2019). The stereochemical change from tryptophan to phenylalanine made a slight side chain rotation change which lead to F525 more facing to α-face of the subsite+1 glucose molecule (Figure 4.3.10 and Figure 4.3.11). This promoted the stacking interaction in subsite +1 site between F525 and Y445 that furtherly improved the stabilization of substrate molecule binding, including with cellotriose substrate with a longer chain. The carbohydrate-aromatic interaction strengths are moderately affected by changes in the stereochemistry and identity of the substituents on the pyranose rings of the sugars, in which the intrinsic tendency of aromatic moieties to interact with certain sugars in water solution (Del Carmen Fernandez-Alonso et al., 2005; Hsu et al., 2016).

For the stereochemical change of pi-pi interaction from W525F mutant also affects its turnover rate change when it reacts with both pNPGlc and oligosaccharide substrates, because it had 35-40% kcat value reduction in each case (Table 4.3.2). This means the pi-pi interaction of the F525 residue in W525F mutant with the phenyl group or glucosyl moiety in subsite +1 site changed, leading to improved substrate binding ability but decreased hydrolysis rate in W525F mutant. The kinetic assays of pNPGlc and oligosaccharide substrate detect a different part of the mechanism: detection of p-nitrophenyl group released from subsite +1 for pNPGlc, two glucose released from both subsites -1 and +1 for cellobiose, but one glucose released from subsite -1 for cellotriose. Nonetheless, in a steady-state kinetics assay, the rates from measuring different products are expected to be the same, due to averaging of multiple rounds of catalysis. Since wild type and W525 residue mutants got comparative kcat values for all three substrates (Table 4.3.2), the product release

from subsite -1 may be supposed to be the rate limiting step, since this step is common to all the substrates. This phenomenon maybe can be compared with another GHs enyzme, processive cellobiohydrolase, for which catalytic efficiency enhancement was more likely attributed to the increase in the binding affinity constant (kon) rather than the more stable catalytic constant (kcat) (Kurasin and Valjamae, 2011), this lead to a substrate limited processivity. The W732F and Y62A mutants had 50 and 90% increases in KM compared to WT and smaller increases in kcat (Table 4.3.2 and Table 4.3.3) for cellobiose and cellotriose, respectively. This is consistent with small contributions to substrate binding to and product release from subsites +1 and +2, respectively.

This kinetics results indicated that removal of aromatic residues at the entrance of the TxGH116 slot like active site entrance dramatically impacts the enzyme substrate binding affinity, suggesting that these residues may play roles in substrate binding and product release (Payne et al., 2011).

4.3.2.6 TLC transglycosylation comparison between WT and subsite +1 and +2 aromatic residue mutants with pNPGlc and oligosaccharide substrates

Transglycosylation level analysis of Y445L, W525F and W525L shows the low hydrolysis activity from Y445L and apparent relative transglycosylation and hydrolysis rates of the W525 mutants. The eliminated transglycosylation products from Y445L mutant (Figure 4.3.5) indicates the essential role of Y445 in acceptor molecule binding by aromatic/CH stacking interactions in TxGH116 alternative transglycosylation process, and its simultaneous contribution to oxycarbonium transition state stability. Since during the glycosylation step the oxocarbonium cation-like transition state is very high energy, which needs not only nucleophile E441 residue but also other
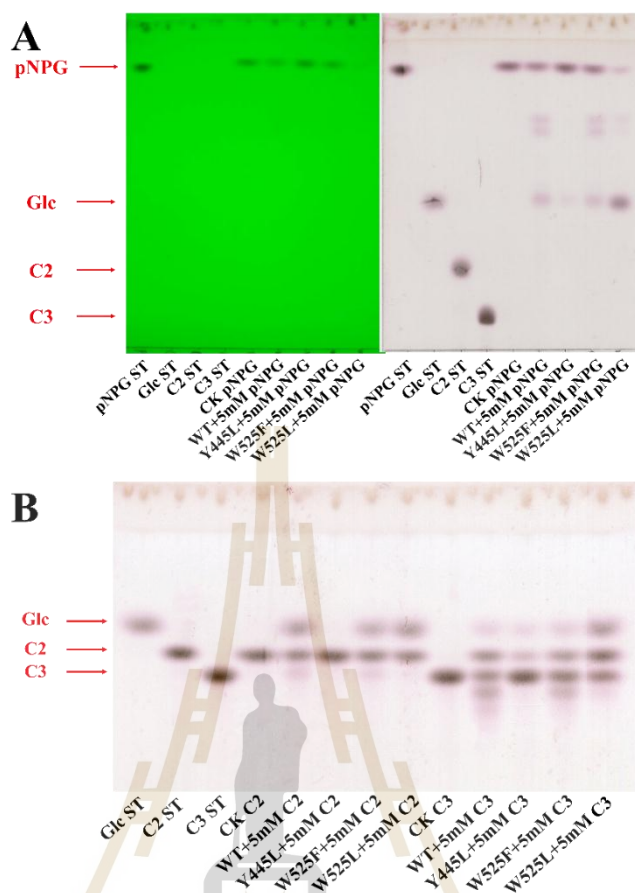
Figure 4.3.5 TLC results of transglycosylation level comparison. A, Reactions with pNPGlc (pNPG) substrate. The reaction conditions were: 0.025 μg enzyme with 5 mM pNPGlc incubated at 60°C for 3 hours in 50 mM NaOAc buffer, pH 5.5, with final reaction volume 50 μl. The mobile phase was ethylOAc:acetic acid:H_2O in the ratio of 11:5:4 (v/v/v), developed twice. B, Reactions with oligosaccharide substrates. The reaction conditions were: 0.025 μg enzyme with 5 mM cellobiose (C2) & cellotriose (C3) incubated at 60C for 5 hours in 50 mM NaOAc buffer, pH 5.5, with final reaction volume 50 μl. The mobile phase was ethylOAc:acetic acid:H_2O in theratio of 8:5:7 (v/v/v), developed twice. CK represent the control reactions without enzyme inside.

negatively charged residues and the pi electron cloud of the planar aromatic group for stabilization. Similar mechanism was previous elucidated from CGTase with two phenylalanine residues from the +1 acceptor binding site that play important and distinct roles in the transglycosylation reactions (van der Veen et al., 2001). The W525F and W525L mutants show comparable activity to WT to breakdown both pNPGlc and cellooligosaccharide substrates, although W525L produces glucose more rapidly (Figure 4.3.5). This is consistent with their kinetics constants (Table 4.3.1, Table 4.3.2, Table 4.3.3). Interestingly, W525L shows very less transglycosylation level compare with WT and W525F mutant, especially with oligosaccharide substrates. This indicated the aromatic platform spanning subsites +1 and +2 is directly related with TxGH116 transglycosylation function performance. The increased hydrolysis activity from W525L (Table 4.3.1, Table 4.3.2, Table 4.3.3, Figure 4.3.5) may also come from the low ratio transglycosylation level which means more hydrolyzed glycone molecule is transferred to water molecules, although this may not be clearly differentiated from rapid hydrolysis of transglycosylation products. The release of product molecule like p-nitrophenyl or glucose may be facilitated by the elimination of aromatic pi-pi or CH-pi interactions to p-nitrophenyl glycoside (Figure 4.3.5 A) and oligosaccharide substrates (Figure 4.3.5 B). This indicates the pivotal role of W525 in determining the ratio of hydrolysis to transglycosylation, which has been characterized from other β-glucosidases (Frutuoso and Marana, 2013).
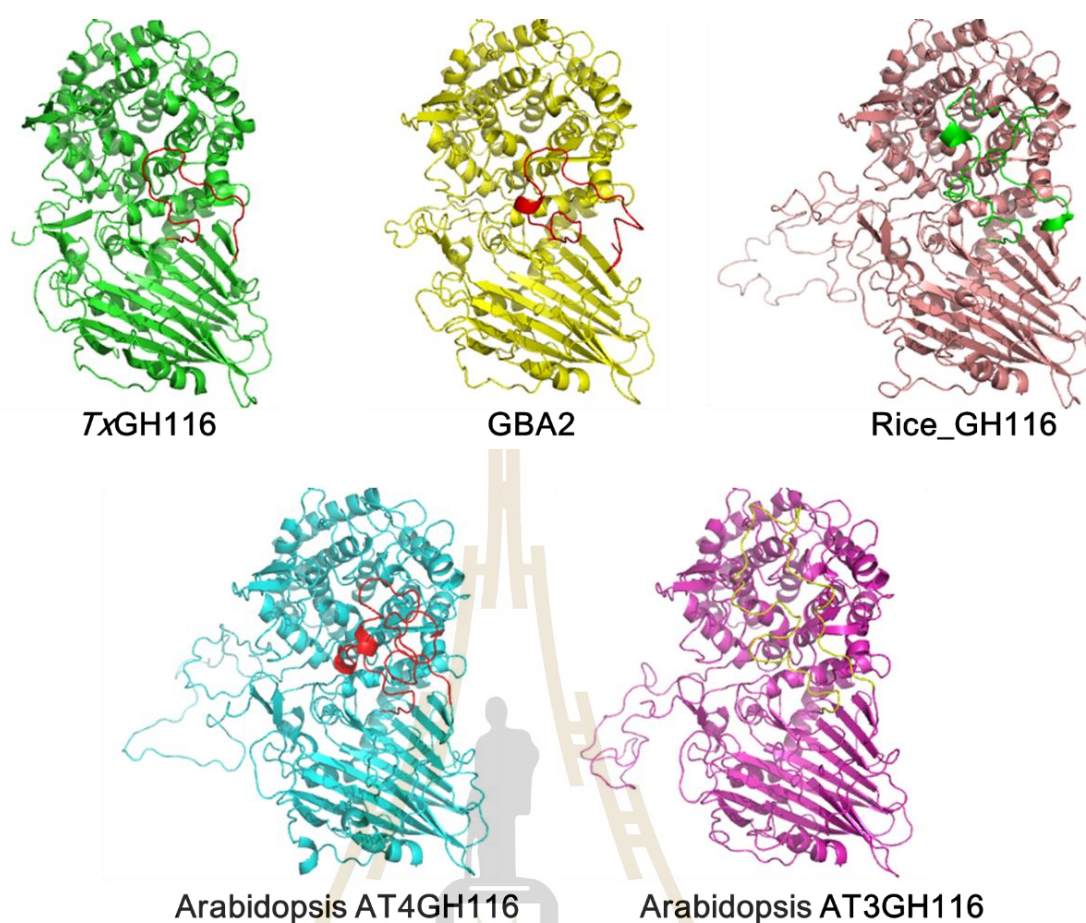
Figure 4.3.6 Comparison of the structure of *Tx*GH116 to homology models of GH116 family enzymes from human and plants. *Tx*GH116 (PDB: 5BX5), human GBA2, rice GH116, and Arabidopsis At4GH116 and At3GH116 isoforms, which were created by SWISS model based on the X-ray crystal structure of the *Tx*GH116 β-glucosidase (PDB: 5BX5). The flexible loop beside subsite +2 is colored red in *Tx*GH116, GBA2 and At4GH116, green in rice GH116, yellow in At3GH116.

