

**PROTEIN CRYSTALLIZATION AND PRELIMINARY
DETERMINATION OF THREE-DIMENSIONAL
STRUCTURE OF RICE β -GLUCOSIDASE**

Sompong Sansenya

A Thesis Submitted in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Biochemistry

Suranaree University of Technology

Academic Year 2008

การเตรียมผลิตและศึกษาโครงสร้างเบื้องต้นของเอนไซม์เบตากลูโคซิเดสของ
ข้าว

นายสมพงษ์ แสนเสนา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาชีวเคมี
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2551

สมพงษ์ แสนเสนา : การเตรียมผลิตและศึกษาโครงสร้างเบื้องต้นของเอนไซม์

เบตากลูโคซิเดสของข้าว (PROTEIN CRYSTALLIZATION AND PRELIMINARY DETERMINATION OF THREE-DIMENSIONAL STRUCTURE OF RICE β -GLUCOSIDASE) อาจารย์ที่ปรึกษา : ผู้ช่วยศาสตราจารย์ ดร.รจนา โอภาสศิริ, 104 หน้า

เอนไซม์ Os4BGlu12 ซึ่งเป็นเอนไซม์เบตากลูโคซิเดสที่จัดอยู่ในตระกูล Glycosyl hydrolase 1 (GH1) สามารถย่อยโอลิโกแซคคาไรด์ที่ประกอบด้วยกลูโคส 3 ถึง 6 หน่วย เชื่อมต่อกันด้วยพันธะ β -(1,4) และไดแซคคาไรด์ที่เชื่อมต่อกันด้วยพันธะ β -(1,3) ได้ เอนไซม์ Os4BGlu12 ถูกผลิตในเชื้อ *Escherichia coli* สายพันธุ์ OrigamiB(DE3) โดยที่ปลายอะมิโนของโปรตีนมีโปรตีนไทโอรีดอกซินและบริเวณที่มีกรดอะมิโนฮิสติดีนเรียงต่อกัน 6 ตัวต่ออยู่ ได้แยกโปรตีนสายผสมนี้ให้บริสุทธิ์ด้วยโครมาโตกราฟีแบบจำเพาะที่มีโคบอลต์เป็นลิแกนด์ต่ออยู่กับเรซิน (IMAC) หลังจากที่ใช้เอนไซม์เอนเทอโรโคเนสตัดโปรตีนไทโอรีดอกซินและบริเวณที่มีกรดอะมิโนฮิสติดีนเรียงต่อกัน 6 ตัวออกจากโปรตีนด้วยแล้ว ได้แยกส่วนนี้ออกไปด้วยวิธี IMAC และได้เอนไซม์ Os4BGlu12 ขนาด 55 kDa ที่มีความบริสุทธิ์มากกว่า 95% ออกมา เอนไซม์ Os4BGlu12 ในสภาพธรรมชาติและที่มีสารยับยั้ง 2,4-dinitrophenyl 2-fluoro-2-deoxy- β -D-glucopyranoside (G2F) ได้ถูกตกผลึกที่ 15 องศาเซลเซียส ผลึกโปรตีนในสภาพธรรมชาติได้ถูกผลิตขึ้นจากวิธี microbatch ในสารละลายตกผลึกที่ประกอบด้วย 25% (w/v) polyethylene glycol (PEG) 4000, 0.1 M Tris-HCl, pH8.5, 0.20 M NaCl ต่อมาได้ปรับเปลี่ยนส่วนประกอบในสารละลายที่พบผลึกนี้เพื่อผลิตผลึกที่มีคุณภาพด้วยวิธี hanging-drop vapour diffusion ร่วมกับเทคนิค microseeding จนได้ผลึกของโปรตีนในสภาพธรรมชาติ และผลึกโปรตีนที่มีสารยับยั้ง G2F ในสารละลายตกผลึกที่ประกอบด้วย 19% (w/v) PEG 3350, 0.1 M Tris-HCl, pH 8.5, 0.16 M NaCl และ 19% (w/v) PEG 2000 0.1 M Tris-HCl, pH 8.5, 0.16 M NaCl ตามลำดับ ผลึกทั้งสองชนิดสามารถหักเหรังสีเอกเรย์ได้ความละเอียดถึง 2.50 และ 2.45 อังสตรอม ตามลำดับ โดยมีกลุ่มสมมาตรโมเลกุลเป็น tetragonal และมี space group แบบ $P4_32_1$ โครงสร้างสามมิติของโปรตีนในสภาพธรรมชาติถูกสร้างขึ้นด้วยวิธี molecular replacement โดยอาศัยโครงสร้างของ 1CBG เป็นโครงสร้างแม่แบบ และพบว่าในโครงสร้างสามมิติของโปรตีนมีโปรตีนสองโมเลกุลในหน่วยสมมาตร มีปริมาณน้ำในผลึกประมาณ 49.98% และค่าสัมประสิทธิ์ของ Matthews คือ $2.46 \text{ \AA}^3 \text{ Da}^{-1}$ สำหรับโครงสร้างสามมิติของโปรตีนกับตัวยับยั้ง G2F ถูกสร้างขึ้นด้วยวิธี rigid body refinement โดยใช้โครงสร้างของโปรตีนในสภาพธรรมชาติเป็นโครงสร้างแม่แบบ ซึ่งมีค่าสัมประสิทธิ์ของ Matthews เท่ากับ $2.68 \text{ \AA}^3 \text{ Da}^{-1}$ และมีปริมาณน้ำในผลึกประมาณ 54.17% จากการเปรียบเทียบโครงสร้างสามมิติกับเอนไซม์ชนิดอื่นที่อยู่ในตระกูลเดียวกันพบว่าโครงสร้างโดยรวมมีลักษณะเหมือนกัน แต่แตกต่างกันตรงโครงสร้างรอบๆ บริเวณทางเข้าของตำแหน่งเร่งปฏิกิริยา ตำแหน่งเร่งปฏิกิริยาของเอนไซม์มีลักษณะเป็นช่อง และมีความลึก

ประมาณ 20 อังสตรอม ตรงบริเวณที่ลึกสุดของตำแหน่งเร่งปฏิกิริยาของโปรตีนกับตัวยับยั้ง G2F มีกรดอะมิโนลำดับอนุรักษ์ทำพันธะไฮโดรเจนอยู่กับกลูโคสซึ่งพบเช่นเดียวกับเอนไซม์อื่นในตระกูลเดียวกัน แต่กลับพบความหลากหลายของกรดอะมิโนที่อยู่รอบๆบริเวณ aglycone binding site ของเอนไซม์ ซึ่งชี้ให้เห็นว่าบริเวณดังกล่าวเกี่ยวข้องกับความจำเพาะต่อสับสเตรตของเอนไซม์ เมื่อเปรียบเทียบระหว่างโปรตีนที่จับกับตัวยับยั้ง G2F กับโปรตีนในสภาพธรรมชาติพบว่าตำแหน่งของกรดอะมิโนที่ทำหน้าที่เป็น nucleophile (Glu393) ไม่แตกต่างกัน ซึ่งแตกต่างจากเอนไซม์เบตาไกลูโคซิเดสชนิดอื่นที่อยู่ในตระกูลเดียวกัน นอกจากนี้ตำแหน่งของกรดอะมิโน Glu393 ในโครงสร้างของ Os4BGlu12 ที่จับอยู่กับตัวยับยั้งมีความคล้ายคลึงกับที่พบในเอนไซม์ S-glycosidase ทั้งขนาดของมุมและระยะทางระหว่างกรดอะมิโนกับคาร์บอนตำแหน่งที่หนึ่งของ G2F ซึ่งสอดคล้องกับการที่เอนไซม์สามารถย่อยสับสเตรต S-glycoside ได้

สาขาวิชาชีวเคมี
ปีการศึกษา 2551

ลายมือชื่อนักศึกษา _____
ลายมือชื่ออาจารย์ที่ปรึกษา _____
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม _____

SOMPONG SANSENYA : PROTEIN CRYSTALLIZATION AND
PRELIMINARY DETERMINATION OF THREE-DIMENSIONAL
STRUCTURE OF RICE β -GLUCOSIDASE. THESIS ADVISOR :
ASST. PROF. RODJANA OPASSIRI, Ph.D. 104.PP.

GLYCOSYL HYDROLASE FAMILY 1/ β -GLUCOSIDASE/RICE/PROTEIN
CRYSTALLIZATION/THREE DIMENSIONAL STRCUTURE

Rice Os4BGlu12, a glycosyl hydrolase family 1 (GH1) β -glucosidase hydrolyzes β -(1,4)-linked oligosaccharides of 3-6 glucosyl residues and the β -(1,3)-linked disaccharide laminaribiose. Os4BGlu12 was expressed as an N-terminal thioredoxin/His₆ fusion protein in OrigamiB(DE3) *Escherichia coli*. The fusion protein was purified by immobilized metal-affinity chromatography (IMAC) with cobalt resin. After the thioredoxin/His₆ tag was excised from the protein with enterokinase, the fusion tag was removed by adsorption to the IMAC resin, which yielded a 55 kDa Os4BGlu12 with >95% purity. The free Os4BGlu12 enzyme and a complex with 2,4-dinitrophenyl 2-fluoro-2-deoxy- β -D-glucopyranoside (G2F) inhibitor were crystallized at 15°C. Microbatch crystallization screening generated the native crystals in 25% (w/v) polyethylene glycol (PEG) 4000, 0.1 M Tris-HCl, pH 8.5, 0.20 M NaCl. The conditions were further optimized by the hanging-drop vapor-diffusion method with microseeding. Native crystals and crystals of Os4BGlu12 complexed with G2F were obtained in 19% (w/v) PEG 3350, 0.1 M Tris-HCl, pH 8.5, 0.16 M NaCl and 19% (w/v) PEG 2000, 0.1 M Tris-HCl, pH 8.5, 0.16 M NaCl, respectively. Crystals of free Os4BGlu12 and Os4BGlu12-G2F complex were diffracted to 2.50 and 2.45 Å resolution, respectively, and their unit cell symmetry

determined to be in the tetragonal $P4_32_12$ space group. The structure of native Os4BGlu12 was solved by molecular replacement with the 1CBG structure as a search model and had two molecules per asymmetric unit with a solvent content of 49.98% and a Mathews Coefficient (V_M) of $2.46 \text{ \AA}^3 \text{ Da}^{-1}$. The native Os4BGlu12 structure further served as a template for rigid body refinement to solve the Os4BGlu12 with G2F data set, which had a V_M of $2.68 \text{ \AA}^3 \text{ Da}^{-1}$ and 54.17% solvent content. The structures were similar to previous known GH1 enzymes, but the significant differences were seen at the main-chain trace of loop surrounding the active site. The active site is located at the bottom of an approximately 20 Å deep slot-like pocket surrounded by a large surface loop. In the innermost part of the active site in the crystal structure of G2F complex, the surrounding conserved amino acid residues seen in other GH1 enzymes formed hydrogen bonds with the glucosyl unit. On the other hand, residues around the aglycone binding site are not conserved, which might indicate the substrate specificity of the enzyme. There was no movement of Glu393 nucleophile residue in glycosyl-enzyme intermediate complex when compared to free Os4BGlu12 structure, which is different from the other known GH1 β -glucosidases. In addition, the position of the Glu393 nucleophilic residue of Os4BGlu12-G2F has the angle and distance to the anomeric carbon of G2F similar to an S-glycosidase, consistent with its ability in the hydrolysis of S-glycosides.

School of Biochemistry

Student's Signature _____

Academic Year 2008

Advisor's Signature _____

Co-advisor's Signature _____

ACKNOWLEDGEMENTS

I would like to express the deepest gratitude to my thesis advisor, Asst. Prof. Dr. Rodjana Opassiri for her kindly advice, guidance and encouragement along this project. I am also deeply grateful to my co-advisor, Assoc. Prof. Dr. James R. Ketudat-Cairns, for his idea to start this project and for his valuable guidance and advice on X-ray crystallography and structural analysis.

I am very grateful to Dr. Buabarn Kuaprasert at the Synchrotron Light Research Institute (SLRI), Thailand for her guidance on X-ray crystallography. I am also grateful to Assoc. Prof. Dr. Chun-Jung Chen, Dr. Phimonphan Chuankhayan and the staffs at National Synchrotron Radiation Research Center (NSRRC), Taiwan for their kindly help to solve the problems during X-ray diffractions and data collection.

Special thanks are extended to all friends in the school of Biochemistry, Suranaree University of Technology for their help, support, and valuable hints. I am obliged to Miss. Salila Pengthaisong for her guidance on the crystallization technique, and Miss. Supriya Seshadri for her value suggestion for improvement of my thesis writing.

This work was supported by a grant from the Synchrotron Light Research Institute (SLRI). Data collection was carried out at the NSRRC, a national user facility supported by the National Science Council of Taiwan, R.O.C. The Synchrotron Radiation Protein Crystallography Facility is supported by the National Research Program for Genomic Medicine of Taiwan.

Finally, I would like to give my special thanks to my family whose patience and love enabled me to complete this work.

Sompong Sansenya

CONTENT

	Page
ABSTRACT IN THAI.....	I
ABSTRACT IN ENGLISH	III
ACKNOWLEDGEMENT	V
CONTENTS.....	V
LIST OF TABLES	XII
LIST OF FIGURES	XIV
LIST OF ABBREVIATIONS.....	XVII
 CHAPTER	
I INTRODUCTION.....	1
1.1 Glycosyl Hydrolase	1
1.2 GH Clan A.....	1
1.3 GH family 1 β -glucosidases	2
1.4 Catalytic Machanism.....	4
1.5 Crystal structure of GH1 enzymes	5
1.6 Active site and substrate specificity of GH1 enzymes.....	7
1.6.1 Glycone specificity	8
1.6.1 Glycone specificity	9
1.7 Rice β -glucosidases and Os4BGlu12 β -glucosidase	11
1.8 Research Objectives	16

CONTENTS (Continued)

	Page
II MATERIALS AND METHODS	17
2.1 Materials	17
2.1.1 Recombinant plasmid.....	17
2.1.2 Chemicals and Reagents	17
2.2 General methods.....	18
2.2.1 Transformation and of recombinant expression plasmid into expression host cells.....	18
2.2.2 Expression of recombinant β -glucosidase was produced as thioredoxin-Os4BGlu12 fusion protein in E.Coli system.....	19
2.2.3 Protein extraction from E.Coli	19
2.2.4 Purification of recombinant thioredoxin-Os4BGlu12 nonfusion protein	20
2.2.5 Purification of recombinant-Os4BGlu12 nonfusion protein.....	20
2.2.6 Protein analysis	21
2.2.6.1 SDS-PAGE electrophoresis	21
2.2.6.2 Enzyme activity	22
2.2.6.3 Bio-Rad protein assay	23
2.3 Protein Crystallization	23
2.3.1 Initial screening for crystallization conditions.....	23

CONTENTS (Continued)

	Page
2.3.2 Optimization of crystallization conditions	
by vapour diffusion method	24
2.3.3 Microseeding method.....	27
2.3.4 Crystallization of complexes.....	28
2.3.5 Determination that crystals were protein	28
2.4 Data collection and processing.....	29
2.4.1 Freezing in cryo-protectant.....	29
2.4.2 Synchrotron X-ray diffraction.....	29
2.5 Structure solution by molecular replacement.....	30
2.6 Model building, structure refinement and validation of final model	31
III RESULTS	32
3.1 Protein expression and purification	32
3.1.1 Purification of recombinant Os4BGlu12	32
3.1.2 Protein purification for crystallization	35
3.1.3 Dynamic light scattering (DLS).....	36
3.2 Crystallization of Os4BGlu12	37
3.2.1 Initial screening of crystallization of freeOs4BGlu12	37
3.2.2 Optimization of freeOs4BGlu12 crystallization	41
3.2.3 Optimization with seeding	44

CONTENTS (Continued)

	Page
3.2.4 Optimization of Os4BGlu12-G2F complex crystal	49
3.3 Crystal characterization	50
3.4 Preliminary X-ray diffraction analysis of native Os4BGlu12 and complex Os4BGlu12 crystals	51
3.5 Molecular replacement	53
3.6 Structure determination of free-Os4BGlu12 and Os4BGlu12 with G2F inhibitor	54
3.6.1 The model quality of freeOs4BGlu12 and G2F complex	54
3.6.2 The Overall structure of Os4BGlu12 with and without G2F	57
3.6.3 Comparison of Os4BGlu12 structure with other known GH1 structures	61
3.6.4 Enzyme active site	66
IV DISCUSSION	75
4.1 Protein expression and purification	75
4.2 Crystallization of Os4BGlu12 crystal	76
4.2.1 Initial crystallization	76
4.2.2 Optimization of crystal	77
4.3 Crystal structure of Os4BGlu12	79
4.3.1 Overall structure	79

CONTENTS (Continued)

	Page
4.3.2 Dimerization by zinc ion.....	79
4.3.3 Disulfide bond formation.....	80
4.3.4 The presence of Tris molecule in the active site.....	82
4.3.5 Active site structure	82
V CONCLUSION	89
REFERENCES	92
CURRICULUM VITAE	138

LIST OF TABLES

Table	Page
3.1 List of JY Screen conditions that produced native Os4BGlu12 crystals at 15°C	38
3.2 Data collection and processing statistics.....	53
3.3 Refinement statistics for native Os4BGlu12 and Os4BGlu12 with G2F	55
3.4 Root-mean-square deviation between the structures of Os4BGlu12 and other plant β -glucosidases.....	63
3.5 Hydrogen bonds occurring between two conformations of Tris molecules and amino acid residues in the enzyme's active site	70
3.6 Hydrogenbonding interaction between G2F and amino acid residues in the enzyme's active site	72

LIST OF FIGURES

Figure	Page
1.1 Generally accepted endocyclic pathway of the double displacement mechanism proposed for the retaining β -glucosidases.....	5
1.2 Ribbon representation of β -D-glucosidase from wheat.....	7
1.3 Alignment of the predicted mature protein sequences of rice Os4BGlu12 with related β -glucosidases.....	15
2.1 Construct of the protein-coding sequence of recombinant pET32a+/DEST	17
2.2 Schematically diagrams of microbatch crystallization trials.....	24
2.3 The strategy for optimization of Os4BGlu12 crystallization	27
3.1 SDS-PAGE of recombinant thioredoxin-Os4BGlu12 protein purified by Co^{2+} -IMAC. Lanne M, molecular weight protein marker.....	33
3.2 The SDS-PAGE of purified Os4BGlu12 nonfusion protein fractions from second IMAC step	34
3.3 Elution profile of the Os4BGlu12 nonfusion protein obtained from Sephacryl S200 chromatography.....	35
3.4 SDS-PAGE analysis of Os4BGlu12 protein from two steps of purification. Lanne M, molecular weight protein marker	36