4.3.2.7 Structure analysis of subsite +1 and +2 aromatic residue mutants with ligand complex
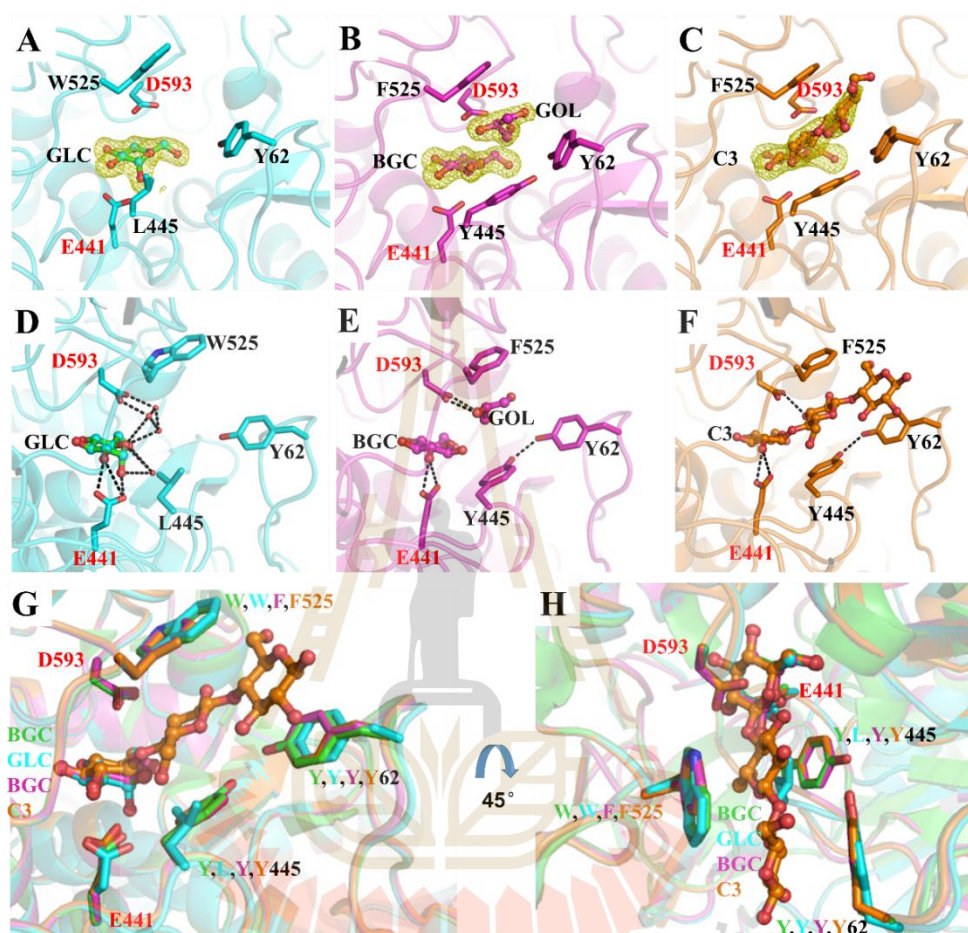
Figure 4.3.7 Structure comparison and intermolecular interactions analysis of subsite +1 and +2 residue mutants. A, B and C show the weighted Fo-Fc electron density omit maps of glucose (A, Y445L in cyan, B, W525F in purple) and cellobiose (C, W525F in brown) ligands in active sites in yellow mesh. D-F show the hydrogen bond interactions (dashed lines) of -1, +1 and +2 subsite ligand molecules with TxGH116 active site aromatic residues. Catalytic residues are labeled in red. G and H show the superimposed active sites of WT (green) and mutants.

To clarify the structural basis for the aromatic residue mutation effects, the structures of TxGH116 Y445L alone and in complex with glucose and W525F alone and in complexes with glucose and cellotriose were determined (Figure 4.3.7). In the complexes with glucose, Y445L was found to bind α-glucose, which may be because of an orientation change of the nucleophile residue E441 side chain (Figure 4.3.7 A), and W525F bound with β-glucose (Figure 4.3.7 B) in their -1subsites. Another difference was seen at subsite +1 site molecule, where Y445L had only waters (Figure 4.3.7 D), but W525F bound a glycerol molecule (Figure 4.3.7 E), which is similar to WT. This confirmed the importance of stacking interaction from Y445 for binding of ligands in subsite +1. In the W525F mutant, Y445 forms a hydrogen bond interaction with Y62 from flexible loop at the edge of the active site (Figure 4.3.7 E and F). This interaction plays an important role in maintaining the orientation of Y62 from the flexible loop to form correct orientation for the stacking interaction with the subsite +2 site glucose moiety (Pengthaisong et al., 2021). The Y445L mutant eliminated the stabilization interaction with Y62 which led to a slight shift of the flexible loop compared with WT and W525F (Figure 4.3.7 G and H).

Interestingly, when W525F soacked with cellobiose ligand, which potentially could be hydrolyzed into two glucose molecules, presented one cellotriose molecule in the active site with a low occupancy (Figure 4.3.7 C). It clearly shows the 1,4-glycoside bond link between the -1 subsite glucose and +1 subsite glucose residues, with the -1 site glucose in quite a different conformation compared to the WT glucose complex (Figure 4.3.7 G). Mutation from tryptophan to phenylalanine at residue 525 led to a slight orientation change. This slight change makes the side chain benzyl group more specific to the subsite +1 glucosyl residue (Glc2) alpha face stacking interaction formation (Figure 4.3.7 G). Glc2 from cellobiose ligand shows a

similar orientation as the glucosyl residue in subsite +1 in the nucleophile mutation structures (Pengthaisong et al., 2021), but the different rotation of C6OH group, which is in trans,gauche (tg) conformation (Quirke and Crich, 2020), whereas, the E441G mutant has glucosyl residues from cellotriose in subsites +1 to +2 in gauche,trans (gt) conformations (Pengthaisong et al., 2021). This indicates the sidechain conformation of Glc2 changed (Figure 4.3.7 F) with the mutation of the +1 to +2 subsite aromatic residue W525. Consideration of the side chain difference may be related with the glucose conformation change (Quirke and Crich, 2020), this means the new orientation of new phenylalanine residue determines the acceptor glucose molecule conformation during TxGH116 transglycosylation reactions. The similar mechanism was elucidated from cyclodextrin-glycosyltransferases, in which the hydrophobicity of Phe-183 and Phe-259 in subsite +1 and +2 plays a pivotal role in directing enzyme hydrolysis and transglycosylation activity (van der Veen et al., 2001). However, the W525F mutant presented comparable transglycosylation ability to WT, while the W525L mutant has a large transglycosylation product reduction (Figure 4.3.5). We can summarize that the W525 residue is responsible for the oligosaccharide chain stabilization during the TxGH116 transglycosylation process.

Glc3 (+2 subsite glucosyl residue) from the cellotriose produced when TxGH116 W445F was soaked with cellobiose is linked with Glc2 by a 1,4-glycoside bond and was observed with quite low occupancy (Figure 4.3.7 C). It was stabilized by the Y62 residue with a strong CH/pi interaction, since the aromatic ring of Y62 show perfect parallel orientation of its face to the Glc3 alpha face (Figure 4.3.7 H and Figure 4.3.8). This indicates the important role of Y62 located in the flexible loop in oligosaccharide chain binding, which is consistent with previous work (Pengthaisong et al., 2021)
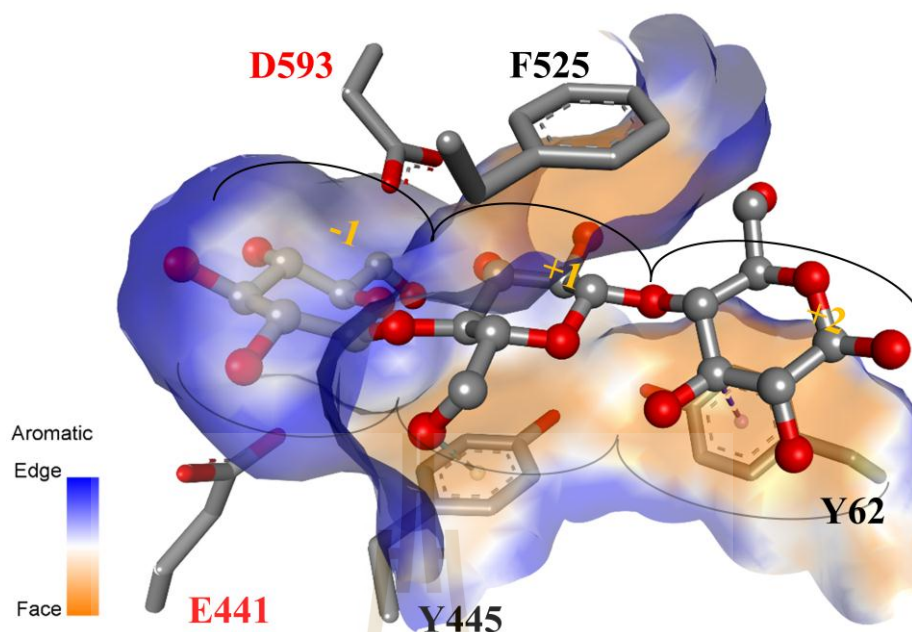
Figure 4.3.8 Structure model of W525F with cellotriose in the active site. Glc 1 binding in the -1 subsite, Glc2 interact with F525 and Y445 in subsite +1, and Glc3 interacts with Y62 in subsite +2. Yellow color show aromatic faces around glucose residues of the ligand. Catalytic amino acid residue are labeled in red. Active site subsites were indicated by black semicircular curve. The dark dashed lines represent CH/pi interaction in Glc3 and OH/pi interactions in Glc2.
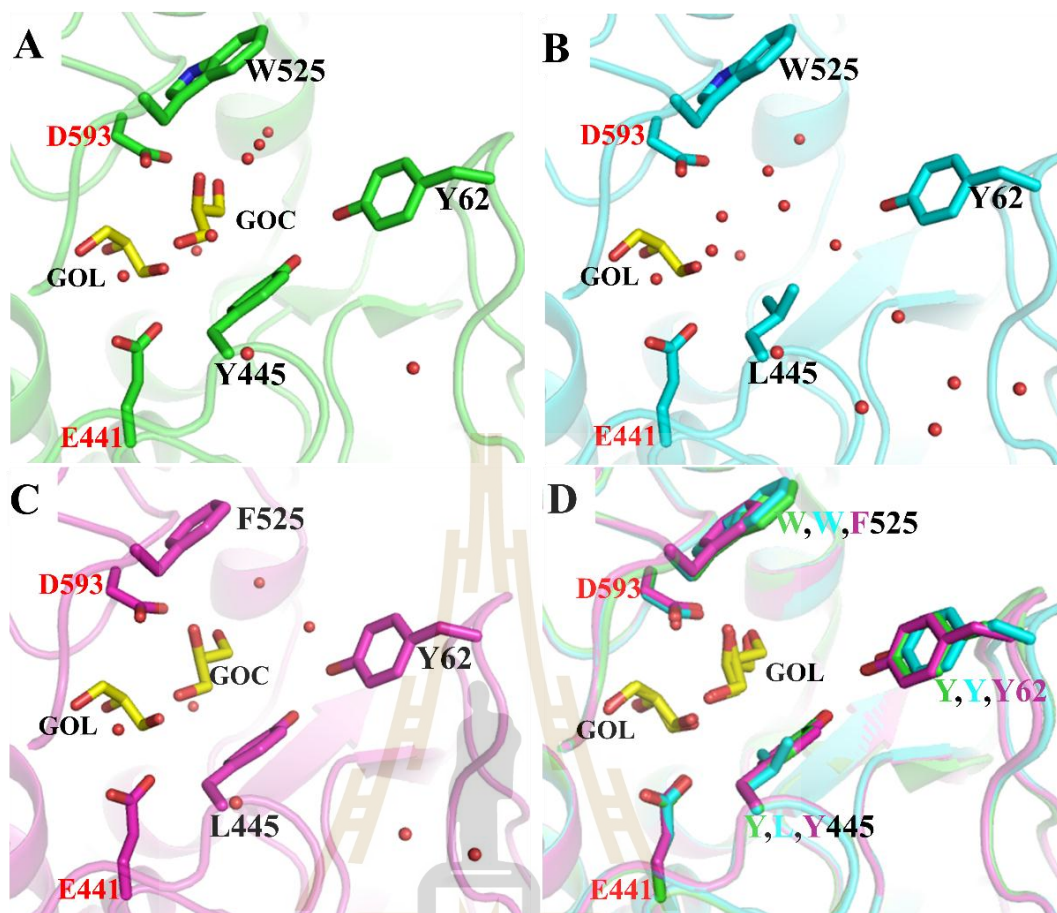
Figure 4.3.9 Structural comparison of WT TxGH116 and subsite +1 and +2 aromatic residue mutants. A:WT, B: Y445L, and C: W525F show the glycerol molecules (yellow sticks) in -1 and +1 site (WT in green, Y445L in cyan, W525F in purple). Water molecules are presented as red balls. Catalytic residues are labeled in red. D, Aromatic residues from subsites +1 and +2 and glycerol molecules in the superposition of the three structures.

Native crystal structures from WT and the Y445L and W525F mutants show an equivalent glycerol molecule in the -1 site (Figure 4.3.9). Whereas the WT and W525F mutant structures had another glycerol molecule in subsite +1, the Y445L

lacked that glycerol molecule in subsite +1 (Figure 4.3.9 A), which is apparently because of the elimination of aromatic ring from Y445. This indicated that the CH-pi interaction from Y445 is likely critical for subsite +1 carbohydrate molecule CH/pi interactions, since glycerol has abundant hydroxyl groups and polarized CH, as in carbohydrates. The W525F and WT structures have very similar distances from aromatic residues to glycerol molecules (Figure 4.3.9 B and C), which is consistent with the positions of the ligands. In the outer part of the active site cleft, Y62 and its loop are also shifted away from the active site in the Y445L mutation (Figure 4.3.7 G and H).

4.3.2.8 Autodocking analysis of pNPGlc substrate interactions with aromatic residue mutants
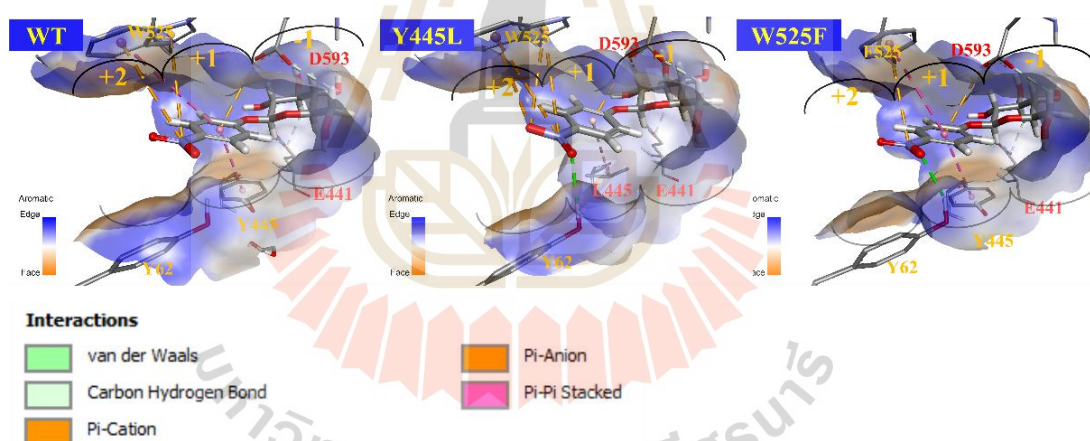
Figure 4.3.10 Autodock analysis of interactions between pNPGlc substrate ligand with TxGH116 WT and aromatic residue mutants. Yellow color show aromatic faces around the glucosyl moiety of the ligand. Catalytic residue was labeled in red. Active site subsites were indicated by black semicircular curves.

From autodocking of the pNPGlc molecule in the active site, we found the mostly interaction change in the Y445L subsite +1 is the elimination of pi-pi interaction between original tyrosine residue with nitrophenyl molecule from pNPGlc substrate (Figure 4.3.10). The original pi-pi interaction changed to CH-pi interaction between L445 isobutyl sidechain carbon and the nitrophenyl group, which is a weaker interaction. This is consistent with the kinetics analysis that showed TxGH116 Y445L hydrolyzed pNPGlc with very low activity (Table 4.3.1). The W525F mutant shows similar pi-pi and aromatic stacking interactions between the F525 residue accompanied by those with Y445 as WT, but the phenylalanine residue made a small rotation shift to the active site, relative to the indole sidechain of W525, that improved the pi-pi interaction between F525 and nitrophenyl group. This may also increase the aromatic platform stacking support in subsite +1 from both F525 and Y445. This is consistents with the kinetic higher specificity constant of the W525F mutant (Table 4.3.1). In addition, the stronger pi-pi and improved stacking interactions in W525F subsite +1 further increased its thermostability, based on its higher temperature optimum and specificity constant at high temperature (Table 4.3.1).

4.3.2.9 Autodocking analysis of cellobiose substrate interactions with aromatic residue mutants

From cellobiose molecule autodock results, we noticed that the original CH-pi interaction between the cellobiose subsite +1 glucose C4H with Y445 residue disappeared because of the Y445L mutation (Figure 4.3.11). In the Y445L active site, the cellobiose molecule docked in a similar position and configuration as in WT. This indicates that the glucose molecule in subsite +1 may be stabilized by the W525 residue. The autodock result from the W525F mutant furtherly

confirmed the improtant role of W525 residue in supporting subsite+1 glucose molecule stability, since the glucose in W525F subsite +1 had a higher energy configuration than in TxGH116 WT and Y445L mutant (Figure 4.3.11). This change may come from the less frequency of phenylalanine residue to form CH-pi aromatic stacking compare with tryptophan (Hudson et al., 2015). This also appeared in cellulase enzyme, in which aromatic residue in the active sites tunnel display only small differences in binding affinity, but the ligand flexibility and enzyme-ligand interactions are affected when an aromatic residue is mutated (Taylor et al., 2013). The natural substrate of TxGH116 is supposed to be oligosaccharides. In addition, sequence alignment results also got variety of amino acid residue types in the residue 525 position in different GH116 enzymes (Figure 4.3.2). GH116 from rice and Arabidopsis which might hydrolyse glycolipid or oligosaccharide substrates, do not have tryptophan in this position. Interestingly, they have one nonaromatic, hydrophobic amino acid like isoleucine or leucine residue, similar to human GBA2 enzyme. This may indicate that the substrate specificity of GH116 enzymes from plants and human mainly act on less polar substrate such as glucosylceramide (Dai et al., 2020; Jatooratthawichot et al., 2020). This is similar to GH116 enzyme from S. solfataricus, which is specific for the gluco- and xylosides β-bound to hydrophobic groups (Boot et al., 2007). These results suggest that nature employs aromatic-carbohydrate interactions with a wide range of binding affinities for diverse functions (Payne et al., 2011). Substrate specificity directly determined by binding site traits in GH enzymes (Aspeborg et al., 2012; Pollet et al., 2010). The specificity of TxGH116 in hydrolysing oligosaccharides indicates its potential role in industral applications.
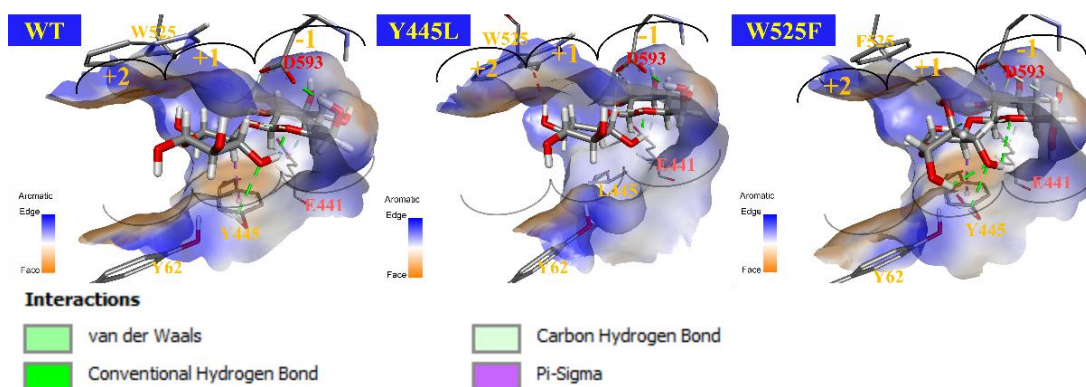
Figure 4.3.11 Autodock analysis of interactions between cellobiose substrate ligand with TxGH116 WT and aromatic residue mutants. Yellow color shows aromatic faces around glucosyl residues of the ligand. Catalytic residues are labeled in red. Active site subsites were indicated by black semicircular curve.

4.3.3 Glucose inhibition kinetics of subsite +1 and +2 aromatic residue mutant

Both W525 residue mutants displayed competitive inhibition by glucose with lower inhibition constants, Ki, than WT, which means they are more sensitive to glucose than WT, especially W525F (Figure 4.3.12). This is consistant with substrate hydrolysis kinetics, in which the W525F mutant had stronger glucoside molecule binding than WT. W525L mutant show strange glucose inhibition mechanism, because it seems like a little bit shift to mainly competitive mixed inhibition.
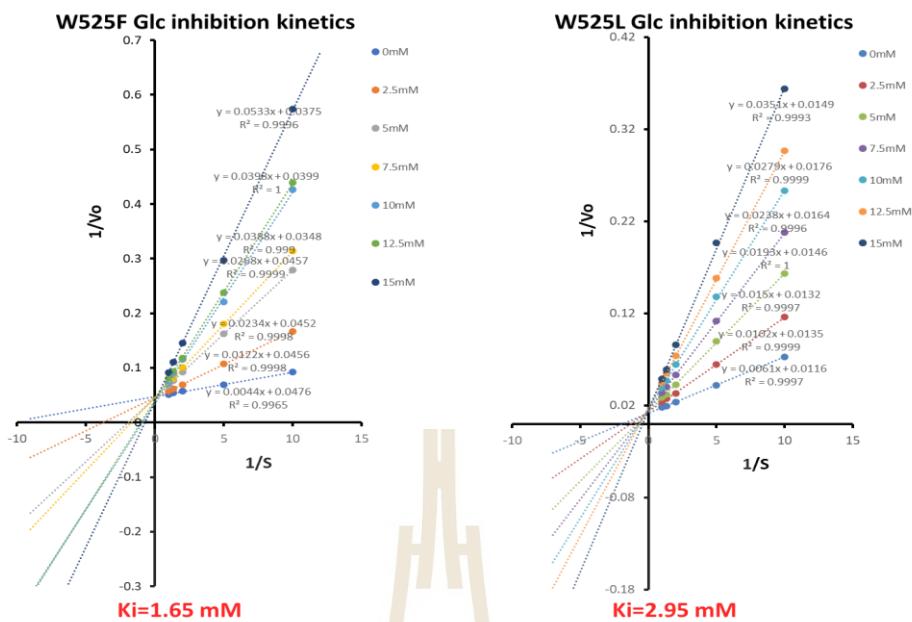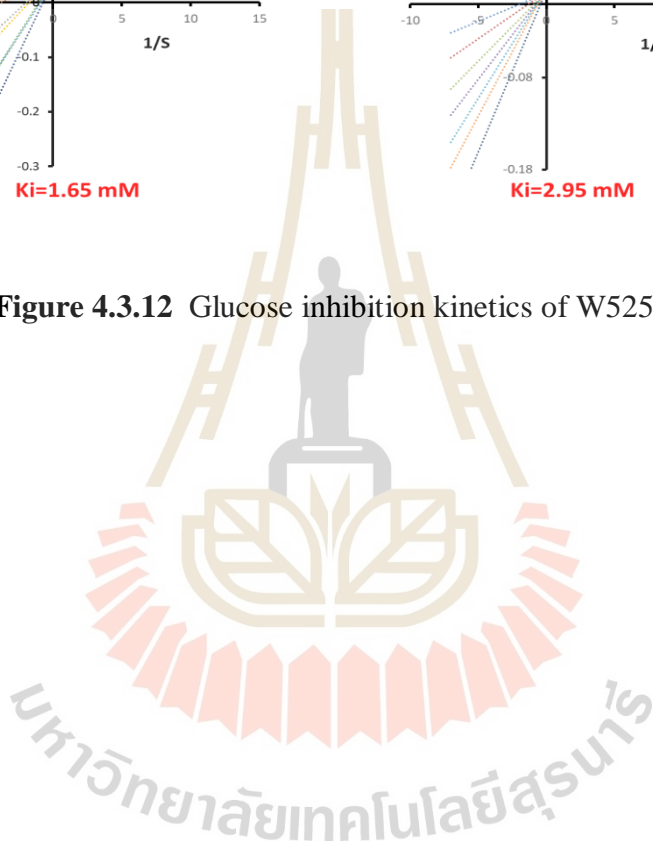
Figure 4.3.12 Glucose inhibition kinetics of W525 mutants.

Table 4.3.4 Data collection and refinement statistics of subsite +1 and +2 related

mutants. Values in parentheses are for the outer resolution shell.