LIST OF FIGURES (Continued)

Figure	Page
3.5 Dynamic light scattering (DSL) spectrum of tag-free Os4Bglu12 protein. DSL was used to determine the mean particle size and size distribution in a solution of 1 mg/mL protein in 20 mM Tris-HCl, pH 7.2, and was performed at 293K.....	37
3.6 Photographs of crystals obtained from the initial crystallization trials using JYS Screen kit solutions in micro-batch plates	39
3.7 Photographs of crystals in condition 7F	40
3.8 Variable concentrations of PEG 4000 and NaCl for the Optimization of freeOs4BGlu12 crystals.....	42
3.9 Variation of protein and NaCl concentrations shown at the fixed concentration of 19% PEG 4000 and 0.1 M Tris-HCl, pH8.5	44
3.10 The seeding lines of Os4BGlu12 crystals obtained from the dilution series of seed stock of seed stock prepared from the crystals obtained from condition number 1 in figure 3.9.....	45
3.11 The positive conditions of the NaCl concentration optimized.....	46
3.12 Optimization of freeOs4BGlu12 with the variation of PEG molecular weight, and concentration of NaCl	71
3.13 Optimization of Os4BGlu12-G2F complex crystals with the variation of PEG molecular weight (19%) and NaCl concentration in 0.1 M Tris-HCl, pH 8.5, and 3.0 mg/mL protein.....	71

LIST OF FIGURES (Continued)

Figure	Page
3.14 SDS-PAGE of protein from dissolved crystals of Os4BGlu12. Lane 2, dissolved Os4BGlu12 crystals and lane 1, molecular weight protein marker	51
3.15 $2F_{\text{obs}}-F_{\text{cal}}$ maps of 2.50 Å resolution free-Os4BGlu12 (A) and 2.45 Å resolutions Os4BGlu12 with G2F (B)	56
3.16 Ribbon diagram representation of the dimeric structure of freeOs4BGlu12 (A) and G2F complex (B)	59
3.17 Stereo views of the Zn^{2+} ion binding site. Key residues interacting with the zinc ion are labeled and depicted as sticks with carbon, nitrogen and oxygen atoms colored in white	60
3.18 Ribbon representation of the superimposition of rice Os4BGlu12 (green) with other GH1 enzymes	64
3.19 Sequence alignment of rice Os4BGlu12, rice BGlu1 β -glucosidase (2RGM), <i>Trifolium repens</i> (1CBG), <i>Sinapis alba</i> (1MYR), β -glucosidase isozyme I from <i>Zea mays</i> (1E1E), dhurrinase isozyme I from <i>Sorghum bicolor</i> (1VO2) and β -glucosidase from <i>Triticum aestivum</i> (2DGA)	65
3.20 Electrostatic surface representation of Os4BGlu12 with G2F located deep in the active site and the surface showing positively charged region in blue, negatively charged region in red and neutral region in white	66

LIST OF FIGURES (Continued)

Figure	Page
3.21 Two Tris molecules from A and B monomer with the surrounding electron density maps.....	68
3.22 Stereo view of active site region of freeOs4BGlu12 with Tris bound. Tris molecules are shown in the active site of A and B.....	69
3.23 Structure of glycosyl-enzyme intermediate.....	73
3.24 Stereo view of protein-ligand interaction in the active site of Os4BGlu12-G2F complex	74
4.1 Superimposition of the 3D structure of different plant glycosyl hydrolase family 1 β -glucosidase.....	81
4.2 Amino acid sequence alignment of Os4BGlu12 and rice BGlu1.....	83
4.3 Comparision of Os4BGlu12 and rice BGlu12 E176Q ligand interaction	85
4.4 Comparison of the position of the nucleophilic residues covalently bound between ligand at the -1 subsite of O-glycosidase (Os4BGlu12-G2F complex and rice BGlu1-G2F complex (2RGM) and S-glycosidase (myrosinase-G2F-ascorbate complex (1E73)).....	88

LIST OF ABBREVIATIONS

BSA	Bovine Serum Albumin
cDNA	Complementary deoxynucleic acid
°C	Degree Celsius
DNase	Deoxyribonuclease
EDTA	Ethylene diamine tetraacetic acid
hr	Hour
kDa	Kilo Dalton
LB	Luria petroni broth
min	Minute
(m, μ)g	(milli, micro) Gram
(m, μ) l	(milli, micro) liter
(m, μ)M	(milli, micro) Molar
(μ)mol	(micro) Mole
MW	Molecular Weight
NCS	noncrystallographic symmetry
nm	nanometers
IPTG	Isopropyl- β -D-thiogalactopyranoside
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PEG	Polyethyleneglycol

LIST OF ABBREVIATIONS (Continued)

<i>p</i> NP	<i>p</i> -Nitrophenolate
<i>p</i> NPG	<i>p</i> -Nitrophenyl- β -D-glucopyranosid
rpm	rotations per minute
S200	Superdex 200
SDS	Sodium dodecyl sulfate
s	second
TEMED	Tetramethylenediamine
Tris	Tris-(hydroxymethyl)-aminoethane
V_m	Matthew' coefficient
v/v	Volume/volume
w/v	Weight/volume

CHAPTER I

INTRODUCTION

1.1 Glycosyl Hydrolases

Glycosyl hydrolases (GH; EC 3.2.1-3.2.3) are a widespread group of enzymes that responsible for the cleavage of glycosidic bonds between sugars or between sugar and nonsugar aglycone moieties. Glycosyl hydrolases have been classified based on the basis of two independent criteria. The first criteria is based on their EC (Enzyme Commission) numbers or catalytic specificity, while the second criteria is based on sequence homology (Rojan *et al.*, 2004). GH have been classified into more than 115 families based on amino acid sequence similarities (Henrissat, 1991; Henrissat and Bairoch, 1993, 1996; Henrissat and Davies, 1997). The up-to-date information is available on Carbohydrate-Active Enzymes database (CAZY at <http://www.cazy.org/CAZY/index.html>). The three-dimensional structures of the enzymes recently have been found to be more strongly conserved than the sequence. Based on their three-dimensional structures, GH can be grouped into clans of related structures (or superfamilies), except for GH family 45 which cannot be classified into any clan (Henrissat and Davies, 1997).

1.2 GH Clan A

GH Clan A is the largest group and the proteins share a similar core (β/α)₈ barrel structure. The active sites of these enzymes are located at the C-terminal

portion of the β -barrels and are surrounded by loops connecting the α -helices to the β -strands. Two catalytic amino acid residues are located at the ends of strands 4 and 7 of the barrel and the enzymes hydrolyze the glycosidic bond with retention of the anomeric configuration (Jenkins *et al.*, 1995; Henrissat *et al.*, 1995; Coutinho and Henrissat, 1999). This group comprises more than 1000 enzymes with various substrate specificities from 17 different families, including GH families 1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79 and 86 (Henrissat *et al.*, 1995; Jenkins *et al.*, 1995; Bolam *et al.*, 1996; Henrissat and Bairoch, 1996).

1.3 GH family 1 β -glycosidases

β -Glycosidases catalyze the hydrolysis of β -*O*-glycosidic bonds at the nonreducing end of glycoside molecules to release a glycone and an aglycone (Reese, 1977). GH family 1 (GH1) contains a wide range of β -glycosidases with a total of 15 different EC numbers (or activities), including β -glucosidases, β -galactosidases, β -mannosidases, thioglucosidases, phospho- β -galactosidases, and phospho- β -glucosidases (Marana, 2006). Currently, there are more than 652 different GH1 β -glycosidase sequences in the protein databases (Marana, 2006).

In GH family 1 (GH1), most catalytic sites are in the form of a pocket or crater. The two glutamic acid residues which serve as the catalytic amino acids are located in the highly conserved peptide motifs Thr-Phe-Asn-Glu-Pro (TFNEP) and Ile-Thr-Glu-Asn-Gly (ITENG) of β -stands 4 and 7, respectively. In GH1 myrosinases, the acid/base residue is replaced by glutamine, and ascorbate is a cofactor and replaces the catalytic acid base (Burmeister *et al.*, 2000). GH1 enzymes also share similar

environmental surrounding the active site center, which include many polar and aromatic residues located around the end of the glycone binding pockets (Barrett *et al.*, 1995; Burmeister *et al.*, 1997 and Sanz-Aparicio *et al.*, 1998). Although the enzymes in GH1 have similar structures and the same stereospecificity and mechanism of glycoside hydrolysis, only small sequence differences affect substrate specificity of the enzymes (Esen, 1993; Warren *et al.*, 2000).

GH1 β -glycosidases are found in all types of organisms, including fungi, bacteria, archaea, animals and plants, with a variety of functions in different organisms. These enzyme have been identified to be involved in several important biochemical and physiological roles in plants, such as hydrolysis of cell wall-derived oligosaccharides during germination (Leah *et al.*, 1995; Hrmova *et al.*, 1996), release of toxic aglycones and aromatic volatiles such as hydrogen cyanide, isoflavones, and flavonoids from inactive glycosides for defense against plant pathogens and herbivores (Poulton, 1990; Duroux *et al.*, 1998; Mizutani *et al.*, 2002), release of plant hormones from glycosides, such as the auxin, cytokinin, gibberellin and ABA (Smith and Van Staden, 1978; Zouhar, 1999; Schliemann, 1984, Brzobohaty *et al.*, 1993; Dietz *et al.*, 2000), lignification by the hydrolysis of coniferin glycoside to the moltiprenol coniferyl alcohol (Dharmawardhana *et al.*, 1995), and release of metabolic intermediates by cleaving glucose blocking groups from the inactive glucosides, such as strictosidine in the synthesis of monoterpene indole alkaloids (Barleben *et al.*, 2005). It seems evident that the roles of these enzymes are related to the substrate specificity and the activities of the substrates and products released.

1.4 Catalytic Mechanism

Two mechanisms of GH, inverting and retaining mechanisms, have been identified (McCarter and Withers, 1994). The mechanism is classified by the anomeric configuration of products generated from the reaction which is inverted or the same as the substrate. Two active site carboxylic acids residues are involved in both mechanisms with different roles. Inverting glycosidases catalyze via a direct displacement of the aglycone (leaving group) by water (which acts as a nucleophile). During catalysis, one carboxylated residue acts as general base to activate water as a nucleophile. The activated water then attacks the anomeric carbon of the sugar, while another carboxylated residue acts as general acid to assist the departure of the aglycone part. The distance between these two catalytic residues is $\sim 9\text{\AA}$, which allows the water and substrate to bind simultaneously (Zechel and Withers, 2000).

In retaining glycosidases including GH1, the cleavage of the glycosidic bond is performed by the double displacement mechanism (Figure 1.1). The distance between these two catalytic residues is $\sim 5-5.5\text{\AA}$. The catalysis is performed in two separated steps. In the first step, one carboxylated residue acts as an acid catalyst, protonating the glycosidic oxygen, while the other carboxylated residue acts as a nucleophile by attacking at the sugar anomeric center to form a covalent glycosyl-enzyme intermediate. In the second step, the resulting glycosyl enzyme is hydrolyzed by a water molecule and this second nucleophilic substitution at the anomeric carbon generates a product with the same stereochemistry as the substrate (Davies and Henrissat, 1995; Zechel and Withers, 2001).

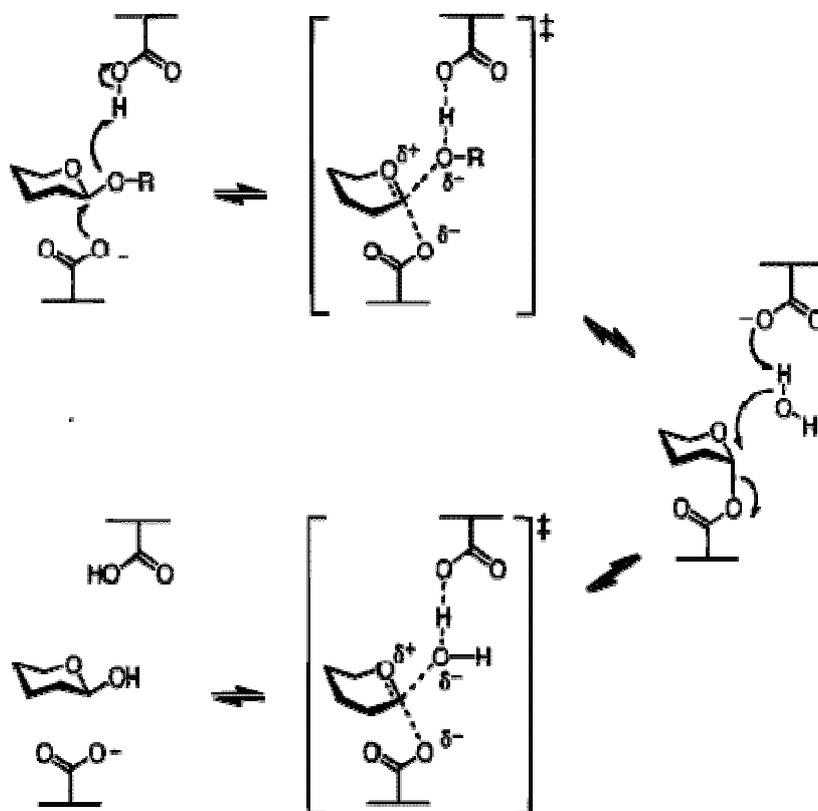


Figure 1.1. Generally accepted endocyclic pathway of the double displacement mechanism proposed for the retaining β -glucosidases (Zechel and Wither, 2001).

1.5 Crystal structures of GH1 enzymes

To date, the three-dimension structures of 21 GH1 β -glucosidases have been deposited in the Protein Databank (PDB). The PDB structures include 1 from a fungus, *Phanerochaete chrysosporium* β -glucosidase (Nijikken *et al.*, 2007); 2 from animals, *Brevicoryne brassicae* myrosinase (Husebye *et al.*, 2005) and human neutral β -glycosylceramidase (Tribolo *et al.*, 2007); 3 from archaea, *Pyrococcus horikoshii* alkyl β -glucosidase (Akiba *et al.*, 2004), *Sulfolobus solfataricus* β -glucosidase

(Aguilar *et al.*, 1997) and *Thermospaera aggregans* β -glycosidase (Chi *et al.*, 1999); 8 from bacteria, *Bacillus circulans* β -glucosidase (Hakulinen *et al.*, 2000), *Lactococcus lactis* 6-phospho- β -galactosidase (Wiesmann *et al.*, 1995), *Paenibacillus polymyxa* β -glucosidases (Sanz-Aparicio *et al.*, 1998; Isoma *et al.*, 2007), *Streptomyces* sp. β -glucosidase (Guasch *et al.*, 1999), *Thermotoga maritima* β -glucosidase (Zechel *et al.*, 2003), *Thermus nonproteolyticus* β -glycosidase (Wang *et al.*, 2003), and *Thermus thermophilus* β -glycosidase (Lakanath *et al.*, to be published). In plants, 8 structures have been determined including *Rauvolfia serpentina* strictosidine β -glucosidase (Barleben *et al.*, 2007), *Sinapis alba* myrosinase (Burmeister *et al.*, 1997), *Sorghum bicolor* dhurrinase (Verdoucq *et al.*, 2004), *Trifolium repens* cyanogenic β -glucosidase (Barrett *et al.*, 1995), *Triticum aestivum* β -glucosidase (Sue *et al.*, 2006), *Zea mays* β -glucosidase (Czjzek *et al.*, 2000), rice BGlu1 β -glucosidase (Chuenchor *et al.*, 2008) and rice Os3BGlu6 β -glucosidase (Seshadri, 2009).

Among above structures, the complex with inhibitor and mutants with substrates have also been deposited in the database. All structures of these enzymes share a common $(\beta/\alpha)_8$ barrel motif (Figure 1.2). The structural differences among GH1 enzymes have being found mainly in the loops at the C-terminal end of the β -barrel around the active site, which vary in both length and sequences and dictate the substrate-specificity of the enzyme (Henrissat *et al.*, 1995; Burmeister *et al.*, 1997; Czjzek *et al.*, 2000; 2001).

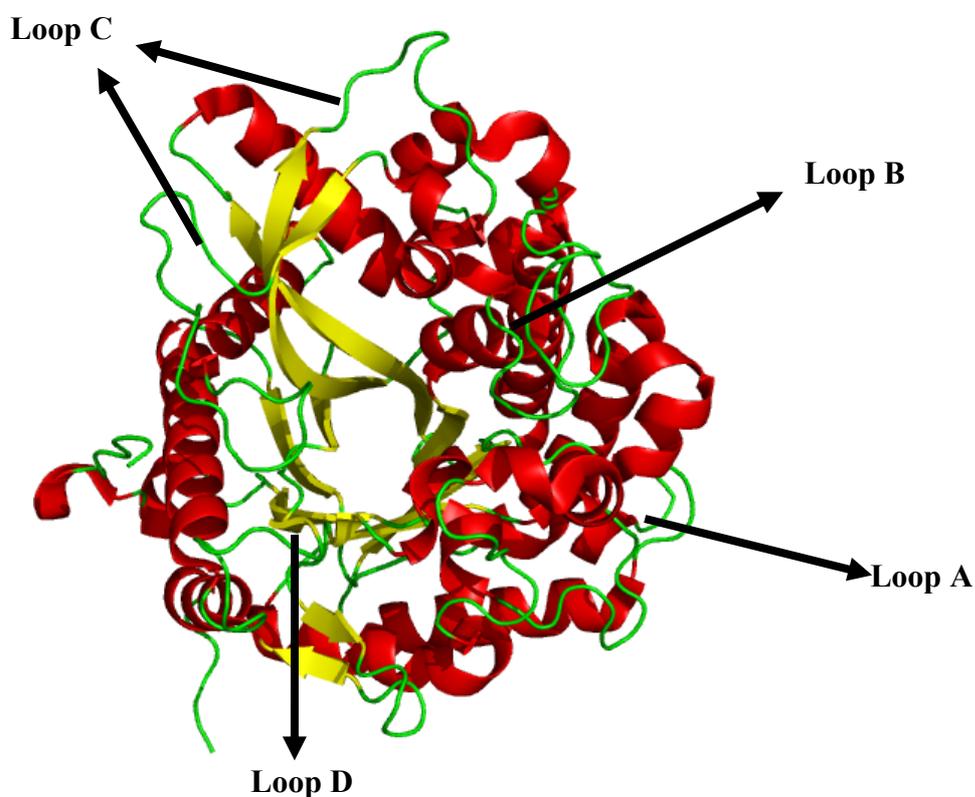


Figure 1.2 Ribbon representation of β -D-glucosidase from wheat.