Dataset	Y445L-native crystal	Y445L-glucose complex	W525F-native crystal	W525F-glucose complex	W525F-cellotriose complex
PDB code					
Beamline	BL13B1	BL15A1	BL13B1	TPS05A1	BL13B1
Wavelength (Å)	1.00	1.00	1.00	1.00	1.00
Data collection					
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions					
a, b, c (Å)	a = 80.098 b = 173.071 c = 83.597	a = 83.964 b = 173.219 c = 79.951	a = 173.389 b = 80.268 c = 121.942	a = 173.736 b = 80.284 c = 123.339	a = 80.589 b = 123.666 c = 173.930
α, β, γ (°)	90.00, 116.22, 90.00	90.00, 116.14, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution range (Å)	50.00 – 1.67	50.00 – 2.03	50.00 – 1.85	30.00 – 1.90	50.00 – 2.20
Resolution outer shell (Å)	1.73 – 1.67	2.10 – 2.03	1.92 – 1.85	1.97 – 1.90	2.28 – 2.20
No. Unique reflections	236519	130723	138812	135734	87473
No. Observed reflections	893547	505012	934847	884655	566916
Completeness (%)	99.9 (99.4)	99.3 (98.6)	95.7 (80.3)	99.8 (99.2)	99.9 (100.0)
Average redundancy per shell	3.8 (3.5)	3.9 (3.8)	6.8 (5.0)	6.5 (4.0)	6.5 (5.6)
I/σ (I)	19.52 (1.84)	12.10 (2.08)	20.19 (1.79)	14.41 (1.98)	11.17 (2.87)
$R_{(merge)}$ (%)	0.066 (0.619)	0.115 (0.658)	0.085 (0.794)	0.138 (0.642)	0.213 (0.740)
CC1/2	0.713	0.718	0.801	0.744	0.711
Refinement					
R_{factor} (%)	14.71	14.74	15.86	14.68	14.96
R_{free} (%)	16.92	18.85	20.02	18.94	19.90
No. of residues in proteins	769/771	769/769	770/769	773/773	771/771
No. Protein atoms	6289/6348	6221/6232	6260/6252	6265/6274	6252/6251
No. Ligand atoms	-	12/12	-	12/12	34/34
No. Other hetero atoms	140	56	82	88	50
No. Waters	1244	701	685	802	484
Mean B-factor					
Protein	16.743/17.661	30.104/30.799	22.964/26.816	21.421/24.958	31.880/35.816
Ligand	-	24.488/25.192	-	17.412/19.647	59.474/67.090
Other hetero atoms	33.475	45.794	40.654	38.528	54.337
Waters	31.612	37.262	30.050	31.310	35.713
r.m.s. bond deviations (length)	0.0093	0.0141	0.0132	0.0128	0.0069
r.m.s angle deviations (degrees)	1.5502	1.8831	1.8199	1.7586	1.4620
Ramachandran plot					
Ramachandran favored (%)	96.21	95.74	95.88	95.65	95.57
Ramachandran outliers (%)	0.00	0.00	0.26	0.19	0.07

CHAPTER V

CONCLUSION

5.1 Glycone sugar binding related mutations conclusion

We conducted the systematic functional analysis of TxGH116 active site glycone sugar binding residues: D452, H507, T591, E730, W732, E777, R786 and R792 in substrate binding and catalysis by analysis of their mutations. From enzyme substrate kinetics and structural analysis, we clarified the roles of each amino acid residue. We observed that residue R786 forms hydrogen bonds to the C6OH of glucose. The R786A mutation eliminates these H-bonds and gave higher K_M and k_{cat} values for pNPGlc hydrolysis than WT. The TxGH116 R786K mutant displays a similar K_M value as R786A with pNPGlc substrate. Glucose competitively inhibited R786A and R786K with K_i values which are similar to the WT. The X-ray crystal structure of R786K has a similar active-site structure as WT, but the structure of R786A has a wider entrance to the active site, which changes interactions between the nitrophenyl moiety and subsites +1 and +2. These changes may facilitate product molecule release from active site. Another arginine residue, R792, forms hydrogen bonds to the glucose C4OH, as do E777 and T591, while D452 appears to form hydrogen bonds to the glucose C3OH and C4OH, and H507 forms a hydrogen bond to the C3OH. W732 forms CH-pi interactions with C6 and a hydrogen bond to E730, which forms 2 hydrogen bonds to the R786 guanidinyl group. Mutations of these residues generally decreased catalytic efficiency. This work begins to tease out the

relative importance of GH116 active site residues in substrate and inhibitor binding, which may allow the modification of *Tx*GH116 β-glucosidase for improved application.

5.2   Acid/base and nucleophile related residues conclusion

With site directed mutation of residues adjacent to the *Tx*GH116 catalytic acid/base and nucleophile residues, we identified the importance of these residues. The Y523, D593, Y790, T450 and E441 residues are highly conserved in the GH116 family, and the Q727 residue is less conserved as it is substituted in some species by histidine and tyrosine, which may play similar roles as hydrogen bond donors. The two triad systems play important roles in hydrogen bond interactions to tune the activity of the catalytic residues. The T450, Q727 and Y790 mostly act as proton donors, while Y523 may be mainly a proton acceptor from the acid/base residue, until the proton is transferred to the glycosidic bond oxygen. So, in the "Tyr-acid/base-Gln" triad, the hydrogen bond proton chain is supposed from Q727 to D593 then Y523. The mutations T450A and Y790F cause large scale hydrolysis activity reduction (Table 4.2.1), because of the imbalance of the E441 residue without either of these two residues, which leads to deglycosylation steps occuring at slow rates. The Y523F mutation has a high optimum pH and 5-fold hydrolysis activity reduction, indicating its main role in proton balance to maintain the D593 acid/base residue in the appropariate protonated or unprotonated state. The effect on k_{cat} suggests it may act in the deglycosylation reaction, which is likely to be rate-limiting or partially rate-limiting with pNPGlc. Interestingly, *Tx*GH116 Q727A and Q727E have relatively high hydrolysis activity (Table 4.2.1) and even higher transglycosylation levels than

WT (Figure 4.2.7 and 4.2.8). Structural analysis indicated a glucose ligand conformation change in the TxGH116 T450A active site and binding change in TxGH116 Q727A and Q727E subsite +1 site, which lead to hydrolysis and transglycosylation activity changes, respectively (Figure 4.2.10). Y523 and D593 both reside in a long loop (Figure 4.2.1), which allows their proximity and interaction. This work presented the effects of changing the residues that interact with the catalytic residues, which may be useful in engineering TxGH116 for more applications by balancing glycosylation and deglycosylation.

5.3 Subsite+1 and +2 aromatic related residue mutation conclusion

We elucidated the roles of aromatic residues in the *Tx*GH116 slot like active site subsites +1 and +2. Enzyme kinetics and structural analysis showed that the Y445 residue plays a pivotal role in transition state interaction and stacking support for subsite +1 substrate molecule binding based on its special location that is parallel with the glucose moiety $^4C_1$ chair conformation. Elimination of tyrosine aromatic platform results in a large reduction in activity (Table 4.3.1 and 4.3.2). The lack of transglycosylation products in the Y445L mutant (Figure 4.3.5) indicated the low activity and possibly an important tyrosine role in acceptor molecule binding in the glucosyl transferase process. The W525 residue in subsites +1 and +2 not only donates stacking support in substrate binding but also affets product release and substrate selectivity. The W525F mutant has improved substrate binding, while W525L appparently increased the rate of product release. The W525F effect came from the aromatic ring rotation change after mutation. The increased hydrolysis activity from the W525L mutant indicated the pivotal role of W525 in determine the

ratio of hydrolysis to transglycosylation in *Tx*GH116. W525 residue mutants show more sensitive to glucose inhibition than WT that may because of the glucose molecule block binding at the +1 subsite. This is supported by the structure from the crystal of the W525F soaked with cellobiose, in which a cellotriose liagnd was observed in the active site (Figure 4.3.7). Our effort to get the crystal structure of the W525L mutant resulted in a low resolution dataset and structure (data not show), in which the electron density was observed for the -1 site glucose only without any ligand in subsites +1 and +2.

The W732F mutant show very less reduction in hydrolyse activity, so its aromatic interaction to substrate molecule in -1 site not that significant as aromatic residues W525 and Y445. The main reduction from W732F mutation maybe come from the hydrogen bond elimination with its neighbor E730 which further affects the R786 residue that interacts with the C6OH group (glycone sugar binding).

The fact that the Y445 residue show high conservation in GH116 family enzymes from different species provides further evidence for the essential role of Y445 in *Tx*GH116 enzyme catalytic function performance. The lack of conservation of W525 in multiple species supports its specific role in substrate recognization that is function dependent. Despite their relatively modest influence on the stability of carbohydrate/protein complexes, the aromatic platforms play a major role in determining the specificity of the molecular recognition process. This can be an efficient supplement for designment of the emerging medical drug design application that related with protein–carbohydrate recognition.

Overall, the functional and structural studies provided in this thesis have greatly improved our understanding of the significance of the amino acid residue

around the active site in the GH116 family and in *Tx*GH116 specifically. This information will aid in the engineering of *Tx*GH116 and other GH116 enzymes for specific applications and in design of inhibitors for GH116 enzymes.

REFERENCES

Andrew Karplus, P. (1997). Hydrophobicity regained. **Protein Science.** 6(6): 1302-1307.

Andrews, S. R., Charnock, S. J., Lakey, J. H., Davies, G. J., Claeyssens, M., Nerinckx, W., Underwood, M., Sinnott, M. L., Warren, R. A., and Gilbert, H. J. (2000). Substrate specificity in glycoside hydrolase family 10. Tyrosine 87 and leucine 314 play a pivotal role in discriminating between glucose and xylose binding in the proximal active site of Pseudomonas cellulosa xylanase 10A. **Journal of Biological Chemistry.** 275(30): 23027-23033.

Ardèvol, A., and Rovira, C. (2015). Reaction mechanisms in carbohydrate-active enzymes: glycoside hydrolases and glycosyltransferases. Insights from ab initio quantum mechanics/molecular mechanics dynamic simulations. **Journal of the American Chemical Society.** 137(24): 7528-7547.

Artola, M., Wu, L., Ferraz, M. J., Kuo, C. L., Raich, L., Breen, I. Z., Offen, W. A., Codee, J. D. C., van der Marel, G. A., Rovira, C., Aerts, J., Davies, G. J., and Overkleeft, H. S. (2017). 1,6-Cyclophellitol Cyclosulfates: A New Class of Irreversible Glycosidase Inhibitor. **ACS Central Science.** 3(7): 784-793.

Aspeborg, H., Coutinho, P. M., Wang, Y., Brumer, H., and Henrissat, B. (2012). Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). **BMC Evolutionary Biology.** 12(1): 1-16.

Baker, E. G., Williams, C., Hudson, K. L., Bartlett, G. J., Heal, J. W., Porter Goff, K. L., Sessions, R. B., Crump, M. P., and Woolfson, D. N. (2017). Engineering

protein stability with atomic precision in a monomeric miniprotein. **Nature Chemical Biology.** 13(7): 764-770.

Barrett, T., Suresh, C. G., Tolley, S. P., Dodson, E. J., and Hughes, M. A. (1995). The crystal structure of a cyanogenic beta-glucosidase from white clover, a family 1 glycosyl hydrolase. **Structure.** 3(9): 951-960.

Bender, M. L. (1960). Mechanisms of catalysis of nucleophilic reactions of carboxylic acid derivatives. **Chemical Reviews.** 60(1): 53-113.

Bertozzi, C. R., and Kiessling, L. L. (2001). Chemical glycobiology. **Science.** 291(5512): 2357-2364.

Bischoff, R., and Schluter, H. (2012). Amino acids: chemistry, functionality and selected non-enzymatic post-translational modifications. **Journal of Proteomics.** 75(8): 2275-2296.

Blow, D. M., Birktoft, J. J., and Hartley, B. (1969). Role of a buried acid group in the mechanism of action of chymotrypsin. **Nature.** 221(5178): 337-340.

Boot, R. G., Verhoek, M., Donker-Koopman, W., Strijland, A., van Marle, J., Overkleeft, H. S., Wennekes, T., and Aerts, J. M. (2007). Identification of the non-lysosomal glucosylceramidase as beta-glucosidase 2. **Journal of Biological Chemistry.** 282(2): 1305-1312.

Borders Jr, C., Broadwater, J. A., Bekeny, P. A., Salmon, J. E., Lee, A. S., Eldridge, A. M., and Pett, V. B. (1994). A structural role for arginine in proteins: multiple hydrogen bonds to backbone carbonyl oxygens. **Protein Science.** 3(4): 541-548.

Braman, J., Papworth, C., and Greener, A. (1996). Site-directed mutagenesis using double-stranded plasmid DNA templates. **Methods in Molecular Biology.** 57: 31-44.

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

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. **Nucleic Acids Research.** 37(Database issue): D233-238.

Cargile, B. J., Sevinsky, J. R., Essader, A. S., Eu, J. P., and Stephenson Jr, J. L. (2008). Calculation of the isoelectric point of tryptic peptides in the pH 3.5-4.5 range based on adjacent amino acid effects. **Electrophoresis.** 29(13): 2768-2778.

Charoenwattanasatien, R., Pengthaisong, S., Breen, I., Mutoh, R., Sansenya, S., Hua, Y., Tankrathok, A., Wu, L., Songsiriritthigul, C., Tanaka, H., Williams, S. J., Davies, G. J., Kurisu, G., and Cairns, J. R. (2016). Bacterial beta-Glucosidase Reveals the Structural and Functional Basis of Genetic Defects in Human Glucocerebrosidase 2 (GBA2). **ACS Chemical Biology.** 11(7): 1891-1900.

Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. **Acta Crystallographica. Section D, Biological Crystallography.** 66(Pt 1): 12-21.

Chen, X. (2017). Chemoenzymatic synthesis of carbohydrates and glycoconjugates. **Abstracts of Papers of the American Chemical Society.** 253.

Chuenchor, W., Pengthaisong, S., Robinson, R. C., Yuvaniyama, J., Svasti, J., and Cairns, J. R. (2011). The structural basis of oligosaccharide binding by rice BGlu1 beta-glucosidase. **Journal of Structural Biology.** 173(1): 169-179.

Cobucci-Ponzano, B., Aurilia, V., Riccio, G., Henrissat, B., Coutinho, P. M., Strazzulli, A., Padula, A., Corsaro, M. M., Pieretti, G., Pocsfalvi, G., Fiume, I., Cannio, R., Rossi, M., and Moracci, M. (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. **Journal of Biological Chemistry.** 285(27): 20691-20703.

Crespim, E., Zanphorlin, L. M., de Souza, F. H., Diogo, J. A., Gazolla, A. C., Machado, C. B., Figueiredo, F., Sousa, A. S., Nobrega, F., Pellizari, V. H., Murakami, M. T., and Ruller, R. (2016). A novel cold-adapted and glucose-tolerant GH1 beta-glucosidase from Exiguobacterium antarcticum B7. **International Journal of Biological Macromolecules.** 82: 375-380.

Crich, D., and Li, L. (2007). 4,6-O-benzylidene-directed beta-mannopyranosylation and alpha-glucopyranosylation: the 2-deoxy-2-fluoro and 3-deoxy-3-fluoro series of donors and the importance of the O2-C2-C3-O3 interaction. **Journal of Organic Chemistry.** 72(5): 1681-1690.

Dai, G. Y., Yin, J., Li, K. E., Chen, D. K., Liu, Z., Bi, F. C., Rong, C., and Yao, N. (2020). The Arabidopsis AtGCD3 protein is a glucosylceramidase that preferentially hydrolyzes long-acyl-chain glucosylceramides. **Journal of Biological Chemistry.** 295(3): 717-728.

Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. **Structure.** 3(9): 853-859.

Davies, G. J., Dauter, M., Brzozowski, A. M., Bjornvad, M. E., Andersen, K. V., and Schulein, M. (1998). Structure of the Bacillus agaradherans family 5 endoglucanase at 1.6 A and its cellobiose complex at 2.0 A resolution. **Biochemistry.** 37(7): 1926-1932.

Day, A. J., Cañada, F. J., Díaz, J. C., Kroon, P. A., Mclauchlan, R., Faulds, C. B., Plumb, G. W., Morgan, M. R., and Williamson, G. (2000). Dietary flavonoid and isoflavone glycosides are hydrolysed by the lactase site of lactase phlorizin hydrolase. **FEBS Letters.** 468(2-3): 166-170.

De Giuseppe, P. O., Souza Tde, A., Souza, F. H., Zanphorlin, L. M., Machado, C. B., Ward, R. J., Jorge, J. A., Furriel Rdos, P., and Murakami, M. T. (2014). Structural basis for glucose tolerance in GH1 beta-glucosidases. **Acta Crystallographica. Section D, Biological Crystallography.** 70(Pt 6): 1631-1639.

Decker, C. H., Visser, J., and Schreier, P. (2000). beta-glucosidases from five black Aspergillus species: study of their physico-chemical and biocatalytic properties. **Journal of Agricultural and Food Chemistry.** 48(10): 4929-4936.

Del Carmen Fernandez-Alonso, M., Canada, F. J., Jimenez-Barbero, J., and Cuevas, G. (2005). Molecular recognition of saccharides by proteins. Insights on the origin of the carbohydrate-aromatic interactions. **Journal of the American Chemical Society.** 127(20): 7379-7386.

Desjardins, M., Mak, W. S., O'Brien, T. E., Carlin, D. A., Tantillo, D. J., and Siegel, J. B. (2017). Systematic Functional Analysis of Active-Site Residues in l-

Threonine Dehydrogenase from Thermoplasma volcanium. **ACS Omega.** 2(7): 3308-3314.

Dixon, M. (1953). The determination of enzyme inhibitor constants. **Biochemical Journal.** 55(1): 170-171.

Dodson, G., and Wlodawer, A. (1998). Catalytic triads and their relatives. **Trends in Biochemical Sciences.** 23(9): 347-352.

Dopitova, R., Mazura, P., Janda, L., Chaloupkova, R., Jerabek, P., Damborsky, J., Filipi, T., Kiran, N. S., and Brzobohaty, B. (2008). Functional analysis of the aglycone-binding site of the maize beta-glucosidase Zm-p60.1. **FEBS Journal.** 275(24): 6123-6135.

Dvir, H., Harel, M., McCarthy, A. A., Toker, L., Silman, I., Futerman, A. H., and Sussman, J. L. (2003). X-ray structure of human acid-beta-glucosidase, the defective enzyme in Gaucher disease. **EMBO Reports.** 4(7): 704-709.

Dwek, R. A. (1996). Glycobiology: Toward Understanding the Function of Sugars. **Chemical Reviews.** 96(2): 683-720.

Edelheit, O., Hanukoglu, A., and Hanukoglu, I. (2009). Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. **BMC Biotechnology.** 9.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. **Acta Crystallographica. Section D, Biological Crystallography.** 60(Pt 12 Pt 1): 2126-2132.

Ferrara, M. C., Cobucci-Ponzano, B., Carpentieri, A., Henrissat, B., Rossi, M., Amoresano, A., and Moracci, M. (2014). The identification and molecular characterization of the first archaeal bifunctional exo-beta-glucosidase/N-

acetyl-beta-glucosaminidase demonstrate that family GH116 is made of three functionally distinct subfamilies. **Biochimica et Biophysica Acta (BBA) - Bioenergetics.** 1840(1): 367-377.

Finn, R. D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S. R., Sonnhammer, E. L., and Bateman, A. (2006). Pfam: clans, web tools and services. **Nucleic Acids Research.** 34(Database issue): D247-251.

Frutuoso, M. A., and Marana, S. R. (2013). A single amino acid residue determines the ratio of hydrolysis to transglycosylation catalyzed by beta-glucosidases. **Protein & Peptide Letters.** 20(1): 102-106.

Gallivan, J. P., and Dougherty, D. A. (1999). Cation-pi interactions in structural biology. **Proceedings of the National Academy of Sciences.** 96(17): 9459-9464.

Gebler, J., Gilkes, N. R., Claeyssens, M., Wilson, D. B., Beguin, P., Wakarchuk, W. W., Kilburn, D. G., Miller, R. C., Jr., Warren, R. A., and Withers, S. G. (1992). Stereoselective hydrolysis catalyzed by related beta-1,4-glucanases and beta-1,4-xylanases. **Journal of Biological Chemistry.** 267(18): 12559-12561.

Geronimo, I., Payne, C. M., and Sandgren, M. (2018). The role of catalytic residue pKa on the hydrolysis/transglycosylation partition in family 3 beta-glucosidases. **Organic and Biomolecular Chemistry.** 16(2): 316-324.

Gorantla, J. N., Pengthaisong, S., Choknud, S., Kaewpuang, T., Manyum, T., Promarak, V., and Ketudat Cairns, J. R. (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(11): 6211-6220.

Guo, B., Amano, Y., and Nozaki, K. (2016). Improvements in Glucose Sensitivity and Stability of *Trichoderma reesei* beta-Glucosidase Using Site-Directed Mutagenesis. **PLoS One.** 11(1): e0147301.

Hamre, A. G., Jana, S., Reppert, N. K., Payne, C. M., and Sorlie, M. (2015). Processivity, Substrate Positioning, and Binding: The Role of Polar Residues in a Family 18 Glycoside Hydrolase. **Biochemistry.** 54(49): 7292-7306.

Hass, M. A., and Mulder, F. A. (2015). Contemporary NMR Studies of Protein Electrostatics. **Annual Review of Biophysics.** 44: 53-75.

Hefferon, K. L., Cantero-Tubilla, B., Brady, J., and Wilson, D. (2019). Aromatic residues surrounding the active site tunnel of TfCel48A influence activity, processivity, and synergistic interactions with other cellulases. **Biotechnology and Bioengineering.** 116(10): 2463-2472.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. **Biochemical Journal.** 280 (Pt 2): 309-316.

Holliday, G. L., Bartlett, G. J., Almonacid, D. E., O'Boyle, N. M., Murray-Rust, P., Thornton, J. M., and Mitchell, J. B. (2005). MACiE: a database of enzyme reaction mechanisms. **Bioinformatics.** 21(23): 4315-4316.

Holliday, G. L., Mitchell, J. B., and Thornton, J. M. (2009). Understanding the functional roles of amino acid residues in enzyme catalysis. **Journal of Molecular Biology.** 390(3): 560-577.

Houser, J., Kozmon, S., Mishra, D., Hammerova, Z., Wimmerova, M., and Koca, J. (2020). The CH-pi Interaction in Protein-Carbohydrate Binding: Bioinformatics and In Vitro Quantification. **Chemistry.** 26(47): 10769-10780.

Hsu, C. H., Park, S., Mortenson, D. E., Foley, B. L., Wang, X., Woods, R. J., Case, D. A., Powers, E. T., Wong, C. H., Dyson, H. J., and Kelly, J. W. (2016). The Dependence of Carbohydrate-Aromatic Interaction Strengths on the Structure of the Carbohydrate. **Journal of the American Chemical Society.** 138(24): 7636-7648.

Hudson, K. L., Bartlett, G. J., Diehl, R. C., Agirre, J., Gallagher, T., Kiessling, L. L., and Woolfson, D. N. (2015). Carbohydrate-Aromatic Interactions in Proteins. **Journal of the American Chemical Society.** 137(48): 15152-15160.

Iglesias-Fernandez, J., Hancock, S. M., Lee, S. S., Khan, M., Kirkpatrick, J., Oldham, N. J., McAuley, K., Fordham-Skelton, A., Rovira, C., and Davis, B. G. (2017). A front-face 'SNi synthase' engineered from a retaining 'double-SN2' hydrolase. **Nature Chemical Biology.** 13(8): 874-881.

Ishii, T. M., Zerr, P., Xia, X. M., Bond, C. T., Maylie, J., and Adelman, J. P. (1998). Site-directed mutagenesis. **Methods in Enzymology.** 293: 53-71.

Jatooratthawichot, P., Talabnin, C., Ngiwsara, L., Rustam, Y. H., Svasti, J., Reid, G. E., and Ketudat Cairns, J. R. (2020). Effect of Expression of Human Glucosylceramidase 2 Isoforms on Lipid Profiles in COS-7 Cells. **Metabolites.** 10(12).

Jencks, W. P. (1975). Binding energy, specificity, and enzymic catalysis: the circe effect. **Advances in Enzymology and Related Areas of Molecular Biology.** 43: 219-410.

Juan, L. A., Ana, A., Francisco, J. C., and Jesús, J.-B. (2013). Carbohydrate Aromatic Interactions. **Accounts of Chemical Research.** Vol. 46(No. 4): 946–954.

Kallemeijn, W. W., Witte, M. D., Voorn-Brouwer, T. M., Walvoort, M. T., Li, K. Y., Codee, J. D., van der Marel, G. A., Boot, R. G., Overkleeft, H. S., and Aerts, J. M. (2014). A sensitive gel-based method combining distinct cyclophellitol-based probes for the identification of acid/base residues in human retaining beta-glucosidases. **Journal of Biological Chemistry.** 289(51): 35351-35362.

Kamiya, Y., Yagi-Utsumi, M., Yagi, H., and Kato, K. (2011). Structural and molecular basis of carbohydrate-protein interaction systems as potential therapeutic targets. **Current Pharmaceutical Design.** 17(17): 1672-1684.

Kancharla, P. K., and Crich, D. (2013). Influence of side chain conformation and configuration on glycosyl donor reactivity and selectivity as illustrated by sialic acid donors epimeric at the 7-position. **Journal of the American Chemical Society.** 135(50): 18999-19007.

Ketudat Cairns, J. R., and Esen, A. (2010). beta-Glucosidases. **Cellular and Molecular Life Sciences.** 67(20): 3389-3405.

Ketudat Cairns, J. R., Mahong, B., Baiya, S., and Jeon, J. S. (2015). beta-Glucosidases: Multitasking, moonlighting or simply misunderstood? **Plant Science.** 241: 246-259.

Korotkova, O. G., Semenova, M. V., Morozova, V. V., Zorov, I. N., Sokolova, L. M., Bubnova, T. M., Okunev, O. N., and Sinitsyn, A. P. (2009). Isolation and properties of fungal beta-glucosidases. **Biochemistry (Mosc).** 74(5): 569-577.

Körschen, H. G., Yildiz, Y., Raju, D. N., Schonauer, S., Bönigk, W., Jansen, V., Kremmer, E., Kaupp, U. B., and Wachten, D. (2013). The non-lysosomal β-glucosidase GBA2 is a non-integral membrane-associated protein at the

endoplasmic reticulum (ER) and Golgi. **Journal of Biological Chemistry.** 288(5): 3381-3393.