The loops surrounding the active site pocket include loop A, loop B, loop C and loop D (Sue *et al.*, 2006).

1.6 Active site and substrate specificity of GH1 enzymes

Based on the three-dimensional structures, the active site of GH1 enzymes range from narrow slots to wide pockets and can be divided into two parts, the glycone and aglycone binding subsites. The glycone subsite (or called subsite -1) binds the nonreducing end monosaccharide of the substrate, whereas the aglycone binding site binds with the remaining part of the substrate. The aglycone binding

region may be formed by several subsites (+1, +2, +3 and so on). The substrate cleavage point is between subsites -1 and +1 (Marana, 2006).

The glycones recognized by β -glycosidases include glucose, galactose, fucose, mannose, xylose, 6-phospho-glucose and 6-phospho-galactose (Czjzek *et al.*, 2000; Verdoucq *et al.*, 2004). The diversity of aglycones is higher, including monosaccharides, oligosaccharides, and aryl or alkyl groups (Coutinho and Henrissat, 1999). The enzymes may hydrolyze a broad range of different glycones or aglycones of substrates even with different specificity, but some enzymes may be specific for only one type of glycone or aglycone. The fundamental substrate specificity of these enzymes depends on their overall binding shape, dimensions and geometry of binding site, and conformation of the active site amino acids that are important for the substrate recognition and binding, as well as the structure of the aglycone and glycone moieties of the substrate (Czjzek *et al.*, 2000; Verdoucq *et al.*, 2004).

1.6.1 Glycone specificity

Since several crystal structures of enzyme-ligand complexes have been determined, the residues forming the glycone binding site (subsite -1) and the non-covalent interaction with the glycone are well identified (Burmeister *et al.*, 1997; Czjzek *et al.*, 2000; 2001; Zechel *et al.*, 2003; Verdoucq *et al.*, 2004; Isorna *et al.*, 2007; Chuenchor *et al.*, 2008). The interactions between the residues at subsite -1 and the monosaccharide of the substrate nonreducing end is involved a hydrogen bond network of at least 5 active site amino acids and 4 substrate hydroxyls (Marana, 2006). All hydrogen bonds and residues forming these interactions are conserved among GH1 enzymes. In *Z. mays*, amino acid residues critical for glucose binding, as

identified by Czjzek *et al.* (2001) include Gln38, Trp465 and His142 interact with OH3, Gln38, Trp457 and Glu464 which interact with OH4, Glu464, which interacts with OH6, and the basal plate form residue Trp457. This basal platform establishes stacking interactions with the glycone ring in the form of soft hydrogen bonds between the pi electron of the indole ring and the slightly positively charge C-H protons of the glucose ring. In *T. maritima*, *Bacillus polymyxa*, *Sinapis alba*, *Lactococcus lactis*, *Z. mays* and *S. bicolor* β -glycosidases, the OH2 forms a bidentate hydrogen bond with an asparagine and a glutamate residue (eg. Asn188 and Glu404 in *S. bicolor*). This glutamate also acts as the catalytic nucleophile of the β -glycosidases. A histidine residue also forms a hydrogen bond with OH2 in *B. polymyxa* and *S. bicolor*. Although the amino acid residues forming the hydrogen bond network at subsite -1 are conserved, differences of glycone specificity of the GH1 enzymes have been reported (Czjzek *et al.*, 2000, Opassiri *et al.*, 2004; 2006).

1.6.2 Aglycone specificity

The high diversity of aglycones recognized by β -glycosidases is correlated with the structural variability of the aglycone binding subsite (Marana, 2006). The study of the active site structure and the amino acids which are responsible for substrate binding and recognition should reveal the substrate specificity of these enzymes. Just three β -glycosidases crystallized in complex with a ligand occupying the aglycone binding region have as yet been published (Czjzek *et al.*, 2000; Verdoucq *et al.*, 2004, Isoma *et al.*, 2007).

The structural data of *Z. mays* ZmGlu1 and *S. bicolor* SbDhr1 with their natural substrates DIMBOA-Glc and dhurrin revealed differences in the shape

and noncovalent interactions of aglycone binding site (Czjzek *et al.*, 2000; 2001; Verdoucq *et al.*, 2004). ZmGlu1 hydrolyzes a broad range of substrates, while SbDhr1 hydrolyzes only dhurrin. The aglycone binding site of both enzymes is formed by two walls called the basal platform and ceiling. The basal platform is formed by the a highly conserved tryptophan (Trp378 in ZmGlu1 and Trp376 in SbDhr1). The differences between these enzymes were seen on their ceilings. In the ceiling site of ZmGlu1-DIMBOAGlc ES complex, the hydrophobic amino acids (Phe198, Phe205, Phe466 and Trp378) interact with the aglycone of the substrate. In the ES complex between SbDhr1 and dhurrin, the aglycone forms indirect hydrogen bonding with two polar residues, Ser462 and Asn259, in addition to the hydrophobic interaction with amino acid residues Val196, Leu203 and Tyr376. This indicates the strict substrate specificity of SbDhr1 due to its tighter binding of the aglycone moiety than in the case of ZmGlu1.

The residues that do not directly contact the aglycone are also involved with the substrate binding by contributing the correct positioning of the residues that interact with the aglycone. In ZmGlu1, residues Trp53, Phe56, Trp143, Phe195 and W465 help in the positioning of the residues that form the ceiling site, while Tyr473 forms a hydrogen bond with Trp378 (Czjzek *et al.*, 2000; 2001). A sequence alignment of crystallized β -glycosidases reveals that the residues forming the basal platform and the group of residues involved in the positioning of the ceiling site forming residues are highly conserved, whereas residues that form the ceiling are more variable and might determine aglycone specificity (Marana, 2006). However, the exchange of the ceiling residues ZmGlu1 and SbDhr1 only reduced their activity upon their own substrates but did not yield an aglycone specificity exchange

(Verdoucq *et al.*, 2003). Therefore, the molecular basis of the specificity of different β -glycosidases based on the interactions with the aglycone remains to be studied.

1.7 Rice β -glucosidases and Os4BGlu12 β -glucosidase

Forty genes homologous to GH1 β -glucosidases have been identified in rice genome, and 33 of which apparently function in rice (Opassiri *et al.*, 2006). To date, only a few rice β -glucosidase isozymes have been characterized for the possible function. Partially purified β -glucosidases from rice were described that hydrolyze gibberillin glucosides (Schlieman, 1984), pyridoxine glucosides (Iwami and Yasumoto, 1986), pantoic acid, and *R*(-) pantoyllactone- β -D-glucopyranoside (Menegus *et al.*, 1995). Akiyama *et al.* (1998) determined the N-terminal sequence of the cell wall-bound β -glucosidase that preferentially hydrolyze cello- and laminari-oligosaccharides. Opassiri *et al.* (2003) cloned the *bglu1* and *bglu2* cDNA from rice seedlings and found they are highly expressed in germinating shoots, and the *bglu1* gene is also highly expressed in flower. The recombinant rice BGlu1 enzyme hydrolyzes a variety of *p*-nitrophenyl β -D-glycosides, natural glycosides at low levels, and showed strong hydrolysis and glucotransferase activity with β -1,3- and β -1,4-linked glucooligosaccharides.

Recently, the cDNA of *Os4bglu12* GH1 β -glucosidase was cloned from rice seedlings by RT-PCR and the recombinant thioredoxin-Os4BGlu12 fusion protein was functionally expressed in *E. coli* (Opassiri *et al.*, 2006). The *Os4bglu12* cDNA encoded a 510 amino acid long precursor protein. The deduced Os4bglu12 N-terminal amino acid sequence was identical to the N-terminal amino acid sequence of the previously purified cell-wall-bound rice β -glucosidase at 40 of 44 residues (Akiyama

et al., 1998). Northern blot analysis revealed that *Os4bglu12* is highly expressed in rice seedling shoot and in the leaf sheath and stem of rice at flowering stage. *Os4bglu12* transcript levels increased in rice seedlings in response to wounding, methyl jasmonate and ethephon (Maneesan, 2007).

The Os4BGlu12 enzyme efficiently hydrolyzed β -(1,4)-linked oligosaccharides of 2-6 glucose residues. Among β -1,3-linked laminari-oligosaccharides, Os4BGlu12 hydrolyzed only laminaribiose (Opassiri *et al.*, 2006). This likely reflects the bent shape of oligosaccharides with the β -1,3-linkage, which is somehow incompatible with the active site for longer chains. Among the artificial *p*-nitrophenyl (*p*NP)-glycosides, Os4BGlu12 hydrolyzed *p*NP- β -D-glucoside and *p*NP- β -D-fucoside with relatively high efficiency. It hydrolyzed *p*NP- β -D-galactoside, and *p*NP- β -D-xyloside and *p*NP- α -L-arabionoside, at 45%, 45% and 26% the rate of *p*NPG, respectively. Hydrolysis of *p*NP- β -D-mannoside, *p*NP- β -D-cellobioside, *p*NP- α -D-glucoside, and *p*NP- β -L-fucoside was not detectable. In addition, this enzyme hydrolyzed cell wall oligosaccharides released from 1,3;1,4- β -glucans of rice cell wall by rice endo-1,3;1,4- β -glucanase (Maneesan, 2007). It could also hydrolyze compounds extracted with methanol from 7-day-old rice seedlings and rice plants at flowering stage (Maneesan, 2007).

The substrate preference of Os4BGlu12 is somewhat similar to rice BGlu1, Os3BGlu7 which has 56% amino acid sequence identity with Os4BGlu12. Rice Os3BGlu7 could also hydrolyze short β -(1,3)-linked and β -(1,4)-linked gluco-oligosaccharides and various natural substrates with low activity (Opassiri *et al.*,

2004). However, BGlu1 could hydrolyze *p*NP- β -mannoside, which Os4BGlu12 could not.

The Os4BGlu12 polypeptide contains Glu residues at positions 203 and 417 that lie within the sequences TFNEP and ITENG, respectively, which match the consensus motifs in GH1 β -glucosidases (Pomthong, 2008). The Os4BGlu12 protein sequence had the highest sequence identity (68%) with *Leucaena leucocephala* β -glucosidase (AC ABY48758) (Pomthong, 2008). It also had sequence identity ranging from 60-64% with other β -glucosidases of many plant species, such as isoflavone β -glucosidases of *Medicago truncatula* (AC ABW76288), *Glycine max* (AC BAF34333), *Dalbergia nigrescens* (AC AAV34606) and *Camellia sinensis* (AC BAC78656), prunasin hydrolase (AC AAF34650) and amygdalin hydrolase (AC AAA93234) of *Prunus serotina* (Pomthong 2008). The sequence alignment between Os4BGlu12 with plant enzymes listed above (Figure 1.3) revealed that rice Os4BGlu12 contains the conserved glucose binding residues found in other GH1 β -glucosidases including Gln38, His142, Glu191, Glu406, Glu464 and Trp465 in maize Glu1 (Czjzek *et al.*, 2000). Pomthong (2008) reported that Os4BGlu12 sequence is somewhat more similar to isoflavone β -glucosidases than rice BGlu1 and barley β II β -glucosidases/exoglucanases. In addition, Os4BGlu12 has some amino acid residues similar to the residues lining the active site of the white clover linamarase structure, 1CBG (Barrett *et al.*, 1995), such as Gln33, His137, Asn190, Phe197, Trp369, Trp446, Glu453 and Trp454 in 1CBG.

```

Os4bglu12 : ---MAAAGAMPGGLLLTFL--LAVVASGAYNS-----ACEPVVSRFSF---PKGTFVGTASASYQYEGAAAGCGRGPSTWD : 69
LCBG      : -----FKPLPISF-----DDFSDLMRSCF---APGTFVGTASASYQYEGAAAGCGRGPSTWD : 49
Medicago : -----MLLSLSIVV--THIDAIKPLHL-----QFSDFMRTSF---PPGTFVGTASASYQYEGAVRECGRGPSTWD : 62
Leucaena  : -----MMKVMVVAVVV--ALITVAAADAT-----NDISLSRFSF---APAFVGTASASYQYEGAAAGCGRGPSTWD : 65
Prunus    : MATKLGSLLLCALLLAGFAL--TNSKAAKTDPP-----IHCASLNRSSDPALEPGTFVGTASASYQYEGAAAGCGRGPSTWD : 75
Camellia  : -----MMAAKGSVVVGLAIVA--YALVVSEVAIA-----AQISSFMRTSF---PDGTFVGTASASYQYEGAAAGCGRGPSTWD : 69
Glycine   : ---MDSNGYLVVGVVAFALPCSFVRSVLTDSVPLFSVPHDAASLTRNSF---PAGTFVGTASASYQYEGAAAGCGRGPSTWD : 75
Dalbergia : ---MIAMTFKVLILLGLLALISTSTSIAPFKVRAITIEVPPFMRSF---PSDTFVGTASASYQYEG---E---GRWPSIWD : 71
barley    : -----DGNPNMPE-----IGMTGGLSRGTF---PAGTFVGTASASYQYEGAAAGCGRGPSTWD : 50
Bglul     : -----W-----LGGLSRAAF---PKRFVGTASASYQYEGAAAGCGRGPSTWD : 45
maizel    : -----SARVGSQNGVQLSPS-----EIPORWF---PSDTFVGTASASYQYEGAAAGCGRGPSTWD : 54

```

```

Os4bglu12 : TTFTHQHEKIRDRSNGDVAADSYHRYKEDVRLKRMGMDAYRFSISWRILPFGSLRGVWKEGIRVYVNLINELLSKGVQP : 151
LCBG      : TTFTHQHEKIRDRSNGDVAADSYHRYKEDVRLKRMGMDAYRFSISWRILPFGSLRGVWKEGIRVYVNLINELLSKGVQP : 131
Medicago : TTFTHQHEKIRDRSNGDVAADSYHRYKEDVRLKRMGMDAYRFSISWRILPFGSLRGVWKEGIRVYVNLINELLSKGVQP : 144
Leucaena  : TTFTHQHEKIRDRSNGDVAADSYHRYKEDVRLKRMGMDAYRFSISWRILPFGSLRGVWKEGIRVYVNLINELLSKGVQP : 147
Prunus    : TTYTHHSEKIRDRSNGDVAADSYHRYKEDVRLKRMGMDAYRFSISWRILPFGSLRGVWKEGIRVYVNLINELLSKGVQP : 157
Camellia  : TTFTHQHEKIRDRSNGDVAADSYHRYKEDVRLKRMGMDAYRFSISWRILPFGSLRGVWKEGIRVYVNLINELLSKGVQP : 151
Glycine   : TTFTHQHEKIRDRSNGDVAADSYHRYKEDVRLKRMGMDAYRFSISWRILPFGSLRGVWKEGIRVYVNLINELLSKGVQP : 157
Dalbergia : NTFTHQHEKIRDRSNGDVAADSYHRYKEDVRLKRMGMDAYRFSISWRILPFGSLRGVWKEGIRVYVNLINELLSKGVQP : 153
barley    : AFWA--IQGMAAGNCTADVTVDEYHRYKEDVRLKRMGMDAYRFSISWRIFPDG--TQVMQGEVQVYVNLINELLSKGVQP : 129
Bglul     : AFAA--TFGVAQNQDVAADSYHRYKEDVRLKRMGMDAYRFSISWRIFPDG--EGRVMQGEVQVYVNLINELLSKGVQP : 124
maizel    : HGHNHSEKIRDRSNGDVAADSYHRYKEDVRLKRMGMDAYRFSISWRILPFGSLRGVWKEGIRVYVNLINELLSKGVQP : 136

```

```

Os4bglu12 : FVILFHMDPQALEDEYGFLLSP---NIDDFROYACCFRFGDRVKHMTDNEPWSYSGGADAGLLAPRCSAFMAF-C : 230
LCBG      : VVILFHMDPQALEDEYGFLLSP---NIDDFROYACCFRFGDRVKHMTDNEPWSYSGGADAGLLAPRCSAFMAF-C : 210
Medicago : VVILFHMDPQALEDEYGFLLSP---RIDDFROYACCFRFGDRVKHMTDNEPWSYSGGADAGLLAPRCSAFMAF-C : 223
Leucaena  : FVILFHMDPQALEDEYGFLLSP---DIDDFROYACCFRFGDRVKHMTDNEPWSYSGGADAGLLAPRCSAFMAF-C : 226
Prunus    : FVILFHMDPQALEDEYGFLLSP---NIDDFROYACCFRFGDRVKHMTDNEPWSYSGGADAGLLAPRCSAFMAF-C : 236
Camellia  : FVILFHMDPQALEDEYGFLLSP---HIDDFROYACCFRFGDRVKHMTDNEPWSYSGGADAGLLAPRCSAFMAF-C : 229
Glycine   : VVILFHMDPQALEDEYGFLLSP---HIDDFROYACCFRFGDRVKHMTDNEPWSYSGGADAGLLAPRCSAFMAF-C : 233
Dalbergia : VVILFHMDPQALEDEYGFLLSP---RVMDDFROYACCFRFGDRVKHMTDNEPWSYSGGADAGLLAPRCSAFMAF-C : 232
barley    : VAMLYHYDPLALEKRYGSLMA---KQADLRTYADFCFRFGDRVKHMTDNEPWSYSGGADAGLLAPRCSAFMAF-C : 204
Bglul     : VVILFHMDPQALEKRYGSLMA---KQADLRTYADFCFRFGDRVKHMTDNEPWSYSGGADAGLLAPRCSAFMAF-C : 199
maizel    : VVILFHMDPQALEKRYGSLMA---KQADLRTYADFCFRFGDRVKHMTDNEPWSYSGGADAGLLAPRCSAFMAF-C : 218