Kuchner, O., and Arnold, F. H. (1997). Directed evolution of enzyme catalysts. **Trends in Biotechnology.** 15(12): 523-530.

Kuntothom, T., and Cairns, J. K. (2020). Expression and characterization of TbCel12A, a thermophilic endoglucanase with potential in biomass hydrolysis. **Biocatalysis and Agricultural Biotechnology.** 30: 101835.

Kurakata, Y., Uechi, A., Yoshida, H., Kamitori, S., Sakano, Y., Nishikawa, A., and Tonozuka, T. (2008). Structural insights into the substrate specificity and function of Escherichia coli K12 YgjK, a glucosidase belonging to the glycoside hydrolase family 63. **Journal of Molecular Biology.** 381(1): 116-128.

Kurasin, M., and Valjamae, P. (2011). Processivity of cellobiohydrolases is limited by the substrate. **Journal of Biological Chemistry.** 286(1): 169-177.

Kuusk, S., and Valjamae, P. (2017). When substrate inhibits and inhibitor activates: implications of beta-glucosidases. **Biotechnology for Biofuels.** 10: 7.

Laskowski, R. A., Moss, D. S., and Thornton, J. M. (1993). Main-chain bond lengths and bond angles in protein structures. **Journal of Molecular Biology.** 231(4): 1049-1067.

Laughrey, Z. R., Kiehna, S. E., Riemen, A. J., and Waters, M. L. (2008). Carbohydrate-pi interactions: what are they worth? **Journal of the American Chemical Society.** 130(44): 14625-14633.

Liu, X., Cao, L., Zeng, J., Liu, Y., and Xie, W. (2019). Improving the cellobiose-hydrolysis activity and glucose-tolerance of a thermostable beta-glucosidase

through rational design. **International Journal of Biological Macromolecules.** 136: 1052-1059.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. **Nucleic Acids Research.** 42(Database issue): D490-495.

Lucas, J. E., and Siegel, J. B. (2015). Quantitative functional characterization of conserved molecular interactions in the active site of mannitol 2-dehydrogenase. **Protein Science.** 24(6): 936-945.

Lundemo, P., Karlsson, E. N., and Adlercreutz, P. (2017). Eliminating hydrolytic activity without affecting the transglycosylation of a GH1 beta-glucosidase. **Applied Microbiology and Biotechnology.** 101(3): 1121-1131.

Madhavan, A., Sindhu, R., Binod, P., Sukumaran, R. K., and Pandey, A. (2017). Strategies for design of improved biocatalysts for industrial applications. **Bioresource Technology.** 245(Pt B): 1304-1313.

Madison, E. L., Kobe, A., Gething, M.-J., Sambrook, J. F., and Goldsmith, E. J. (1993). Converting tissue plasminogen activator to a zymogen: a regulatory triad of Asp-His-Ser. **Science.** 262(5132): 419-421.

Matsuzawa, T., Jo, T., Uchiyama, T., Manninen, J. A., Arakawa, T., Miyazaki, K., Fushinobu, S., and Yaoi, K. (2016). Crystal structure and identification of a key amino acid for glucose tolerance, substrate specificity, and transglycosylation activity of metagenomic beta-glucosidase Td2F2. **FEBS Journal.** 283(12): 2340-2353.

McCarter, J. D., and Withers, S. G. (1994). Mechanisms of enzymatic glycoside hydrolysis. **Current Opinion in Structural Biology.** 4(6): 885-892.

McCarthy, J. K., Uzelac, A., Davis, D. F., and Eveleigh, D. E. (2004). Improved catalytic efficiency and active site modification of 1,4-beta-D-glucan glucohydrolase A from Thermotoga neapolitana by directed evolution. **Journal of Biological Chemistry.** 279(12): 11495-11502.

McIlvaine, T. (1921). A buffer solution for colorimetric comparison. **Journal of Biological Chemistry.** 49(1): 183-186.

Mistry, P. K., Liu, J., Sun, L., Chuang, W.-L., Yuen, T., Yang, R., Lu, P., Zhang, K., Li, J., and Keutzer, J. (2014). Glucocerebrosidase 2 gene deletion rescues type 1 Gaucher disease. **Proceedings of the National Academy of Sciences.** 111(13): 4934-4939.

Montalvillo-Jimenez, L., Santana, A. G., Corzana, F., Jimenez-Oses, G., Jimenez-Barbero, J., Gomez, A. M., and Asensio, J. L. (2019). Impact of Aromatic Stacking on Glycoside Reactivity: Balancing CH/pi and Cation/pi Interactions for the Stabilization of Glycosyl-Oxocarbenium Ions. **Journal of the American Chemical Society.** 141(34): 13372-13384.

Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. **Acta Crystallographica. Section D, Biological Crystallography.** 67(Pt 4): 355-367.

Naim, H. Y. (2001). Molecular and cellular aspects and regulation of intestinal lactase-phlorizin hydrolase. **Histology and Histopathology.** 16(2): 553-561.

Nakazawa, H., Kawai, T., Ida, N., Shida, Y., Kobayashi, Y., Okada, H., Tani, S., Sumitani, J., Kawaguchi, T., Morikawa, Y., and Ogasawara, W. (2012). Construction of a recombinant Trichoderma reesei strain expressing

Aspergillus aculeatus beta-glucosidase 1 for efficient biomass conversion. **Biotechnology and Bioengineering.** 109(1): 92-99.

Namchuk, M. N., and Withers, S. G. (1995). Mechanism of Agrobacterium beta-glucosidase: kinetic analysis of the role of noncovalent enzyme/substrate interactions. **Biochemistry.** 34(49): 16194-16202.

Nerinckx, W., Desmet, T., and Claeyssens, M. (2003). A hydrophobic platform as a mechanistically relevant transition state stabilising factor appears to be present in the active centre of all glycoside hydrolases. **FEBS Letters.** 538(1-3): 1-7.

Newberry, R. W., and Raines, R. T. (2019). Secondary Forces in Protein Folding. **ACS Chemical Biology.** 14(8): 1677-1686.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. **Methods in Enzymology.** 276: 307-326.

Ouairy, C., Cresteil, T., Delpech, B., and Crich, D. (2013). Synthesis and evaluation of 3-deoxy and 3-deoxy-3-fluoro derivatives of gluco- and manno-configured tetrahydropyridoimidazole glycosidase inhibitors. **Carbohydrate Research.** 377: 35-43.

Palackal, N., Brennan, Y., Callen, W. N., Dupree, P., Frey, G., Goubet, F., Hazlewood, G. P., Healey, S., Kang, Y. E., Kretz, K. A., Lee, E., Tan, X., Tomlinson, G. L., Verruto, J., Wong, V. W., Mathur, E. J., Short, J. M., Robertson, D. E., and Steer, B. A. (2004). An evolutionary route to xylanase process fitness. **Protein Science.** 13(2): 494-503.

Park, J.-S., Hitomi, J., Horinouchi, S., and Beppu, T. (1993). Identification of two amino acids contributing the high enzyme activity in the alkaline pH range of

an alkaline endoglucanase from a Bacillus sp. **Protein Engineering, Design and Selection.** 6(8): 921-926.

Payne, C. M., Bomble, Y. J., Taylor, C. B., McCabe, C., Himmel, M. E., Crowley, M. F., and Beckham, G. T. (2011). Multiple functions of aromatic-carbohydrate interactions in a processive cellulase examined with molecular simulation. **Journal of Biological Chemistry.** 286(47): 41028-41035.

Pengthaisong, S., Chen, C. F., Withers, S. G., Kuaprasert, B., and Ketudat Cairns, J. R. (2012). Rice BGlu1 glycosynthase and wild type transglycosylation activities distinguished by cyclophellitol inhibition. **Carbohydrate Research.** 352: 51-59.

Pengthaisong, S., Hua, Y., and Ketudat Cairns, J. R. (2021). Structural basis for transglycosylation in glycoside hydrolase family GH116 glycosynthases. **Archives of Biochemistry and Biophysics.** 706: 108924.

Polgár, L. (2005). The catalytic triad of serine peptidases. **Cellular and Molecular Life Sciences.** 62(19): 2161-2172.

Pollet, A., Delcour, J. A., and Courtin, C. M. (2010). Structural determinants of the substrate specificities of xylanases from different glycoside hydrolase families. **Critical Reviews in Biotechnology.** 30(3): 176-191.

Poveda, A., & Jiménez-Barbero, J. (1998). NMR studies of carbohydrate–protein interactions in solution. **Chemical Society Reviews.** 27(2): 133-144.

Quiocho, F. A., Wilson, D. K., and Vyas, N. K. (1989). Substrate specificity and affinity of a protein modulated by bound water molecules. **Nature.** 340(6232): 404-407.

Quirke, J. C. K., and Crich, D. (2020). Glycoside Hydrolases Restrict the Side Chain Conformation of Their Substrates To Gain Additional Transition State

Stabilization. **Journal of the American Chemical Society.** 142(40): 16965-16973.

Ramani, G., Meera, B., Vanitha, C., Rajendhran, J., and Gunasekaran, P. (2015). Molecular cloning and expression of thermostable glucose-tolerant beta-glucosidase of Penicillium funiculosum NCL1 in Pichia pastoris and its characterization. **Journal of Industrial Microbiology & Biotechnology.** 42(4): 553-565.

Rojel, N., Kari, J., Sorensen, T. H., Badino, S. F., Morth, J. P., Schaller, K., Cavaleiro, A. M., Borch, K., and Westh, P. (2020). Substrate binding in the processive cellulase Cel7A: Transition state of complexation and roles of conserved tryptophan residues. **Journal of Biological Chemistry.** 295(6): 1454-1463.

Romaniouk, A., and vijay, I. K. (1997). Structure-function relationships in glucosidase I: amino acids involved in binding the substrate to the enzyme. **Glycobiology.** 7(3): 399-404.

Sansenya, S., Mutoh, R., Charoenwattanasatien, R., Kurisu, G., and Ketudat Cairns, J. R. (2015). Expression and crystallization of a bacterial glycoside hydrolase family 116 beta-glucosidase from Thermoanaerobacterium xylanolyticum. **Acta crystallographica Section F Structural Biology Communications.** 71(Pt 1): 41-44.

Santos, C. A., Morais, M. A. B., Terrett, O. M., Lyczakowski, J. J., Zanphorlin, L. M., Ferreira-Filho, J. A., Tonoli, C. C. C., Murakami, M. T., Dupree, P., and Souza, A. P. (2019). An engineered GH1 beta-glucosidase displays enhanced glucose tolerance and increased sugar release from lignocellulosic materials. **Scientific Reports.** 9(1): 4903.

Sarian, F. D., Janeček, Š., Pijning, T., Nurachman, Z., Radjasa, O. K., Dijkhuizen, L., Natalia, D., and Van Der Maarel, M. J. (2017). A new group of glycoside hydrolase family 13 α-amylases with an aberrant catalytic triad. **Scientific Reports.** 7(1): 1-10.

Satoh, H., and Manabe, S. (2013). Design of chemical glycosyl donors: does changing ring conformation influence selectivity/reactivity? **Chemical Society Reviews.** 42(10): 4297-4309.

Seeberger, P. H., and Werz, D. B. (2007). Synthesis and medical applications of oligosaccharides. **Nature.** 446(7139): 1046-1051.

Seidle, H. F., and Huber, R. E. (2005). Transglucosidic reactions of the Aspergillus niger family 3 beta-glucosidase: qualitative and quantitative analyses and evidence that the transglucosidic rate is independent of pH. **Archives of Biochemistry and Biophysics.** 436(2): 254-264.

Shanmugam, N. R. S., Selvin, J. F. A., Veluraja, K., and Gromiha, M. M. (2018). Identification and Analysis of Key Residues Involved in Folding and Binding of Protein-carbohydrate Complexes. **Protein & Peptide Letters.** 25(4): 379-389.

Shaw, A., Bott, R., Vonrhein, C., Bricogne, G., Power, S., and Day, A. G. (2002). A Novel Combination of Two Classic Catalytic Schemes. **Journal of Molecular Biology.** 320(2): 303-309.

Shewale, J. G. (1982). Beta-Glucosidase: its role in cellulase synthesis and hydrolysis of cellulose. **International Journal of Biochemistry.** 14(6): 435-443.

Shi, J., Mukhopadhyay, R., and Rosen, B. P. (2003). Identification of a triad of arginine residues in the active site of the ArsC arsenate reductase of plasmid R773. **FEMS Microbiology Letters.** 227(2): 295-301.

Siggel, M. R., Streitwieser, A., and Thomas, T. D. (1988). The role of resonance and inductive effects in the acidity of carboxylic acids. **Journal of the American Chemical Society.** 110(24): 8022-8028.

Sindhu, R., Binod, P., Madhavan, A., Beevi, U. S., Mathew, A. K., Abraham, A., Pandey, A., and Kumar, V. (2017). Molecular improvements in microbial alpha-amylases for enhanced stability and catalytic efficiency. **Bioresource Technology.** 245(Pt B): 1740-1748.

Singhania, R. R., Patel, A. K., Pandey, A., and Ganansounou, E. (2017). Genetic modification: A tool for enhancing beta-glucosidase production for biofuel application. **Bioresource Technology.** 245(Pt B): 1352-1361.

Singhania, R. R., Patel, A. K., Sukumaran, R. K., Larroche, C., and Pandey, A. (2013). Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production. **Bioresource Technology.** 127: 500-507.

Spiwok, V. (2017). CH/pi Interactions in Carbohydrate Recognition. **Molecules.** 22(7).

Stewart, R., and Yates, K. (1960). The Position of Protonation of the Carboxyl Group1. **Journal of the American Chemical Society.** 82(15): 4059-4061.

Sultana, S., Reichbauer, J., Schule, R., Mochel, F., Synofzik, M., and van der Spoel, A. C. (2015). Lack of enzyme activity in GBA2 mutants associated with hereditary spastic paraplegia/cerebellar ataxia (SPG46). **Biochemical and Biophysical Research Communications.** 465(1): 35-40.

Sultana, S., Stewart, J., and van der Spoel, A. C. (2020). Truncated mutants of beta-glucosidase 2 (GBA2) are localized in the mitochondrial matrix and cause mitochondrial fragmentation. **PLoS One.** 15(6): e0233856.

Tagami, T., Okuyama, M., Nakai, H., Kim, Y. M., Mori, H., Taguchi, K., Svensson, B., and Kimura, A. (2013). Key aromatic residues at subsites +2 and +3 of glycoside hydrolase family 31 alpha-glucosidase contribute to recognition of long-chain substrates. **Biochimica et Biophysica Acta (BBA) - Bioenergetics.** 1834(1): 329-335.

Tawfik, D. S. (2014). Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and catalytic efficiency? **Current Opinion in Chemical Biology.** 21: 73-80.

Taylor, C. B., Payne, C. M., Himmel, M. E., Crowley, M. F., McCabe, C., and Beckham, G. T. (2013). Binding site dynamics and aromatic-carbohydrate interactions in processive and non-processive family 7 glycoside hydrolases. **Journal of Physical Chemistry B.** 117(17): 4924-4933.

Teze, D., Daligault, F., Ferrieres, V., Sanejouand, Y. H., and Tellier, C. (2015). Semi-rational approach for converting a GH36 alpha-glycosidase into an alpha-transglycosidase. **Glycobiology.** 25(4): 420-427.

Uchiyama, T., Katouno, F., Nikaidou, N., Nonaka, T., Sugiyama, J., and Watanabe, T. (2001). Roles of the exposed aromatic residues in crystalline chitin hydrolysis by chitinase A from Serratia marcescens 2170. **Journal of Biological Chemistry.** 276(44): 41343-41349.

Vacas, T., Corzana, F., Jimenez-Oses, G., Gonzalez, C., Gomez, A. M., Bastida, A., Revuelta, J., and Asensio, J. L. (2010). Role of aromatic rings in the molecular recognition of aminoglycoside antibiotics: implications for drug design. **Journal of the American Chemical Society.** 132(34): 12074-12090.

Vagin, A., and Teplyakov, A. (2010). Molecular replacement with MOLREP. **Acta Crystallographica. Section D, Biological Crystallography.** 66(Pt 1): 22-25.

Van der Veen, B. A., Leemhuis, H., Kralj, S., Uitdehaag, J. C., Dijkstra, B. W., and Dijkhuizen, L. (2001). Hydrophobic amino acid residues in the acceptor binding site are main determinants for reaction mechanism and specificity of cyclodextrin-glycosyltransferase. **Journal of Biological Chemistry.** 276(48): 44557-44562.

Varghese, J. N., Hrmova, M., and Fincher, G. B. (1999). Three-dimensional structure of a barley beta-D-glucan exohydrolase, a family 3 glycosyl hydrolase. **Structure.** 7(2): 179-190.

Varki, A. (2017). Biological roles of glycans. **Glycobiology.** 27(1): 3-49.

Vocadlo, D. J., and Davies, G. J. (2008). Mechanistic insights into glycosidase chemistry. **Current Opinion in Chemical Biology.** 12(5): 539-555.

Vyas, N. K., Vyas, M. N., Chervenak, M. C., Johnson, M. A., Pinto, B. M., Bundle, D. R., and Quiocho, F. A. (2002). Molecular recognition of oligosaccharide epitopes by a monoclonal Fab specific for Shigella flexneri Y lipopolysaccharide: X-ray structures and thermodynamics. **Biochemistry.** 41(46): 13575-13586.

Weis, W. I., and Drickamer, K. (1996). Structural basis of lectin-carbohydrate recognition. **Annual Review of Biochemistry.** 65: 441-473.

Wheeler, S. E. (2013). Understanding substituent effects in noncovalent interactions involving aromatic rings. **Accounts of Chemical Research.** 46(4): 1029-1038.

White, A., and Rose, D. R. (1997). Mechanism of catalysis by retaining β-glycosyl hydrolases. **Current Opinion in Structural Biology.** 7(5): 645-651.

Wiberg, K. B., Hadad, C. M., Breneman, C. M., Laidig, K. E., Murcko, M. A., and LePage, T. J. (1991). The response of electrons to structural changes. **Science.** 252(5010): 1266-1272.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. **Acta Crystallographica. Section D, Biological Crystallography.** 67(Pt 4): 235-242.

Woeste, M. A., and Wachten, D. (2017). The Enigmatic Role of GBA2 in Controlling Locomotor Function. **Frontiers in Molecular Neuroscience.** 10: 386.

Wolfenden, R., and Snider, M. J. (2001). The depth of chemical time and the power of enzymes as catalysts. **Accounts of Chemical Research.** 34(12): 938-945.

Yun, S. I., Jeong, C. S., Chung, D. K., and Choi, H. S. (2001). Purification and some properties of a beta-glucosidase from Trichoderma harzianum type C-4. **Bioscience, Biotechnology, and Biochemistry.** 65(9): 2028-2032.

Zakariassen, H., Aam, B. B., Horn, S. J., Varum, K. M., Sorlie, M., and Eijsink, V. G. (2009). Aromatic residues in the catalytic center of chitinase A from Serratia marcescens affect processivity, enzyme activity, and biomass converting efficiency. **Journal of Biological Chemistry.** 284(16): 10610-10617.

Zanphorlin, L. M., de Giuseppe, P. O., Honorato, R. V., Tonoli, C. C., Fattori, J., Crespim, E., de Oliveira, P. S., Ruller, R., and Murakami, M. T. (2016). Oligomerization as a strategy for cold adaptation: Structure and dynamics of

the GH1 beta-glucosidase from Exiguobacterium antarcticum B7. **Scientific Reports.** 6: 23776.

Zechel, D. L., and Withers, S. G. (2000). Glycosidase mechanisms: anatomy of a finely tuned catalyst. **Accounts of Chemical Research.** 33(1): 11-18.

Zhang, X., Wang, S., Wu, X., Liu, S., Li, D., Xu, H., Gao, P., Chen, G., and Wang, L. (2015). Subsite-specific contributions of different aromatic residues in the active site architecture of glycoside hydrolase family 12. **Scientific Reports.** 5: 18357.

Zhao, H., Pan, Q., Zhang, W., Carmichael, I., and Serianni, A. S. (2007). DFT and NMR studies of 2JCOH, 3JHCOH, and 3JCCOH spin-couplings in saccharides: C-O torsional bias and H-bonding in aqueous solution. **Journal of Organic Chemistry.** 72(19): 7071-7082.