```

```

Os4bglu12 : SVGDSGREPVTACHHQLLAHAETVRYVYKQYALKQKIGITVSHVFFVFR-SKSNDAARRAIDRQYGYMDFLTRGEV : 311
LCBG      : TCGDSGREPVTAAHYQLLAHAARVRYVYKQYASQKIGITVSHVFFVFR-EKADVDAARRAIDRQYGYMDFLTRGEV : 291
Medicago : TCGDSGREPVTAAHYQLLAHAARVRYVYKQYASQKIGITVSHVFFVFR-AKSQVDAARRAIDRQYGYMDFLTRGEV : 304
Leucaena  : TCGDSGREPVTAAHYQLLAHAARVRYVYKQYASQKIGITVSHVFFVFR-KKSQVDAARRAIDRQYGYMDFLTRGEV : 307
Prunus    : TCGDSGREPVTAAHYQLLAHAARVRYVYKQYASQKIGITVSHVFFVFR-AEEDVDAARRAIDRQYGYMDFLTRGEV : 317
Camellia  : PKGSGTPEPVTACHHQLLAHAARVRYVYKQYALKQKIGITVSHVFFVFR-SKADKDAARRAIDRQYGYMDFLTRGEV : 310
Glycine   : LCGDAGTPEPVTACHHQLLAHAARVRYVYKQYALKQKIGITVSHVFFVFR-LAENSTSDAARRAIDRQYGYMDFLTRGEV : 315
Dalbergia : TCGDAGTPEPVTACHHQLLAHAARVRYVYKQYALKQKIGITVSHVFFVFR-VLQSN-STSOKRAARRAIDRQYGYMDFLTRGEV : 313
barley    : ACGDSGREPVTACHHQLLAHAARVRYVYKQYALKQKIGITVSHVFFVFR-LDADQAAARRAIDRQYGYMDFLTRGEV : 285
Bglul     : ACGDSGREPVTACHHQLLAHAARVRYVYKQYALKQKIGITVSHVFFVFR-LDADQAAARRAIDRQYGYMDFLTRGEV : 280
maizel    : PTCNSLVPEPVTACHHQLLAHAARVRYVYKQYALKQKIGITVSHVFFVFR-SFLQKDAARRAIDRQYGYMDFLTRGEV : 298

```

```

Os4bglu12 : FLSMRSLGKRLPKFTEKSKLWCSFDFGLNMYYSYAAKAPRPNARPAIQDLSLN--ATFEHCKPLGMAASMLC : 371
LCBG      : FLSMRSLGKRLPKFTEKSKLWCSFDFGLNMYYSYAAKAPRPNARPAIQDLSLN--ATFEHCKPLGMAASMLC : 371
Medicago : FLSMRSLGKRLPKFTEKSKLWCSFDFGLNMYYSYAAKAPRPNARPAIQDLSLN--ATFEHCKPLGMAASMLC : 384
Leucaena  : FLSMRSLGKRLPKFTEKSKLWCSFDFGLNMYYSYAAKAPRPNARPAIQDLSLN--ATFEHCKPLGMAASMLC : 387
Prunus    : PHLMSRSLGKRLPKFTEKSKLWCSFDFGLNMYYSYAAKAPRPNARPAIQDLSLN--ATFEHCKPLGMAASMLC : 397
Camellia  : FLSMRSLGKRLPKFTEKSKLWCSFDFGLNMYYSYAAKAPRPNARPAIQDLSLN--ATFEHCKPLGMAASMLC : 390
Glycine   : FLSMRSLGKRLPKFTEKSKLWCSFDFGLNMYYSYAAKAPRPNARPAIQDLSLN--ATFEHCKPLGMAASMLC : 394
Dalbergia : FLSMRSLGKRLPKFTEKSKLWCSFDFGLNMYYSYAAKAPRPNARPAIQDLSLN--ATFEHCKPLGMAASMLC : 393
barley    : FLSMRSLGKRLPKFTEKSKLWCSFDFGLNMYYSYAAKAPRPNARPAIQDLSLN--ATFEHCKPLGMAASMLC : 365
Bglul     : FLSMRSLGKRLPKFTEKSKLWCSFDFGLNMYYSYAAKAPRPNARPAIQDLSLN--ATFEHCKPLGMAASMLC : 360
maizel    : FLSMRSLGKRLPKFTEKSKLWCSFDFGLNMYYSYAAKAPRPNARPAIQDLSLN--ATFEHCKPLGMAASMLC : 380

```

```

Os4bglu12 : VYVPGFRDILLYVKNYGNPPIVITENGDEPMDPK--LSLEESLDITRIDVYFRHLYYQSLIRDCANVAGYFAWSLDD : 471
LCBG      : VYVPGFRDILLYVKNYGNPPIVITENGDEPMDPK--LSLEESLDITRIDVYFRHLYYQSLIRDCANVAGYFAWSLDD : 451
Medicago : VYVPGFRDILLYVKNYGNPPIVITENGDEPMDPK--LSLEESLDITRIDVYFRHLYYQSLIRDCANVAGYFAWSLDD : 464
Leucaena  : VYVPGFRDILLYVKNYGNPPIVITENGDEPMDPK--LSLEESLDITRIDVYFRHLYYQSLIRDCANVAGYFAWSLDD : 467
Prunus    : VYVPGFRDILLYVKNYGNPPIVITENGDEPMDPK--LSLEESLDITRIDVYFRHLYYQSLIRDCANVAGYFAWSLDD : 477
Camellia  : VYVPGFRDILLYVKNYGNPPIVITENGDEPMDPK--LSLEESLDITRIDVYFRHLYYQSLIRDCANVAGYFAWSLDD : 468
Glycine   : VYVPGFRDILLYVKNYGNPPIVITENGDEPMDPK--LSLEESLDITRIDVYFRHLYYQSLIRDCANVAGYFAWSLDD : 474
Dalbergia : VYVPGFRDILLYVKNYGNPPIVITENGDEPMDPK--LSLEESLDITRIDVYFRHLYYQSLIRDCANVAGYFAWSLDD : 473
barley    : VYVPGFRDILLYVKNYGNPPIVITENGDEPMDPK--LSLEESLDITRIDVYFRHLYYQSLIRDCANVAGYFAWSLDD : 443
Bglul     : VYVPGFRDILLYVKNYGNPPIVITENGDEPMDPK--LSLEESLDITRIDVYFRHLYYQSLIRDCANVAGYFAWSLDD : 438
maizel    : VYVPGFRDILLYVKNYGNPPIVITENGDEPMDPK--LSLEESLDITRIDVYFRHLYYQSLIRDCANVAGYFAWSLDD : 462

```

```

Os4bglu12 : FEWSGYIVRFGIMFVDMGRRYPRKSAHNFKFLK----- : 510
LCBG      : FEWSGYIVRFGIMFVDMGRRYPRKSAHNFKFLK----- : 491
Medicago : FEWSGYIVRFGIMFVDMGRRYPRKSAHNFKFLK----- : 504
Leucaena  : FEWSGYIVRFGIMFVDMGRRYPRKSAHNFKFLK----- : 507
Prunus    : FEWSGYIVRFGIMFVDMGRRYPRKSAHNFKFLK----- : 553
Camellia  : FEWSGYIVRFGIMFVDMGRRYPRKSAHNFKFLK----- : 507
Glycine   : FEWSGYIVRFGIMFVDMGRRYPRKSAHNFKFLK----- : 514
Dalbergia : FEWSGYIVRFGIMFVDMGRRYPRKSAHNFKFLK----- : 531
barley    : FEWSGYIVRFGIMFVDMGRRYPRKSAHNFKFLK----- : 485
Bglul     : FEWSGYIVRFGIMFVDMGRRYPRKSAHNFKFLK----- : 476
maizel    : FEWSGYIVRFGIMFVDMGRRYPRKSAHNFKFLK----- : 512

```

Figure 1.3 Alignment of the predicted mature protein sequences of rice Os4BGlu12 with related β -glucosidases (Pomthong, 2008).

The rice cDNA derived sequence is labeled as Os4BGlu12; BGlu1 is rice β -glucosidase 1 (AC U28047 also called Os3BGlu7); maize1 is maize β -glucosidase 1 (accession no. U33816); barley is deduced from barley BGQ60 nucleotide sequence (AC AAA87339), which was previously determined to be the same as β II β -glucosidase (Hrmova *et al.*, 1998; Leah *et al.*, 1995); 1CBG is white clover linamarase (P26205); Medicago is *Medicago truncatula* isoflavone β -glucosidase (AC ABW76288) (Naoumkina *et al.*, 2007); Leucaena is *Leucaena leucocephala* β -glucosidase (AC ABY48758); Prunus is *Prunus serotina* amygdalin hydrolase (AC AAA93234); Camellia is *Camellia sinensis* isoflavone β -glucosidase (AC BAC78656); Glycine is *Glycine max* isoflavone β -glucosidase (AC BAF34333); and Dalbergia is *Dalbergia nigrescens* isoflavone β -glucosidase (AC AAV34606). Sequences corresponding to the catalytic acid/base and nucleophile consensus sequences are marked by a thick line over the top. Residues shown by Czjzek *et al.* (2000) to be in contact with the DIMBOA aglycone in the maize β -glucosidase are indicated by: ▼, while those conserved residues making contacts with the sugar are marked by: Δ above the column. Residues lining the active site of the 1CBG structure are marked with a “<” or “>” under their column in the alignment for residues appearing to reside inside the two catalytic carboxylic acids and those appearing to be outside these residues, respectively (Barrett *et al.*, 1995; Opassiri *et al.*, 2004). The alignment was generated using the Clustal X implementation of Clustal W (Jeanmougin *et al.*, 1998; Thompson *et al.*, 1994), analyzed and manually adjusted by Genedoc (<http://www.psc.edu/biomed/genedoc/>).

In order to understand the enzyme-substrate binding mechanism leading to the substrate preferences of this enzyme and the basis for the different substrate specificity among GH1 enzymes, the three-dimension structure of Os4BGlu12 has been studied. This study would help to provide information to further clarify the enzyme-substrate binding mechanism by its relationship to the active site architecture of the enzyme.

1.8 Research Objectives

1. To optimize recombinant expression and purification conditions to obtain high quality and quantity of rice Os4BGlu12 β -glucosidase for protein crystallization.
2. To screen for the conditions for crystallization of rice Os4BGlu12 β -glucosidase and find optimal conditions for production of diffracting crystals.
3. To study the preliminary three-dimension structure of rice Os4BGlu12 β -glucosidase.

Purchased from Merck (Darmstadt, Germany). Sodium acetate, sodium hydroxide, sodium chloride, sodium carbonate, disodium ethylenediamine tetraacetate (EDTA), piperazine-1,4-bis(2-ethanesulfonic acid) (PIPES), bromophenol blue, methanol, HPLC-grade distilled water, ethanol and glacial acetic acid were purchased from Carlo ERBA (Rodano, Milano, Italy). Bactotrypton, yeast extract and bacto agar were purchased from Difco (BD, NJ, USA). Imidazole, Coomassie Brilliant Blue R250, phenylmethylsulfonyl fluoride (PMSF), calcium chloride, bovine serum albumin (BSA) and dichlorodimethylsilane were purchased from Fluka (Steinheim, Switzerland). Ammonium persulfate, acrylamide, *N,N',N'',N'''*-tetramethylethylenediamine (TEMED), Sodium dodecyl sulfate (SDS), *N,N'*-methylene-bis-acrylamide, Triton X-100, and lysozyme were purchased from Amersham Pharmacia Biotech (Uppsala, Sweden). Isopropyl thio- β -D-galactoside (IPTG), kanamycin, tetracycline, DNase I, 2-deoxy-2-fluoro- β -D-glucoside, Tris base, 2-mercaptoethanol, *p*-nitrophenol β -D-glucoside (*p*NPG) were purchased from Sigma (St. Louis, MO, USA). Cellooligosaccharides of degree of polymerization (DP) 3-6 were purchased from Seikagaku Kogyo Co. (Tokyo, Japan).

2.2 General Methods

2.2.1 Transformation of recombinant expression plasmid into expression host cells

The first expression experiment was initially started with a fresh transformation of recombinant pET32(a+)/DEST-*Os4bglu12* to the expression host cells. The recombinant plasmids were transferred into OrigamiB(DE3) competent cells. One microliter (100 ng) of plasmid was added to 50 μ L thawed OrigamiB(DE3)

competent cells, mixed gently by swirling and the tube stored on ice for 20 min. The mixed cells were heat shocked at 42°C for 60 sec and the tube rapidly transferred to ice for 2 min. The transformed cells were grown by adding 450 µL of LB broth and they were grown at 37°C for 45 min. Then, 200 µL of transformation medium was spread directly onto an LB agar plate containing 50 µg/mL of ampicillin, 15 µg/mL kanamycin and 12.5 µg/mL tetracycline, and incubated overnight at 37°C.

2.2.2 Expression of recombinant β -glucosidase was produced as thioredoxin-Os4BGlu12 fusion protein in *E. coli* system

The positive colony was inoculated into 50 mL LB medium containing 50 µg/mL ampicillin, 15 µg/mL kanamycin and 12.5 µg/mL tetracycline and cultured at 37°C on an incubator shaker for 16 hr. Eight milliliters of starter culture was then transferred into four 2-liter Erlenmeyer flasks containing 800 mL LB medium with the same antibiotics as above and incubated at 37°C until the optical density at 600 nm reached 0.5-0.6. Then, IPTG was added to a final concentration of 0.4 mM to induce the expression of the target proteins and the culture was grown for a further for 16 hr at 20°C. The culture broth was centrifuged at 4°C for 15 min at 3900 xg. The cells pellets were kept at -80°C.

2.2.3 Protein extraction from *E. coli*

Frozen bacterial cells pellets was thawed, and then resuspended in 5 mL per gram cells with freshly prepared extraction buffer containing 50 mM sodium phosphate buffer, pH 7.2, 200 µg/mL lysozyme, 1% Triton-X 100, 1 mM phenylmethylsulfonylfluoride (PMSF) and 0.25 µg/mL DNase1. The cell suspension was then incubated at room temperature for 30 min. The supernatant containing

soluble protein was separated from the cell debris by centrifugation at 14000 xg for 15 min at 4°C and the protein solution was stored at 4°C.

2.2.4 Purification of recombinant thioredoxin-Os4BGlu12 fusion protein

The fusion protein was purified by immobilized metal affinity chromatography (IMAC) on Co²⁺ (Talon resin, Clontech, Palo Alto, CA, USA) at 4°C. In each round of protein purification, fifteen-milliliters of soluble protein extract was loaded onto a 3 mL Co²⁺ IMAC column, which was pre-equilibrated with equilibration buffer (20 mM Tris-HCl, pH 7.2, and 300 mM NaCl). Next, the column was washed with 5 column volumes of equilibration buffer and 5 column volumes of each wash buffer (5 and 10 mM imidazole in equilibration buffer). The bound protein was eluted with 5 column volumes of 250 mM imidazole in equilibration buffer. All fractions of purification steps were assayed for β-glucosidase activity with 1 mM pNP-β-D-glucoside as described by Opassiri *et al.*, (2003). Then the pure protein was pooled and concentrated in a 30 kDa molecular weight cut-off (MWCO) Centricon centrifugal filter (Milipore, Massachusetts, USA) and the buffer was changed to 20 mM Tris/HCl, pH 7.2 in the same device.

2.2.5 Purification of Os4BGlu12 nonfusion protein

Since the thioredoxin and histidine fusion tags might interfere with crystallization, only the non-fusion protein was used for crystallization. To remove the N-terminal fusion tag, Os4BGlu12 fusion protein was cleaved with 0.1 μg enterokinase (New England Biolabs) per 100 μg protein in 20 mM Tris-HCl, pH 7.2

at 23°C for 16 hr. The tag-free protein was separated from the fusion tag and uncleaved fusion protein by loading the digest onto a Co²⁺ IMAC column (1 mL resin per 1 miligram protein). The unbound protein fraction was collected and the resin was washed with 5 and 10 mM imidazole in equilibration buffer (20mM Tris-HCl, pH 7.2, and 300 mM NaCl). The unbound and wash fractions containing tag-free protein were combined. The protein sample was concentrated and the buffer was changed to 20 mM Tris-HCl, pH 7.2, in a 10 kDa MWCO Centricon by centrifugation at 2800 xg at 4°C. Finally, the tag-free protein was further purified by Sephacryl S200 chromatography on an ÄKTA purifier system (GE Healthcare, Sweden). The protein was equilibrated and eluted with 150 mM NaCl, 20 mM Tris-HCl, pH 7.2, at flow rate of 0.25 ml/min. Fractions containing β -glucosidase activity were pooled.

2.2.6 Protein analysis

2.2.6.1 SDS-PAGE electrophoresis

The SDS-PAGE was executed by the method of Laemmli (Laemmli, 1970). The 12% separating gel was prepared for 2 mini-gels containing 4.0 mL of 30% acrylamide mixed, 2.5 mL of 1.5 M Tris, pH 8.8, 3.345 mL of distilled water, 0.1 mL of 10% SDS, 0.050 mL of 10% ammonium persulfate and 0.005 mL of TEMED. The 4% stacking gel contained 0.670 mL of 30% acrylamide mixed, 0.63 mL of 1.5 M Tris, pH 8.8, 3.0 mL of distilled water, 0.050 mL of 10% SDS, 0.0025 mL of 10% ammonium persulfate 0.005 mL of and TEMED. The two gel polymerizations were done in a Mini-PROTEAN 3 cell (BioRAD). The protein sample was mixed with 1X SDS-PAGE loading buffer (50 mM Tris-HCl, pH 6.8, 10% glycerol, 2% (w/v) SDS, 14.4 mM 2-mercaptoethanol and 0.1% bromophenol

blue) and then boiled at 100°C for 5 minute; a 10-15 µL aliquot was loaded onto the gel, and then electrophoresed using running buffer (50 mM Tris base, 250 mM glycine and 0.1% SDS). A constant voltage of 150 V was applied to the electrophoresis until the loading buffer migrated the appropriate distance through the gel (approximately 45 minute). Protein bands were then examined by staining of this gel in Coomassie Brilliant Blue R-250 staining solution (0.1% Coomassie Brilliant Blu R-250, 45% (v/v) methanol, 45% (v/v) distilled water and 10% (v/v) glacial acetic acid) for 1 hour at room temperature, and then destained with destaining solution (45% (v/v) methanol, 45% (v/v) distilled water and 10% (v/v) glacial acetic acid) until excess of dye was removed. The size of protein was estimated by comparing its migration with those of the protein standards in the low molecular weight protein marker (Biomann Scientific, Washington DC, USA).

2.2.6.2 Enzyme activity

The activity of Os4BGlu12 was determined according to the method of Opassiri *et al.*, (2003). Activity toward *p*NP-β-D-glucoside (*p*NPG) was assayed in a reaction mixture containing 90 µL of 50 mM sodium acetate, pH 5.0, 5 µL of 10 mM *p*NPG and 5 µL of crude enzyme. The reaction was incubated at 37°C for 10 minute, and then 50 µL of 0.4 M sodium carbonate was added to stop reaction. The free *p*-nitrophenolate group was measured with iEMS reader MF microplate photometer (Labsystem iMES MF, Finland) at 405 nm. The experiment was controlled by substrate blank (reaction without enzyme).

2.2.6.3 Bio-Rad protein assay

Protein concentrations in the crude extracts and fractions and pools from the purification steps were measured by the method of Bradford (Bradford *et al.*, 1976) with a Bio-Rad kit (Hercules, CA, USA) using bovine serum albumin (BSA) as a standard (0-5 µg). The assay solution contained suitably diluted enzyme in distilled water in a total volume of 800 µl. The solution was mixed and incubated at room temperature for 10 minutes, and then the absorbance was measured at 595 nm with a Genesys 10UV spectrophotometer (Spectronic Instruments, Rochester, NY, USA).

2.3 Protein crystallization

2.3.1 Initial screening for crystallization conditions

The purified Os4BGlu12 protein for crystallization screening was adjusted to 10 mg/mL, in 20 mM Tris/HCl pH 7.2 in a 10 kDa MWCO Centricon by centrifugation at 2800 xg, 4°C and the protein was filtered with 0.2 µm centrifugal filter (Milipore) around 8900 xg for 5 minute. Microbatch screening (Chayen *et al.*, 1992) was applied for screening of crystallization conditions with Sigma Crystallization Basic Kit, JBScreen HTS I and II (Jena Bioscience, Jena, Germany), Crystals Screen High Throughput HR2-130 (Hampton Research, Aliso Viejo, CA, USA), JY Screen screening kit (Mahidol University, Thailand) and Emerald Biosystems (I and II) (Emerald BioSystems, Bainbridge, Island). The initial screening as performed with 60 well plastic plates (Nunclon, Denmark). Firstly, approximately 20 µL of 100% paraffin oil was pipetted into each well to encapsulate and prevent the evaporation of the aqueous drops. Then a 0.5 µL aqueous precipitant cocktail solution was added to each well under the oil. Finally, 0.5 or 1 µL of 10 mg/mL protein

solution was added under the oil, and the aqueous drops were allowed to merge in the hydrophobic environment. A single drop of protein and precipitant should occur spontaneously (Figure 2.2). If this was not the case, a cat whisker was used to mix the drops together under the oil. The plates with crystallization trials were covered with the plate cover and placed on a moist sponge in a plastic box and incubated at 15°C. The crystals trials were observed with a Stemi 2000-C stereomicroscope (Zeiss Crop, Göttingen, Germany).

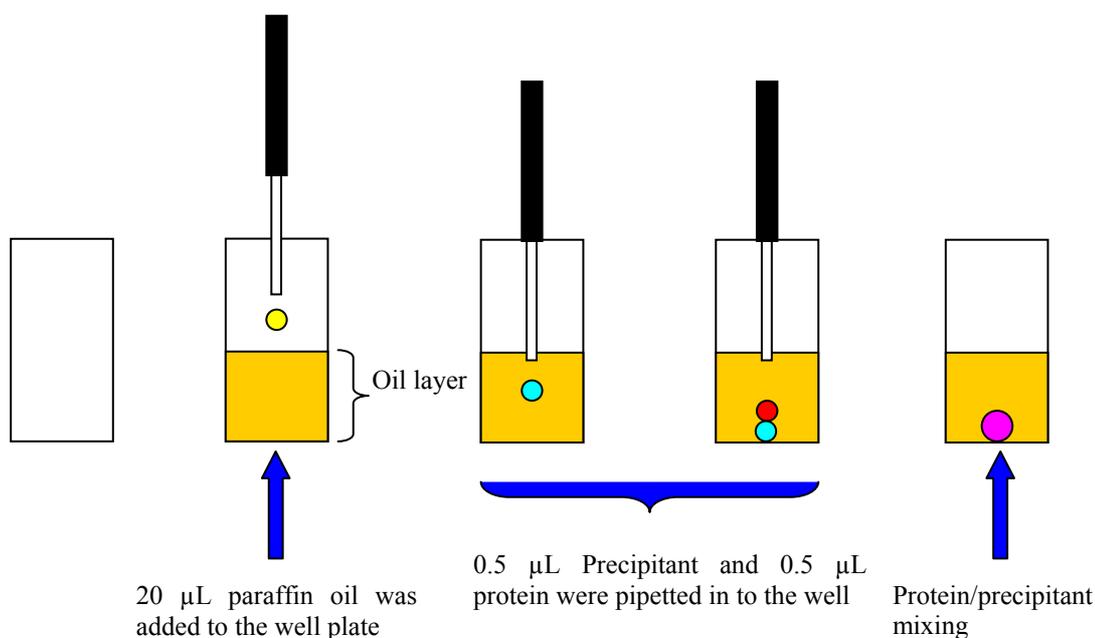


Figure 2.2 Schematic diagrams of microbatch crystallization trials.

2.3.2 Optimization of crystallization conditions by vapour diffusion method

The conditions that yielded small crystals in the initial screening were optimized by hanging drop vapour diffusion (Weber, 1997) to attain sufficient size

crystals suitable for X-ray diffraction. Various factors were adjusted for optimization. Primary variables were the concentrations of PEG, salt and protein, while the secondary any variable was the molecular weight of PEG. The variable table of the optimization is presented in the Figure 2.3. The crystallization trial was performed in 24 well plates (Greiner Bio-One, Frickenhausen, Germany). Glass cover slips (Menzel-Glaser, Braunschweig, Germany) were siliconized to reduce the contact area of the drop protein with the glass surface of the cover slip. For Siliconization, the glass cover slips were washed with the ratio of 1 part of 0.1 M HCl and 3 part of absolute methanol for 30 min, then put in deionized water three time. Finally, 3 mL of Dimethyldichlorosilane was added and the glass slips were put in the oven overnight. Then high vacuum grease (Dow Corning, Michigan, USA) was the top edge of each well. A hanging drop of 2 μ L of 0.5-9 mg/mL pure Os4BGlu12 was mixed with 1 μ L of precipitant solution on the center of a siliconized cover slip. The cover slip with the drop was inverted and placed on to the top of the well containing 0.5 mL of precipitant solution, and the grease seal on the top was twisted a few degrees to complete sealing. The drop was equilibrated against a reservoir of precipitant solution at 15°C.

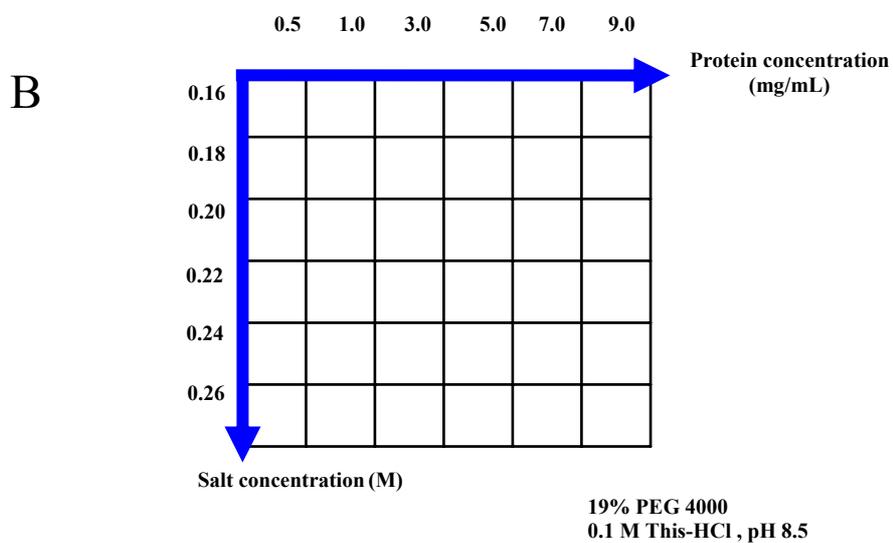
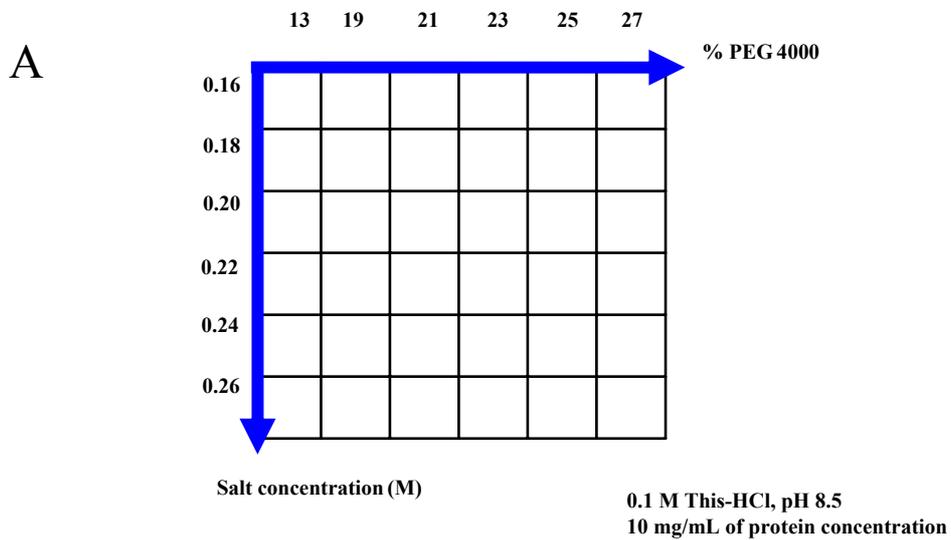


Figure 2.3 The strategy for optimization of Os4BGlu12 crystallization.

A and B, the grid screen of the variable concentrations of salt, %PEG and protein in fixed concentration of 0.1 M Tris-HCl, pH 8.5. C, the grid screen of the variable concentrations of salt and molecular weights of PEG at fixed pH, % PEG and protein concentration.

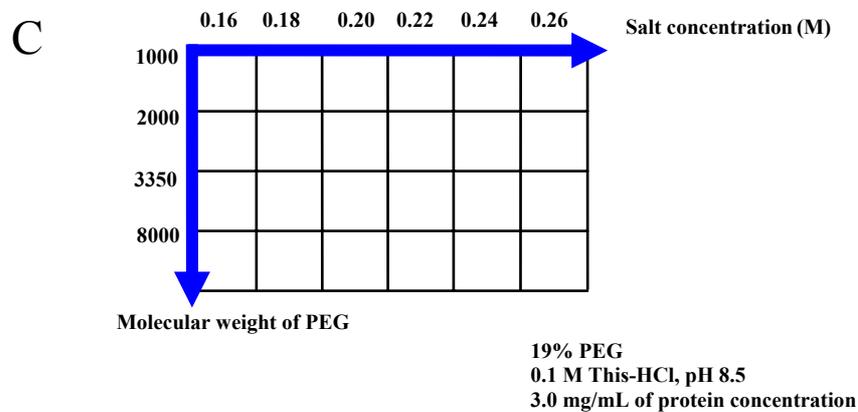


Figure 2.3 (Continued)

2.3.3 Microseeding method

The microseeding technique was performed to produce the large and single crystals. The small crystals or cluster crystals obtained from the first optimization conditions were used as seed stock for microseeding. The crystals were removed from their initial drop, and then transferred to 10 μ L of reservoir solution. The crystal was crushed with a cat whisker, and then the suspension was added to a 1.5 mL microcentrifuge tube containing 100 μ L of reservoir solution. The tube was centrifuged for 5 min at 3400 xg , then the supernatant containing the microcrystal stock was serially diluted with mother liquor to 1/100, 1/500 and 1/1000 and the stock solution was stored at 4°C. Optimization of crystallization was done in 2.3.2 with an additional step of pre-equilibration of the drop as described. The protein precipitant drop was pre-equilibrated with 0.5 mL of reservoir solution for 2 hour at 15°C. Then,

the equilibrated drops were streaked with the diluted microseed suspension using a cat whisker, and then stored at 15°C.

2.3.4 Crystallization of complexes

The complex of Os4BGlu12 with (2,4-dinitrophenyl 2-deoxy-2-fluoro- β -D-glucopyranoside (G2F) was used for hanging drop vapour diffusion crystallization at 15°C. For crystallization, the purified Os4BGlu12 was concentrated to 3.0 mg/mL in 20 mM Tris-HCl, pH 8.5, and filtered with a 0.2 μ m centrifugal filter around 8900 g for 5 minute. A concentrated G2F solution was prepared at 10 mM in the HPLC grade water and stored at -20°C. 3.0 mg/mL of Os4BGlu12 protein solution was incubated for 10 minute with 10 mM G2F at 4°C. The hanging drops consisted of 2 μ L of the protein mixture solution and 1 μ L of precipitant solution. Then, the drop was equilibrated with 0.5 mL of reservoir solution, and kept at 15°C.

2.3.5 Determination that crystals were protein

SDS-PAGE was applied to determine whether Os4BGlu12 crystals were really protein or just salt crystals. Clusters and large crystals were harvested from original crystallization drops (19% PEG 4000, 0.16 M NaCl, 0.1 M Tris-HCl, pH 8.5). They were washed with 10 μ L of precipitant solution 6 times to ensure that no layer of the original protein solution was still present on the surface of the crystals. Then, the crystals were transferred again into a fresh drop of newly prepared reservoir solution. Finally, the crystals were dissolved in 20 μ L of 20 mM Tris-HCl, pH 8.5, and 1X final concentration of protein loading buffer was added and then the mixture

was boiled for 5 minute and loaded onto an SDS-PAGE gel. Electrophoresis and staining were conducted as described in section 2.2.6.1.

2.4 Data collection and processing

2.4.1 Freezing in cryoprotectant

When the crystal was frozen in liquid nitrogen or during X-ray data collection, the temperature of the system is lower than the melting point of the precipitant solution. In this case the crystal might have been damaged by ice formation. The cryosolution was used to protect crystals, by reduction of decadence or separation of the crystals. The components of the cryoprotectant solution were optimized by varying the concentration of glycerol, at 18%, 22%, 24%, 26% and 28% and the concentration of precipitant with the concentrations of its components increased by 18%, 20% and 22%. Crystals were soaked overnight in 5 μ L of different cryoprotectant solutions at room temperature and 15°C. The crystals were mounted in a nylon loop (Hampton Research, 0.1-0.2 mm) and transferred into the 5 μ L of cryoprotectant solution for 15 seconds, then flash-cooled in liquid nitrogen at -180°C for X-ray diffraction.

2.4.2 Synchrotron X-ray diffraction

The free Os4BGlu12 crystal and the complex crystals were diffracted with synchrotron radiation as the X-ray source at the BL13B1 beamline, National Synchrotron Radiation Research Center (NSRRC), Taiwan with the ADSC Quantum 315 CCD detector. The crystal with cryo-loop was placed onto the goniometer head between the X-ray beam and the detector. During X-ray diffraction, the distance

between the crystal and detector was 300.0 mm for free-Os4BGlu12 crystal and 280 mm for Os4BGlu12 complex. The X-ray wavelength was set at 1.00 Å and crystals were kept at 100 K with a nitrogen stream from an Oxford Cryosystems Cryo-stream during diffraction. A crystal was collected for 180° rotation with 0.50° oscillations, and the exposure time per frame of 15-20 second until the data were complete. All the diffraction images were indexed, integrated and scaled with the HKL-2000 program (Otwinowski and Minor, 1997).

2.5 Structure solution by molecular replacement

The crystal structure of free Os4BGlu12 was solved by the molecular replacement method, using the *MOLREP* program (Vagin and Teplyakov, 1997). The first dataset processed to 2.50 Å was solved by the molecular replacement with the cyanogenic β-glucosidase from white clover structure (PDB code 1CBG; 63% identical to Os4BGlu12) as a search model (Barrett *et al.*, 1995). The solution included two molecules per asymmetric unit, and had a solvent content of 49.98% and Mathew's coefficient (V_m) of 2.46 Å³ Da⁻¹ (Matthews, 1968). Non-crystallographic symmetry (NCS) restraints were applied to the two protein molecules in the refinement steps. The rotation function peak yielded the best translation function solution, (resolution 26.19-2.79 Å) with correlation coefficient of 52.4% and *R* factor of 49.4%. The other data set of Os4BGlu12 crystals were solved by using the structure of free Os4BGlu12 structure as a search model for rigid body refinement, including the two molecules per asymmetric unit.

2.6 Model building, structure refinement and validation of the final model

The analysis of the electron density map $F_{\text{obs}}-F_{\text{cal}}$ and $2F_{\text{obs}}-F_{\text{cal}}$, and the model building were done in the Coot program (Emsley and Cowtan, 2004). The refinement was done with REFMAC5 in the CCP4 suit, by varying of the non-crystallographic symmetry (NCS) restraints for structure refinement. The water molecule positions were checked in the refinement stages by inspecting the $2F_{\text{obs}}-F_{\text{cal}}$ electron density map for peaks at a height 1.3σ and seeing if they were within hydrogen-bonding distance of other hydrogen bonding groups. The stereochemistry of the final model was checked with PROCHECK (Laskowski *et al.*, 1993).