Zolotnitsky, G., Cogan, U., Adir, N., Solomon, V., Shoham, G., and Shoham, Y. (2004). Mapping glycoside hydrolase substrate subsites by isothermal titration calorimetry. **Proceedings of the National Academy of Sciences**. 101(31): 11275-11280.
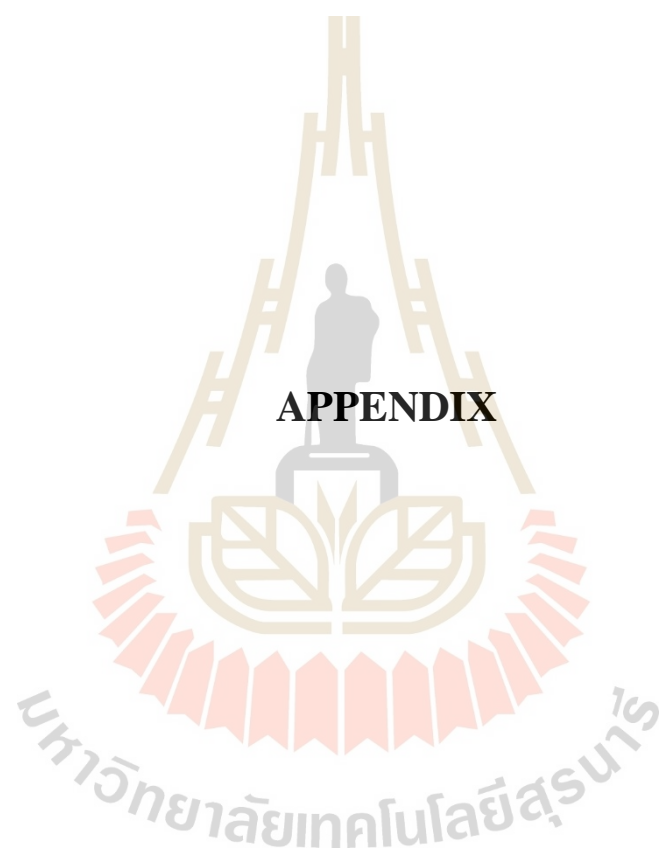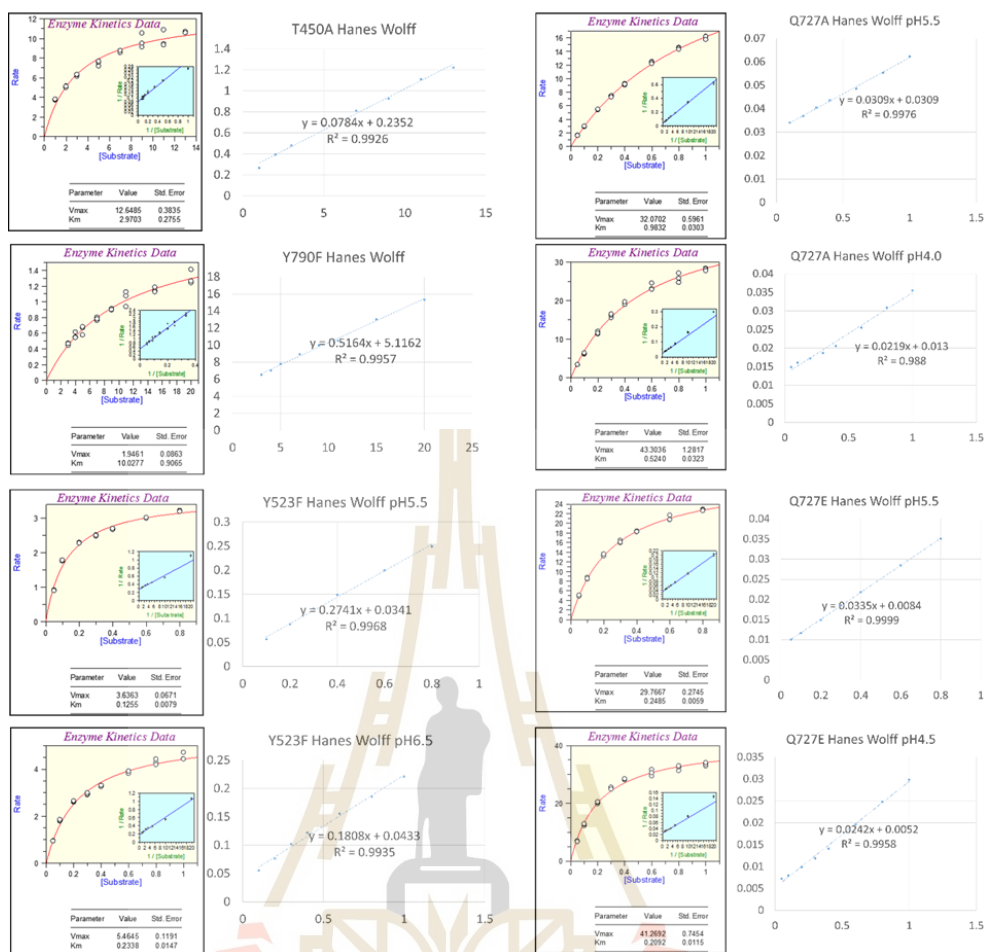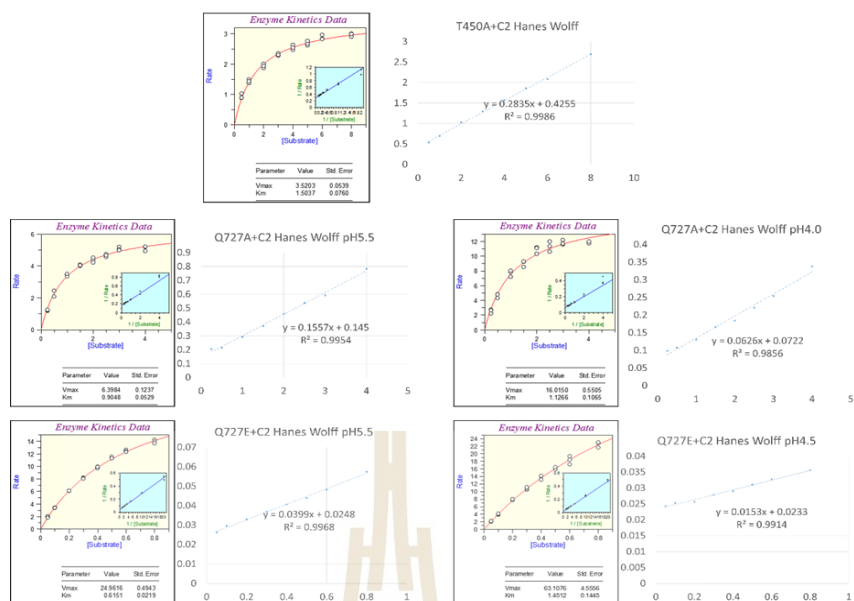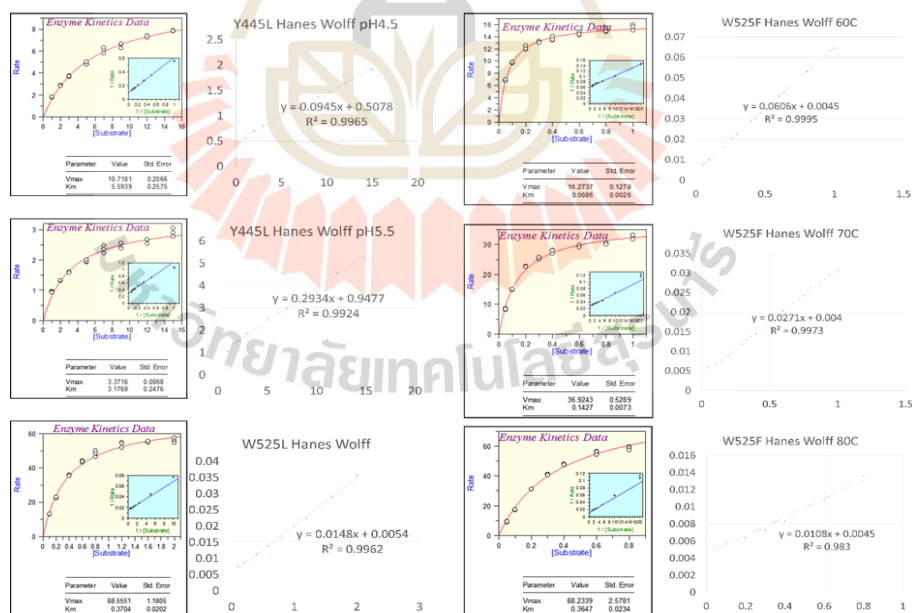
APPENDIX
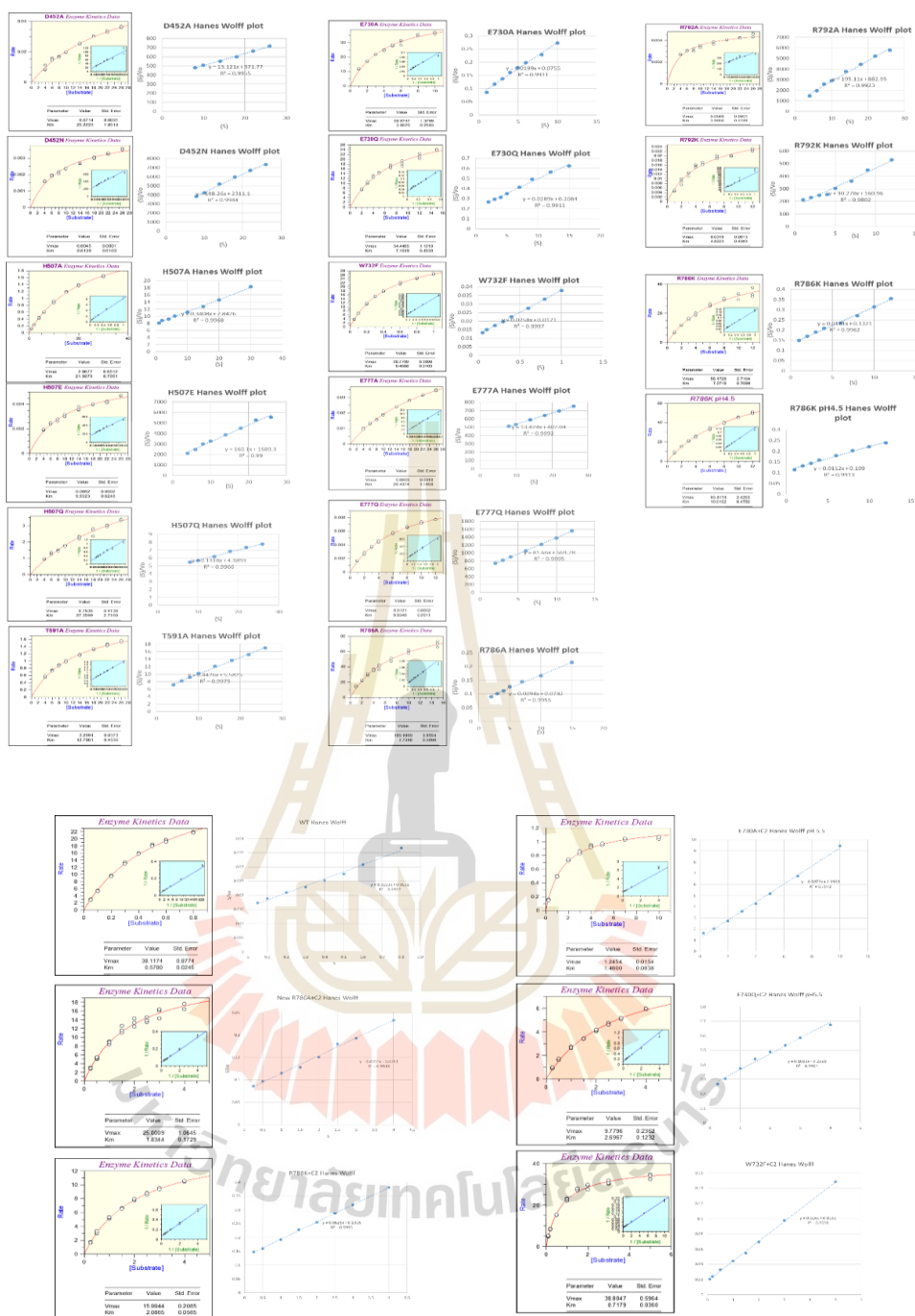
Figure A1　Michaelis–Menten and Hanes Wolff plot (Y-axis: S/V_o, X-axis: S) of pNPGlc substrate kinetics from glycone sugar binding related residues mutants.
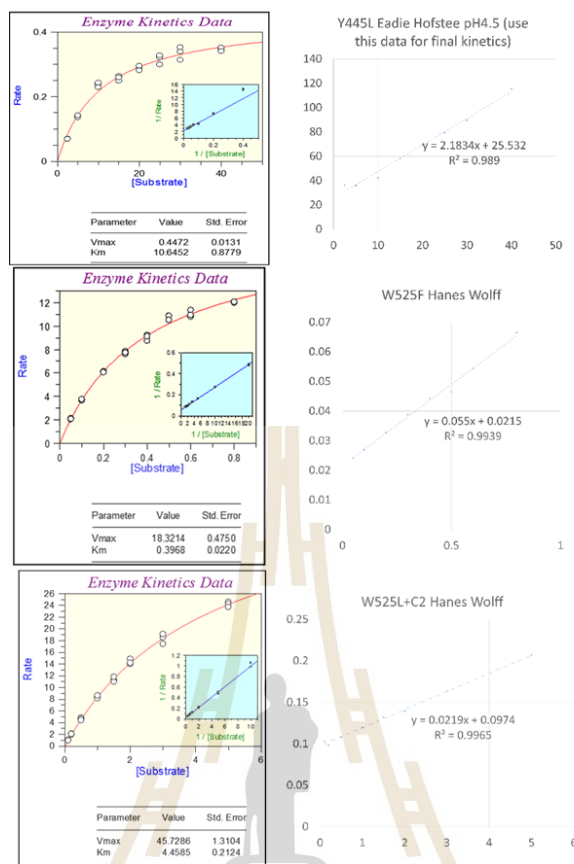
Figure A2 Michaelis–Menten and Hanes Wolff plot (Y-axis: S/V_o, X-axis: S) of pNPGlc substrate kinetics of acid/base and nucleophile related residues mutants.

Figure A3 Michaelis–Menten and Hanes Wolff plot (Y-axis: S/V_o, X-axis: S) of

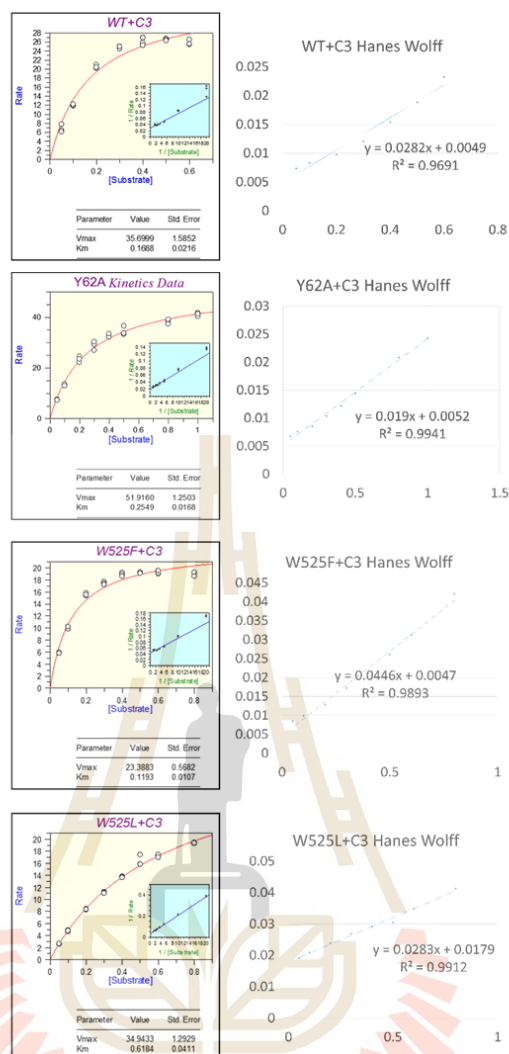cellobiose kinetics of acid/base and nucleophile related residues mutants.

Figure A4 Michaelis–Menten and Hanes Wolff plot (Y-axis: S/V_o, X-axis: S) of

pNPGlc substrate kinetics of subsite+1 and +2 related residues mutants.

Figure A5 Michaelis–Menten and Hanes Wolff plot (Y-axis: S/V_o, X-axis: S) of cellobiose substrate kinetics of subsite+1 and +2 related residues mutants.

Figure A6 Michaelis–Menten and Hanes Wolff plot (Y-axis: S/V_o, X-axis: S) of cellotriose substrate kinetics of subsite+1 and +2 related residues mutants.

CURRICULUM VITAE

Name	Mr. Meng Huang
Date of Birth	December 03, 1990
Place of Birth	Anhui, China
Education	
2016-Present	**PhD of Science** in School of Chemistry, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima
2012-2015	**Master** of Crop Genetics and Breeding in School of Plant Science, Institute of Agriculture, Hainan University, Haikou, China
2008-2012	**Bachelor** of Seed Science and Technology in School of Plant Science, Institute of Agriculture, Anhui Science and Technology University, Fengyang, China