CHAPTER III

RESULTS

3.1 Protein expression and purification

3.1.1 Purification of recombinant Os4BGlu12

Recombinant Os4BGlu12 was expressed as an N-terminal thioredoxin and hexahistidine tag (his-tag) fusion protein in the *E. coli* strain OrigamiB(DE3) using the pET32(a+)/DEST expression vector under the optimized condition. The thioredoxin-Os4BGlu12 fusion protein contains the cleavage site for enterokinase at the N-terminus of the mature Os4BGlu12 protein. The recombinant fusion protein containing his-tag was purified by a single step of IMAC on a Co²⁺ resin column and a major band of fusion protein at 69 kDa and the nonspecifically binding and contaminating proteins were seen on the SDS-PAGE (Figure 3.1).

The yield of protein obtained from purified fractions was approximately 15 mg protein per liter of *E. coli* culture. To remove the fusion tag and thioredoxin at the N-terminus that might interfere with crystallization, the purified protein obtained from the Co²⁺-IMAC was cleaved with enterokinase to generate the Os4BGlu12 non-fusion protein. The thioredoxin fusion tag was removed from the protein solution by adsorption to Co²⁺ bound resin in the second IMAC step. The purification profile of Os4BGlu12 non-fusion protein obtained from the second IMAC is shown in Figure 3.2. The fractions 10-18 contain nonfusion protein detected at approximately 50 kDa

on SDS-PAGE with different amounts of contamination of the tag free proteins. These fractions were pooled and concentrated, and the buffer was changed to 20 mM Tris-HCl, pH 7.2. After the second IMAC step, the nonfusion Os4BGlu12 protein was obtained with >90% purity, with a yield of about 6 mg per liter of culture. The protein from this the second IMAC was used for crystallization screening and the production of sufficient crystals in the optimization experiment.

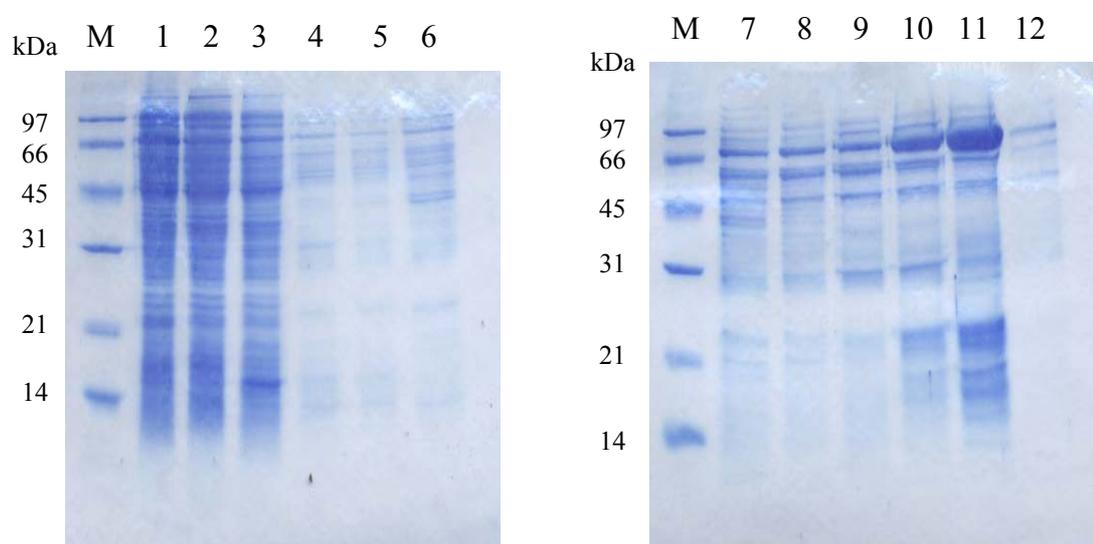


Figure 3.1 SDS-PAGE of recombinant thioredoxin-Os4BGlu12 protein purified by Co^{2+} -IMAC.

Lane M, molecular weight protein marker. Lane 1, soluble fraction of *E. coli* cells containing pET32(a+)/DEST-*Os4bglu12*. Lane 2, flowthrough. Lane 3, the protein fraction eluted in the wash with equilibration buffer. Lanes 4-6 the protein fractions eluted in the wash with 5 mM imidazole. Lanes 7-9, the protein fractions eluted in the wash with 10 mM imidazole. Lane 10-12, the elution fraction of thioredoxin-Os4BGlu12 protein eluted with 250 mM imidazole.

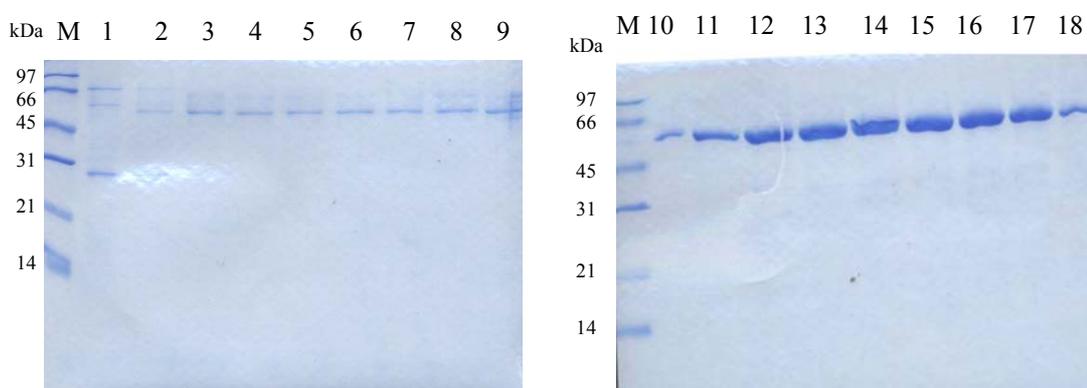


Figure 3.2 The SDS-PAGE of purified Os4BGlu12 nonfusion protein fractions from second IMAC step.

The purified fusion protein was cleaved with enterokinase and purified by Co^{2+} -IMAC to remove the fusion tag. Lane M, molecular weight protein marker. Lane 1, flowthrough. Lane 2, fraction washed from the column with equilibration buffer. Lanes 3-9, the protein fractions eluted with 5 mM imidazole, Lanes 10-17, the protein fractions eluted with 10 mM imidazole. Lane 18, the protein fraction eluted with 250 imidazole.

A further step of purification by Sephacryl S200 gel filtration chromatography was applied to eliminate the contaminating proteins for high purity of nonfusion protein. The purified protein from second IMAC was filtrated and applied onto the S200 gel filtration column. The protein fractions with high purity passed through gel filtration as a single peak on the chromatogram. The major band of nonfusion protein was observed in the fractions 30-38 (Figure 3.3). However, a trace amount of free tag protein that eluted with nonfusion protein could be observed. These eluted fractions were then pooled and concentrated and the buffer was changed

to 20 mM Tris-HCl, pH 7.2. Approximately 0.25 mg of purified protein was obtained from 1 liter of *E. coli* culture after this step of purification. Homogenous protein from S200 gel filtration was also used for crystallization experiment.

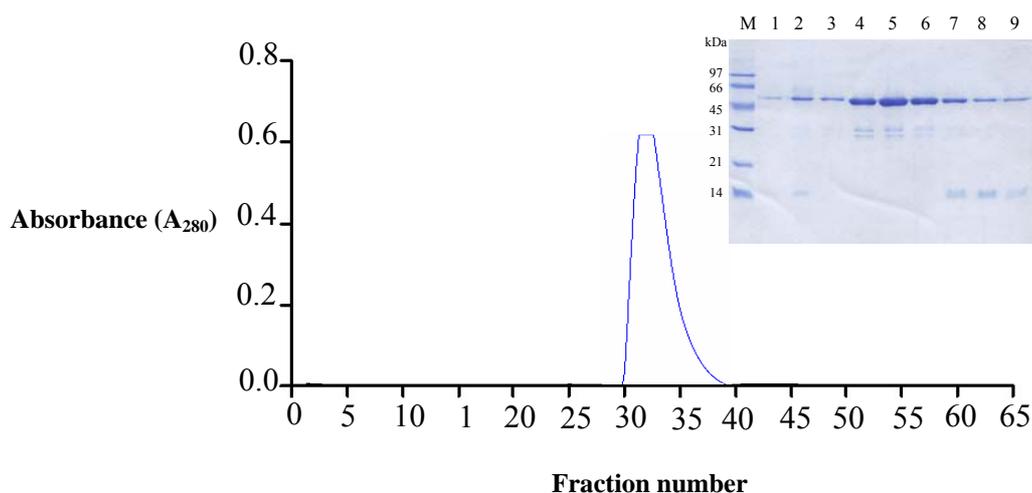


Figure 3.3 Elution profile of the Os4BGlu12 nonfusion protein obtained from Sephacryl S200 chromatography.

The separation was done in a Sephacryl S200 column on an ÄKTA purifier system. The protein was eluted with 150 mM NaCl, 20mM Tris-HCl, pH 7.2, at flow rate of 0.25 ml/min.

3.1.2 Protein purification for crystallization

The high purity protein from second IMAC column and S200 gel filtration was used in the first trial of initial crystallization screening. In the initial crystallization screen, multiple Os4BGlu12 crystals with small size were obtained from the pure protein of the second IMAC and S200 column. In the optimization of the crystallization condition, the pure protein from second IMAC column gave higher

yield and reproducibility of protein crystals than that of the final S200 column. Therefore, only the protein from second IMAC column was used for the optimization experiment. Figure 3.4 showed the SDS-PAGE profile of purified protein from three steps of purification.

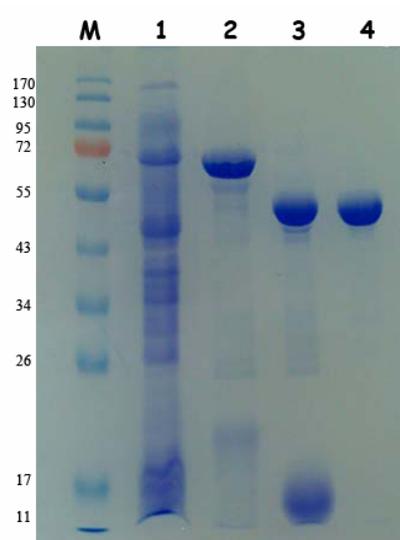


Figure 3.4 SDS-PAGE analysis of Os4BGlu12 protein from two steps of purification. Lane M, molecular weight protein maker. Lane 1, soluble protein extract of *E. coli* cells. Lane 2, thioredoxin-Os4BGlu12 fusion protein after initial IMAC. Lane 3, entorokinase digest of fusion protein. Lane 4, Os4BGlu12 tag free after second IMAC.

3.1.3 Dynamic light scattering (DLS)

DLS was used to estimate the molecular weight of the nonfusion Os4BGlu12 protein in solution (20 mM Tris-HCl, pH 7.2), and gave a single peak (100% of scattering) at 5.454 nm diameter, which corresponds to the estimated

molecular weight of tag-free Os4BGlu12 monomer (55 kDa) (Figure 3.5). This result confirmed that Os4BGlu12 forms a functional monomeric enzyme in solution.

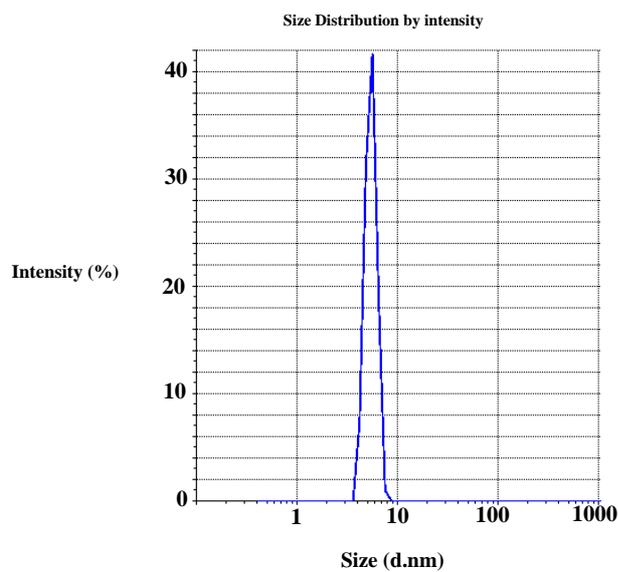


Figure 3.5 Dynamic light scattering (DLS) spectrum of tag-free Os4BGlu12 protein. DLS was used to determine the mean particle size and size distribution in a solution of 1 mg/mL protein in 20 mM Tris-HCl, pH 7.2, and was performed at 293K. The size distribution was plotted as intensity (%) versus size as the diameter of particle (d.nm). The protein gave a single peak with a diameter of 5.454 nm.

3.2 Crystallization of Os4BGlu12

3.2.1 Initial screening of crystallization of free-Os4BGlu12

The first trial for crystallization of freeOs4BGlu12 was screened with the JY Screen and microbatch under oil technique. The pure Os4BGlu12 protein from S200 gel filtration was used for the first crystal screening. Ten milligrams per milliliter of protein solution was used for the initial screening with a protein:

precipitant solution ratio of 1:1. The precipitant solution contained 0.2 M salt, 25% (w/v) PEG 4000 at pH ranges of 4.6 to 8.5 as listed in Table 3.1.

Table 3.1 List of JY Screen conditions that produced native Os4BGlu12 crystals at 15°C.

Condition number	Chemical composition	Day of observation
2A	25% (w/v) PEG 4000, 0.1 M sodium acetate pH 4.6, 0.2 M ammonium sulfate.	7
2C	25% (w/v) PEG 4000, 0.1 M sodium acetate pH 6.5, 0.2 M ammonium sulfate.	7
2D	25% (w/v) PEG 4000, 0.1 M sodium acetate pH 6.5, 0.2 M ammonium sulfate.	7
4A	25% (w/v) PEG 4000, 0.1 M sodium acetate pH 4.6, 0.2 M calcium chloride.	8
4D	25% (w/v) PEG 4000, 0.1 M Tris-hydrochloride pH 6.5, 0.2 M calcium chloride.	10
4F	25% (w/v) PEG 4000, 0.1 M Tris-hydrochloride pH 8.5, 0.2 M calcium chloride.	10
7F	25% (w/v) PEG 4000, 0.1 M Tris-hydrochloride pH 8.5, 0.2 M sodium chloride.	17

Small crystals appeared under several different conditions in the crystallization trials at 15°C on day 7 to day 17. The crystals were fully-grown in approximately three days after appearance (Figure 3.6). The single crystals were observed in the conditions 2A.1, 2A.2, 2C, 4F and 7F, and multiple crystals were observed in the conditions 2D, 4A.1, 4A.2 and 4D. Only a single crystal was observed in condition 7F (Table 3.1). It was broken after 20 days and numerous needle crystals appeared 2 days after this (Figure 3.7).

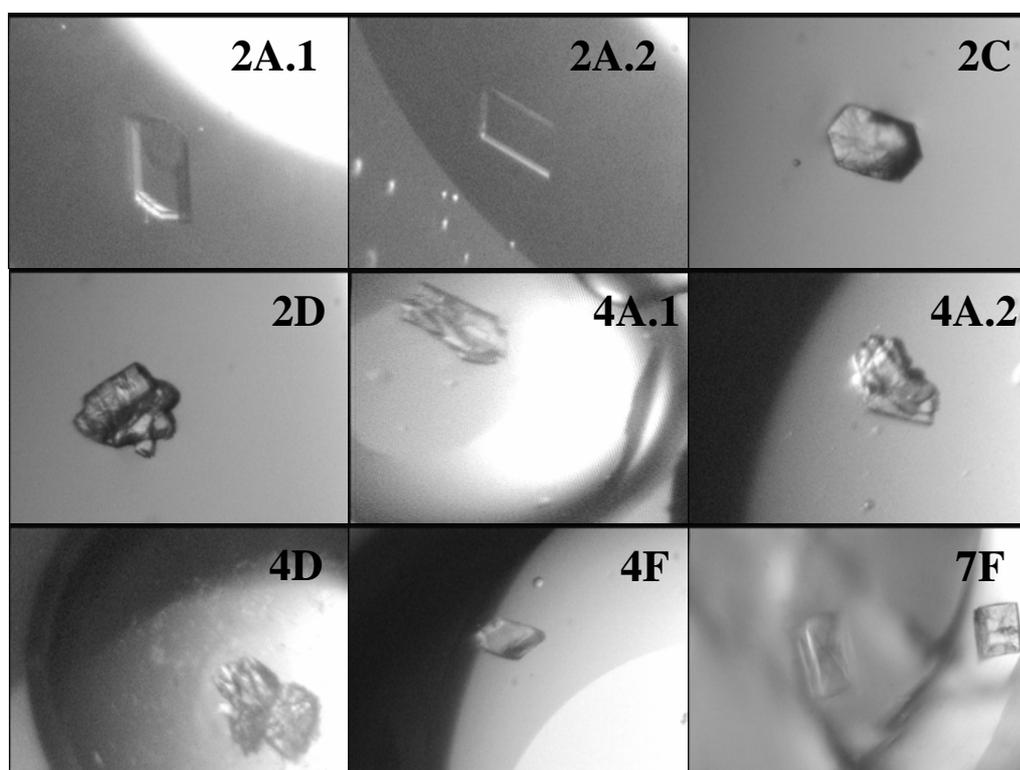


Figure 3.6 Photographs of crystals obtained from the initial crystallization trials with the JY Screen solutions in microbatch plates.

The crystal screen precipitant solution for each condition number is indicated in Table 3.1.

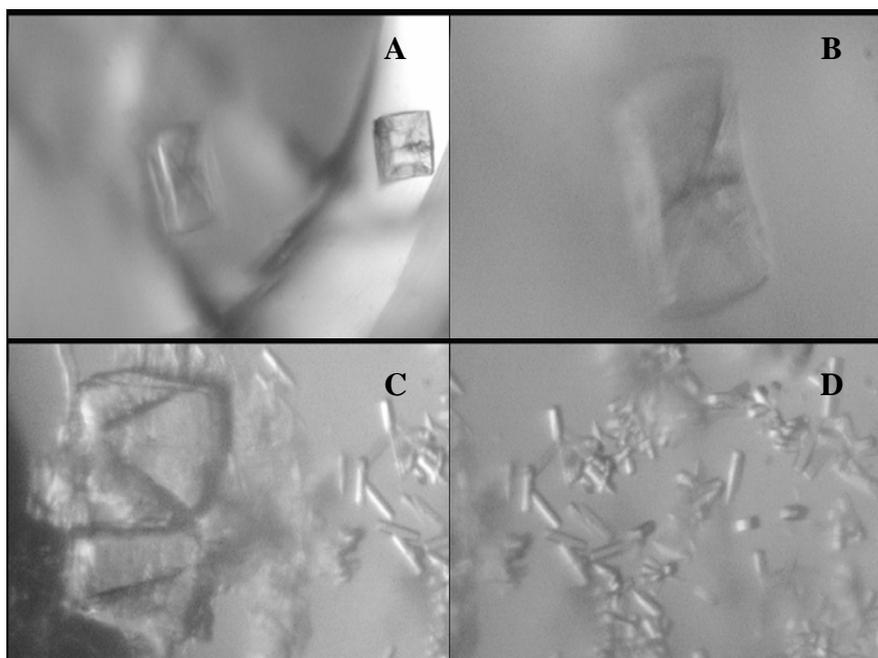


Figure 3.7 Photographs of crystals in condition 7F.

(A) Single Os4BGlu12 crystals that grew in 25% (w/v) PEG 4000, 0.1 M Tris-HCl, pH 8.5, 0.2 M NaCl were observed in day 17. (B) Fully-grown Os4BGlu12 crystal observed in day 20. The photograph shows the cracked crystal surface. The small single crystals were observed in days 22 and 25 in the cracked crystal condition, (Figure (C) and (D), respectively). The single and small size crystals in Figure D were used for diffraction.

All crystals obtained from the above positive conditions were diffracted by synchrotron radiation as an X-ray source. Diffraction patterns of the crystals in conditions 2A, 2C, 2D, 4A, 4D and 4F are similar to salt diffraction patterns and only condition 7F gave a diffraction pattern similar to a protein diffraction pattern (the

images are not shown), so the 7F condition was used for further optimization by hanging drop vapour diffusion.

3.2.2 Optimization of free Os4BGlu12 crystallization

The hanging drop vapour diffusion technique was used in the optimization of the crystallization. Condition 7F from JY Screen was used for optimization. Each drop contained 2 μL of 10 mg/mL protein solution and 1 μL of reservoir solution. The optimizing strategy consisted of two steps describe in the section 2.3.2. Fine-grid matrices were designed to improve the quality of the initially obtained crystals. The solutions of these matrices consisted of chemical compounds that were varied systematically with small increments from the initial crystallization trial condition (25% (w/v) PEG 4000, 0.1 M Tris-HCl, pH 8.5, 0.2 M NaCl). The needle crystals and small size crystals were observed after 5 days (Figure 3.8)

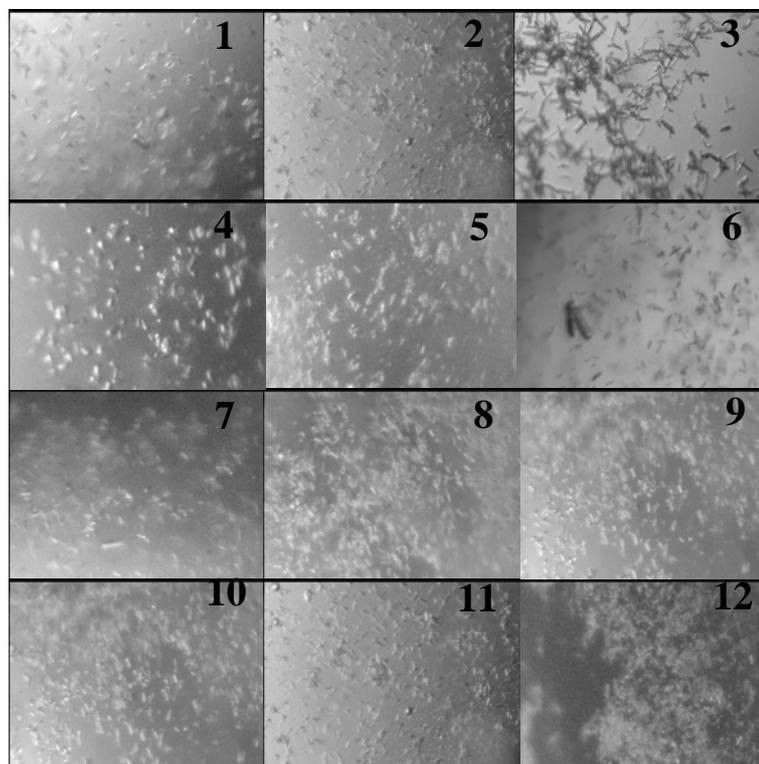


Figure 3.8 Photographs of crystals obtained from the crystallization trials with the variation of PEG 4000 and NaCl concentrations for the optimization of free Os4BGlu12 crystals.

The positive conditions in which crystals of free Os4BGlu12 were obtained are:

- | | |
|-----------------------------|------------------------------|
| 1, 13%PEG 4000, 0.24 M NaCl | 7, 23%PEG 4000, 0.16 M NaCl |
| 2, 13%PEG 4000, 0.26 M NaCl | 8, 23%PEG 4000, 0.26 M NaCl |
| 3, 19%PEG 4000, 0.16 M NaCl | 9, 25%PEG 4000, 0.16 M NaCl |
| 4, 19%PEG 4000, 0.26 M NaCl | 10, 25%PEG 4000, 0.26 M NaCl |
| 5, 21%PEG 4000, 0.16 M NaCl | 11, 27%PEG 4000, 0.16 M NaCl |
| 6, 21%PEG 4000, 0.26 M NaCl | 12, 27%PEG 4000, 0.26 M NaCl |

All NaCl concentrations did provide needle crystals of Os4BGlu12 (Figure 3.8). Small crystals were observed in the condition containing 19-27% PEG 4000 and 0.16-0.26 M NaCl. The conditions containing 13% PEG 4000 gave small crystals only in the condition contained 0.24 M and 0.26 M NaCl.

Crystal size plays an important role, as the intensity of X-ray diffraction is proportional to the volume of the diffracting matter. Acceptable size of a protein crystal for conducting the diffraction experiment is in the range of 0.2-0.5 mm (Rhodes, 2006). However, the size of the crystals of Os4BGlu12 obtained from these optimization conditions was far from these values. Since the condition of 19% PEG 4000 gave a bigger size crystal than the other concentrations of PEG 4000 (Figure 3.8), the NaCl and protein concentrations were further refined under the fixed concentration of 19% PEG 4000. As seen in Figure 3.9, the conditions number 1, 2, 3 and 4, which contain 3.0 and 5.0 mg/mL protein concentrations, gave single and big size crystals within 5 days, and several crystals were obtained in these conditions. Needle crystals were obtained in the conditions containing 7.0 and 9.0 mg/mL protein. Only conditions obtained with the protein concentration of 3.0 mg/mL were used for further optimization. Fewer and bigger crystals with similar shape were observed in the conditions with decreasing concentrations of NaCl (Figure 3.9). The result indicated that low NaCl concentration could produce the high quality of the crystals. However, all NaCl concentrations were applied for further optimization.

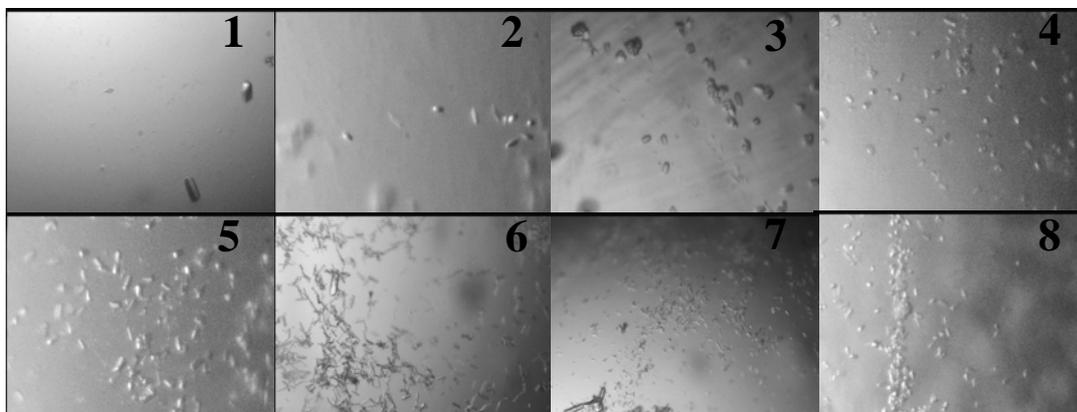


Figure 3.9 Photographs of crystals obtained from the crystallization trials with the variation of protein and NaCl concentrations at the fixed PEG 4000 concentration of 19% and 0.1 M Tris-HCl, pH 8.5.

The positive conditions that gave needle crystals within 5 days are:

- | | |
|-----------------------------------|------------------------------------|
| 1, 3.0 mg/mL protein, 0.16 M NaCl | 5, 7.0 mg/mL protein, 0.16 M NaCl |
| 2, 3.0 mg/mL protein, 0.26 M NaCl | 6, 7.0 mg/mL protein, 0.26 M NaCl |
| 3, 5.0 mg/mL protein, 0.16 M NaCl | 7, 9.0 mg/mL protein, 0.16 M NaCl |
| 4, 5.0 mg/mL protein, 0.26 M NaCl | 8, 9.0 mg/mL protein, 0.26 M NaCl. |

3.2.3 Optimization with seeding

Microseeding was applied to produce high quality crystals. The crystals from the condition number 1 (3.0 mg/mL protein in 19% PEG 4000, 0.16 M NaCl and 0.1 M Tris-HCl, pH 8.5) were picked from the original drop, and washed 6 times with the original precipitant solution to prepare the seed stock. The dilution of seed stock was used to streak seeds into new drops, which had been equilibrated to be a metastable solution. Initially, fewer and larger single crystals appeared within 2 days

in the number 2 and 3 conditions, as shown in Figure 3.10. The biggest single crystals were produced in 19% PEG 4000, 0.16 M NaCl and 0.1 M Tris-HCl, pH 8.5, 10 mg/mL protein with the seed stock diluted to 1/1000.

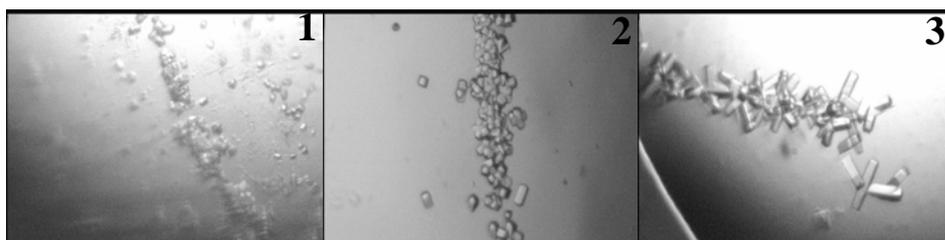


Figure 3.10 The seeding lines of Os4BGlu12 crystals obtained from the dilution series of seed stock prepared from the crystals obtained from condition number 1 in Figure 3.9.

The crystals in condition numbers 1, 2 and 3 grew. In these conditions, the seed stock which was diluted to 1/100, 1/500 and 1/1000, respectively, was streaked into new drops containing 19% PEG 4000, 0.16 M NaCl, 0.1 M Tris-HCl, pH 8.5 and 10 mg/mL protein.

The big single crystals produced in condition 3 (using 1/1000 dilution of seed stock) were used as the seed stock for further optimization. Optimization further refined the concentrations of NaCl in 19% PEG 4000, 0.1 M Tris-HCl, pH 8.5. The successful optimization was obtained in the conditions containing 3.0 mg/mL protein, 19% PEG 4000, 0.1 M Tris-HCl, pH 8.5, and NaCl varied from 0.16-0.26 M (Figure 3.11).

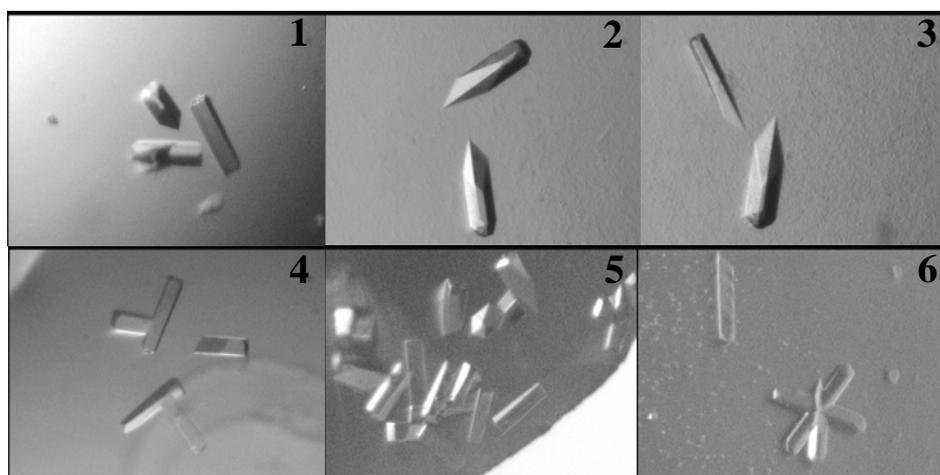


Figure 3.11 Photographs of free Os4BGlu12 crystals produced in various ranges of NaCl concentration.

The NaCl concentration was refined in the fixed protein concentration of 3.0 mg/mL, 19% PEG 4000, 0.1 M Tris-HCl, pH 8.5. FreeOs4BGlu12 crystals produced in various ranges of NaCl concentration: 1, 0.16 M; 2, 0.18 M; 3, 0.20 M; 4, 0.22 M; 5, 0.24 M and 6, 0.26 M.

The hexagonal single crystals were observed after one day and crystals fully grew within fifteen days with approximately 110 x 25 x 20 μm dimensions in conditions number 1, 2 and 3, which contained 0.16 M, 0.18 M and 0.20 M NaCl, respectively, in 19% PEG 4000, 0.1 M Tris-HCl, pH 8.5. While the single crystals with size of approximately 70 x 15 x 15 μm were obtained in the conditions number 4 and 5, which contained 0.22 M and 0.24 M NaCl, respectively, in 19% PEG 4000, 0.1 M Tris-HCl, pH 8.5. All crystals obtained from conditions number 1, 2, 3, 4 and 5 were diffracted by synchrotron radiation as an X-ray source. For the X-ray diffraction,

crystals were soaked for a few seconds in prepared cryoprotectant solution composed of 18% increasing concentration of original reservoir solution and 18% glycerol. However, all crystals were diffracted to low resolution at approximately 4-6 Å and the data is not enough for collection and processing.

Variable parameters, including the molecular weight of PEG at the concentration of 19% and the concentration of NaCl, were optimized in 0.1 M Tris-HCl, pH 8.5, and 3.0 mg/mL protein concentration using the same seed stock as above experiment. The good size and shape crystals appeared in the presence of PEG 2000 and PEG 3350 within 7 and 5 days, respectively. The maximum sizes of crystals with approximate dimensions up to 120 x 25 x 20 μm were obtained in the presence of PEG 2000 and PEG 3350 on the day fifteenth and seventeenth, respectively.

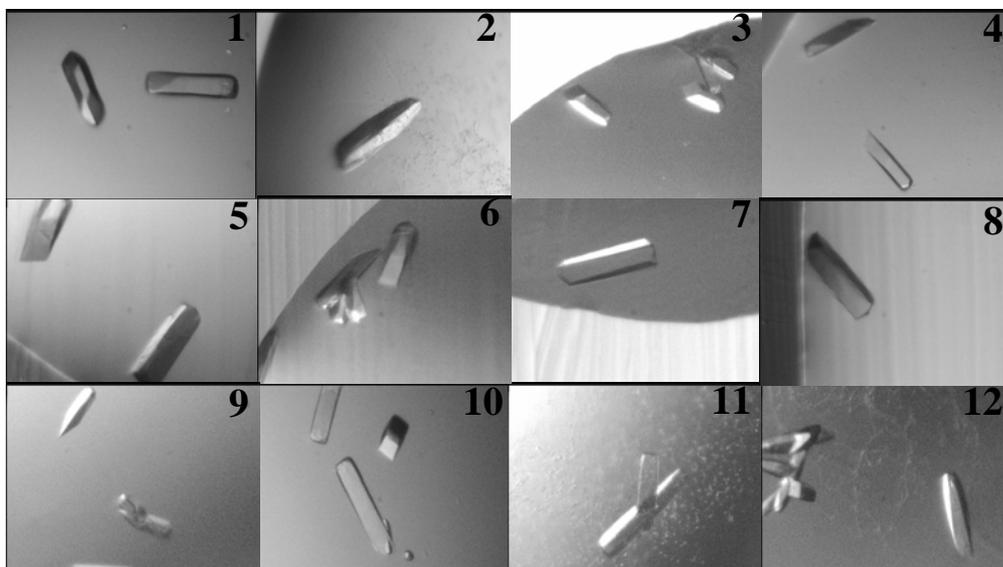


Figure 3.12 Optimization of free Os4BGlu12 crystals with variation of PEG molecular weight, and concentration of NaCl.

The single crystals obtained from the conditions are:

1, 19%PEG 2000, 0.16 M NaCl	7, 19%PEG 3350, 0.16 M NaCl
2, 19%PEG 2000, 0.18 M NaCl	8, 19%PEG 3350, 0.18 M NaCl
3, 19%PEG 2000, 0.20 M NaCl	9, 19%PEG 3350, 0.20 M NaCl
4, 19%PEG 2000, 0.22 M NaCl	10, 19%PEG 3350, 0.22 M NaCl
5, 19%PEG 2000, 0.24 M NaCl	11, 19%PEG 3350, 0.24 M NaCl
6, 19%PEG 2000, 0.26 M NaCl	12, 19%PEG 3350, 0.26 M NaCl

All crystals shown in Figure 3.12 were diffracted at the NSRRC, Taiwan. The cryoprotectant soaking conditions of the crystals for X-ray diffraction were optimized by varying the concentrations of precipitant and glycerol, and the incubation temperature (room temperature and 15°C). Some obtained crystals were incubated in various cryoprotectant solution overnight and the crystal surfaces were observed (Figure 3.12). The best condition was obtained when the crystal was soaked in the cryoprotectant solution containing 28% glycerol and a precipitant solution concentration increased by 22% and performed at 15°C. The soaked crystals were flash-frozen in the liquid nitrogen for 15 second. The crystals in conditions number 1 and 8 (Figure 3.12) were soaked in the cryoprotectant solution containing 2 mM 2,4-dinitrophenyl 2-fluoro-2-deoxy- β -D-glucoopyranoside (G2F) inhibitor and 20 mM cellotetraose, respectively. Then the crystal was mounted in a cryo-loop and was placed onto goniometer head between the X-ray beam and the detector. The diffraction data were processed with the HKL2000 package. The crystals soaked with G2F and with cellotetraose were diffracted to 2.90 Å and 2.80 Å resolutions, respectively. Only one free Os4BGlu12 crystal obtained from 19% PEG 3350, 0.1 M Tris-HCl, pH 8.5, and 0.16 M NaCl gave 2.50 Å resolution.

3.2.4 Optimization of Os4BGlu12-G2F complex crystal

The complex of Os4BGlu12 crystal with G2F was produced by the hanging drop vapour diffusion method at 15°C. Os4BGlu12 protein solution (3 mg/mL in Tris-HCl, pH 7.2) was incubated with 10 mM G2F solution for 10 minute at 4°C. Crystals of Os4BGlu12-G2F were obtained in 19% PEG 2000 and PEG 3350 in 0.1 M Tris-HCl, pH 8.5, and 0.16-0.26 M NaCl by microseeding (Figure 3.13).

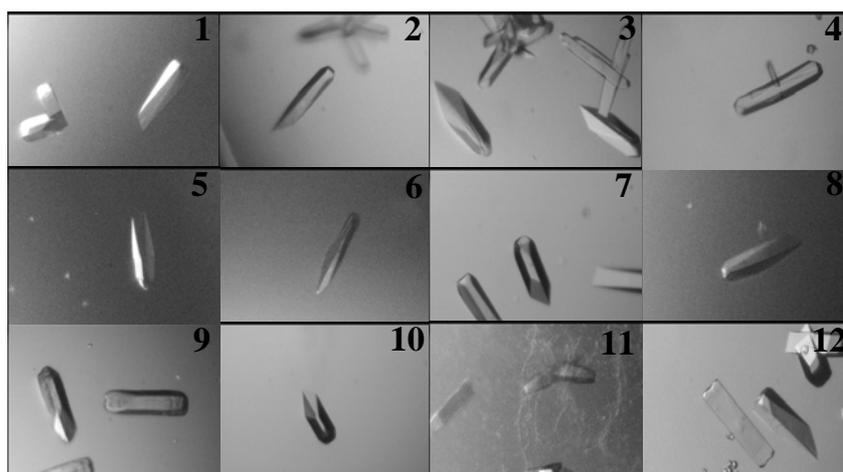


Figure 3.13 Optimization of Os4BGlu12-G2F complex crystals with variation of PEG molecular weight (19%) and NaCl concentration in 0.1 M Tris-HCl, pH 8.5, and 3.0 mg/mL protein.

Crystals were obtained in the following conditions.

- | | |
|-----------------------------|-----------------------------|
| 1, 19%PEG 2000, 0.16 M NaCl | 7, 19%PEG 3350, 0.16 M NaCl |
| 2, 19%PEG 2000, 0.18 M NaCl | 8, 19%PEG 3350, 0.18 M NaCl |
| 3, 19%PEG 2000, 0.20 M NaCl | 9, 19%PEG 3350, 0.20 M NaCl |

4, 19%PEG 2000, 0.22 M NaCl	10, 19%PEG 3350, 0.22 M NaCl
5, 19%PEG 2000, 0.24 M NaCl	11, 19%PEG 3350, 0.24 M NaCl
6, 19%PEG 2000, 0.26 M NaCl	12, 19%PEG 3350, 0.26 M NaCl

The single crystals of Os4BGlu12 with G2F were obtained within 2 days. The crystals suitable for X-ray diffraction with approximate dimensions of 125 x 25x 25 μm were observed on day fifteen. All of the complex crystals obtained from Figure 3.13 were diffracted with synchrotron radiation as the X-ray source. For the X-ray diffraction, single crystals of Os4BGlu12 with G2F were picked up in a nylon loop and then transferred to cryoprotectant solution containing a 122% of precipitant solution, 28% glycerol with 10 mM G2F and cryo-protected for 15 seconds. They were frozen and stored in liquid nitrogen. Only two Os4BGlu12-G2F complex crystals gave good resolution. Crystals of the complex that grew in 19% PEG 2000, 0.1 M Tris-HCl, pH 8.5, 0.16 M NaCl (Figure 3.13 number 1), and 19% PEG 3350, 0.1 M Tris-HCl, pH 8.5, 0.18 M NaCl (Figure 3.13 number 8) diffracted to 2.45 Å and 2.65 Å resolution, respectively.

3.3 Crystal characterization

To prove that the crystals obtained were made of Os4BGlu12 protein, SDS-PAGE was used to confirm the dissolved crystal content. Several crystals from the Os4BGlu12 crystallization trials were collected and washed thoroughly in precipitant solution to remove any dissolved protein solution that may be left on the surface of the crystals. As seen in the SDS-PAGE in Figure 3.13, the protein from dissolved

crystals had molecular weight approximately 55 kDa, which is similar to the size of nonfusion Os4BGlu12 protein.

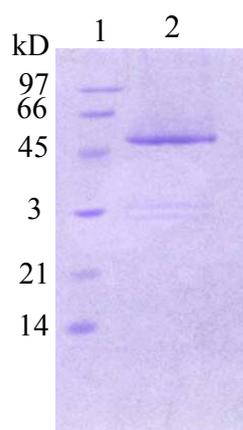


Figure 3.14 SDS-PAGE of protein from dissolved crystals of Os4BGlu12. Lane 1, molecular weight protein marker and lane 2, dissolved Os4BGlu12 crystals.

3.4 Preliminary X-ray diffraction analysis of free Os4BGlu12 and its complex crystals.

Four X-ray diffraction data set of free Os4BGlu12 and its complex crystals with G2F and cellotetraose were collected at 100 K in a nitrogen stream from an Oxford Cryosystems Cryo-stream with an ADSC Quantum 315 CCD detector. From diffraction data collection, the four data sets were determined to have tetragonal $P4_32_12$ space group symmetry with different unit cell parameters, as summarized in the Table 3.2. The free Os4BGlu12 crystal and the crystal of complex with G2F that grew in 0.16 M NaCl data sets were collected to 99.9% completeness over in the resolution ranges of 30.0-2.50 Å and 30.0-2.45 Å, with high numbers of unique

reflections of 39533 and 43131 and R_{sym} values of 10.9% and 9.4%, respectively. The diffraction data of the Os4BGlu12 crystal with C4 and Os4BGlu12-G2F complex crystal that grew in 0.18 M NaCl were collected to completeness 99.9% the resolution ranges of 30.0-2.65 Å and 30.0-2.80 Å, which correlated with low numbers of unique reflections of 34255 and 29438, respectively. The four data sets have two molecules per asymmetric unit. The Matthews coefficients (V_M), were calculated to be 2.46 Å³ Da⁻¹, 2.68 Å³ Da⁻¹, 2.68 Å³ Da⁻¹ and 2.75 Å³ Da⁻¹, indicating solvent contents of approximately 49.98%, 54.17%, 54.20% and 55.29% for the freeOs4BGlu12 with G2F, the crystals of complexes of Os4BGlu12 crystal diffracted to 2.45 Å and 2.65 Å resolution and the Os4BGlu12 crystal with C4, respectively.

Table 3.2 Data collection and processing statistics

Dataset	Os4BGlu12	Os4BGlu12_ G2F	Os4BGlu12_ G2F	Os4BGlu12 _C4
Beamline	BL13B1	BL13B1	BL13B1	BL13B1
Wavelength (Å)	1.00	1.00	1.00	1.00
Space group	P4 ₃ 2 ₁ 2			
	a = 112.656	a = 114.112	a = 114.161	a = 114.492
Unit-cell parameters (Å)	b = 112.656	b = 114.112	b = 114.161	b = 114.492
	c = 182.773	c = 184.541	c = 184.507	c = 184.564
Resolution range (Å)	30-2.50	30-2.45	30-2.65	30-2.80
Resolution of outer shell(Å)	2.50-2.49	2.51-2.45	2.72-2.65	2.86-2.80
No. Unique reflections	39533	43131	34255	29438
No. Observed reflections	408534	381316	257680	315858
Completeness (%)	100.0 (99.9)	100.0 (99.9)	100.0 (99.9)	100.0 (99.9)
Average redundancy per shell	9.7 (10.0)	8.4 (8.6)	7.3 (7.1)	10.2 (10.5)
$I/\sigma(I)$	19.60 (6.4)	21.70 (4.4)	20.34 (4.5)	20.19 (5.1)
$R_{\text{(merge)}}$ (%)	10.6 (41.7)	9.4 (49.9)	9.4 (46.0)	11.6 (50.0)

3.5 Molecular replacement

The Os4BGlu12 structure was solved by molecular replacement with the *MOLREP* program using the cyanogenic β -glucosidase from white clover structure (1CBG chain A, Barrett *et al.*, 1995) as a search model. Only two from four data sets were used for refinement of the final model those of Os4BGlu12 with and without G2F, which were determined at 2.45 Å and 2.50 Å resolutions, respectively. Initially, the free Os4BGlu12 model was built by molecular replacement, and the structures of Os4BGlu12 with G2F was built based on the structure of free Os4BGlu12.

3.6 Structure determination of free Os4BGlu12 and the Os4BGlu12 complex with G2F inhibitor

3.6.1 The model quality of free Os4BGlu12 and G2F complex

As previously noted, the final model of Os4BGlu12 with and without G2F were refined by molecular replacement to 2.45 Å and 2.50 Å resolutions, respectively. The refinement statistics are shown in Table 3.3. In general, the electron density of almost all residues was seen clearly in the final models of both structures (Figure 3.15). The electron density for residues 9 to 486 were observed in free Os4BGlu12 and the Os4BGlu12 complex with G2F, but eight residues, including Ala1, Tyr2, Asn3, Ser4, Ala5, Gly6, Glu7 and Pro8, were not visible. The final $2F_{\text{obs}}-F_{\text{cal}}$ map, contoured at 1.0 σ , shows continuous density for most main-chain and side-chain atoms for both structures. Residues, Asn333, Ser334, Gly335, Leu 336, Asn337 have weak density at the side-chain of both structures. At the C-terminus, the electron density of the side chain of Lys486 of both structures was poorly defined. A few residues on the surface of free Os4BGlu12 and Os4BGlu12-G2F did not have electron density for their side-chains.

Electron density of one glucose unit was observed at -1 subsite of the G2F complex structure. According to the model from the electron density, one G2F unit is attached to the catalytic nucleophile of each molecule in the asymmetric unit. Tris molecule was seen in the structure of freeOs4BGlu12 at -1 subsite, but this was not found in the Os4BGlu12-G2F complex structure.

The final model of Os4BGlu12 with and without G2F contains 7720 nonhydrogen protein atoms (per 2 molecules) and 231 and 322 water molecules were present in the free Os4BGlu12 and Os4BGlu12-G2F structures, respectively. The

final crystallographic R -factor and R_{free} for the free Os4BGlu12 model were 21.36% and 25.44%, respectively, and it had a good stereochemistry with root mean square deviation (rmsd) of 0.007° on bond length and 1.048 \AA on angles. The final refined structure of G2F complex had an R -factor of 21.40% and R_{free} of 25.67% with rmsd of 0.007° on bond length and 1.055 \AA on angles. In the Ramachandran plot computed in PROCHECK, 89.3% of freeOs4BGlu12 structure and 89.6% of the Os4BGlu12 complex with G2F residues are found in the most favorable regions.

Table 3.3 Refinement statistics for native Os4BGlu12 and Os4BGlu12 with G2F

Dataset	Os4BGlu12	Os4BGlu12_G2F
R_{factor} (%)	21.36	21.40
R_{free} (%)	25.44	25.67
No. of residues in protein	486 (2 molecules)	486 (2 molecules)
No. protein atoms	7720 (2 molecules)	7720 (2 molecules)
No. Ligand atoms	16 (Tris)	22 (G2F)
No. Other hetero atoms	None	36 (GOL)
No. waters	231	322
<i>Mean B-factor</i>		
Protein	33.152	33.82
Ligand	53.78 (Tris)	48.40 (G2F)
Other hetero atoms	None	49.77
Waters	31.72	32.96
r.m.s. bond deviations (length)	0.007	0.007
r.m.s angle deviations (degrees)	1.048	1.055
Ramachandran plot		
Residues in most favorable regions (%)	89.3	89.6
Residues in additional allowed regions (%)	9.8	9.6
Residues in generously allowed regions (%)	1.0	0.7
Residues in disallowed regions (%)	0	0

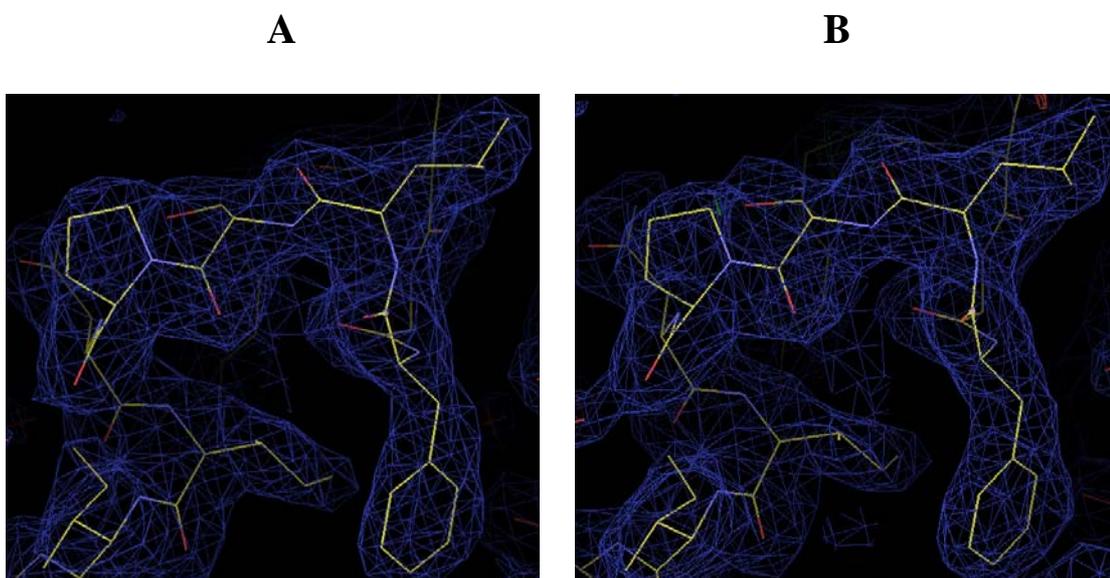


Figure 3.15 $2F_{\text{obs}}-F_{\text{cal}}$ maps of 2.50 Å resolution free Os4BGlu12 (**A**) and 2.45 Å resolutions Os4BGlu12 with G2F (**B**).

A and B, show the quality of the electron density map at amino acid residues 147 to 153 of the final model and these maps are contoured at the 1.0σ level. The figures were drawn with Coot (Emsley P. and Cowtan K, 2004).

3.6.2 The overall structure of Os4BGlu12 with and without G2F

The overall fold of free Os4BGlu12 and the Os4BGlu12-G2F complex has a $(\beta/\alpha)_8$ barrel fold as found in related glycosyl hydrolase family 1 (GH1) glycosidases. The asymmetric units of the free Os4BGlu12 and G2F complex crystals contains a dimer, whereas gel filtration and dynamic light scattering of the freeOs4BGlu12 were consistent with the presence of a monomer (Figure 3.3 and 3.5). The overall structure of Os4BGlu12 shows 478 residues for each molecule in the asymmetric unit starting 9 residues after the predicted signal peptide and ending with residue 486 as the last well refined residue. The catalytic acid/base Glu179 and the nucleophile Glu393 are located at the C-terminal ends of β -strands 4 and 7, respectively. Glu179 and Glu393 are located on opposite sides of a cleft at the bottom of the active site.

A Tris molecule is bound within the -1 subsite of the free Os4BGlu12 structure (Figure 3.16 A). Within the structure of the Os4BGlu12-G2F complex, electron density for G2F inhibitor could be observed at the end of the groove and six hetero atoms (glycerol) are observed around the surface of the G2F complex structure (Figure 3.16 B).

The two molecules in the asymmetric unit of both structures are linked by Zn^{2+} ions, which is chelated by the interaction between His69 and Asp66 from monomer A and His69 and Glu36 from monomer B. The Zn^{2+} ion has tetrahedral coordination, with angles of 101.23° between His69/A and Glu36/B, 115.87° between His69/A and His69/B, 119.42° between His69/A and Asp69/A, with coordination at a ligand-ion distance of 1.91 Å to 2.06 Å (Figure 3.17). The dimer interfaces are also stabilized by side-chain charged-charged interaction between residues Glu108/A and

Lys111/B, Lys470/A and Asp62/B, Asp467/A and Arg469/B, Arg469/A and Asp66/B, which form a salt bridge network of close contact about a two molecules in asymmetric unit. The overall structure of Os4BGlu12 contains several salt bridges and hydrogen bonds. Of the 25 arginine residues, 18 residues form salt bridges, 4 residues form hydrogen bonds with water atom, and three residues do not have hydrogen bond with other protein atoms. Fifteen salt bridges are obtained from twenty-seven residues of aspartic acid, nine of which form hydrogen bonded with other protein atoms and water. Asp66/A has bonded with the Zn^{2+} ion, and another two aspartic residues do not interaction with the other atoms. Thirty lysine residues are found in the structure of Os4BGlu12, fifteen residues of which form salt bridges, ten residues are mainly exposed to solvent, and five residues form hydrogen bond interactions with other protein atoms. Three of ten histidine form salt bridges, one residues (His69/A) has coordinated with Zn^{2+} ion and the remaining residues form hydrogen bond with other protein atoms. Twenty glutamic acid residues are found in the structure of Os4BGlu12, in which fifteen residues form salt bridges, seven residues form hydrogen bond interactions with water atoms and other protein atoms and the remaining one does not interact with other atoms. Two disulfide bridges were found in the crystal structure of Os4BGlu12, involving Cys184-Cys219 and Cys198-Cys206, located at loop B region (Figure 3.16 A). Only Cys198-Cys206 is generally present in other plant family 1 β -glycosidases.

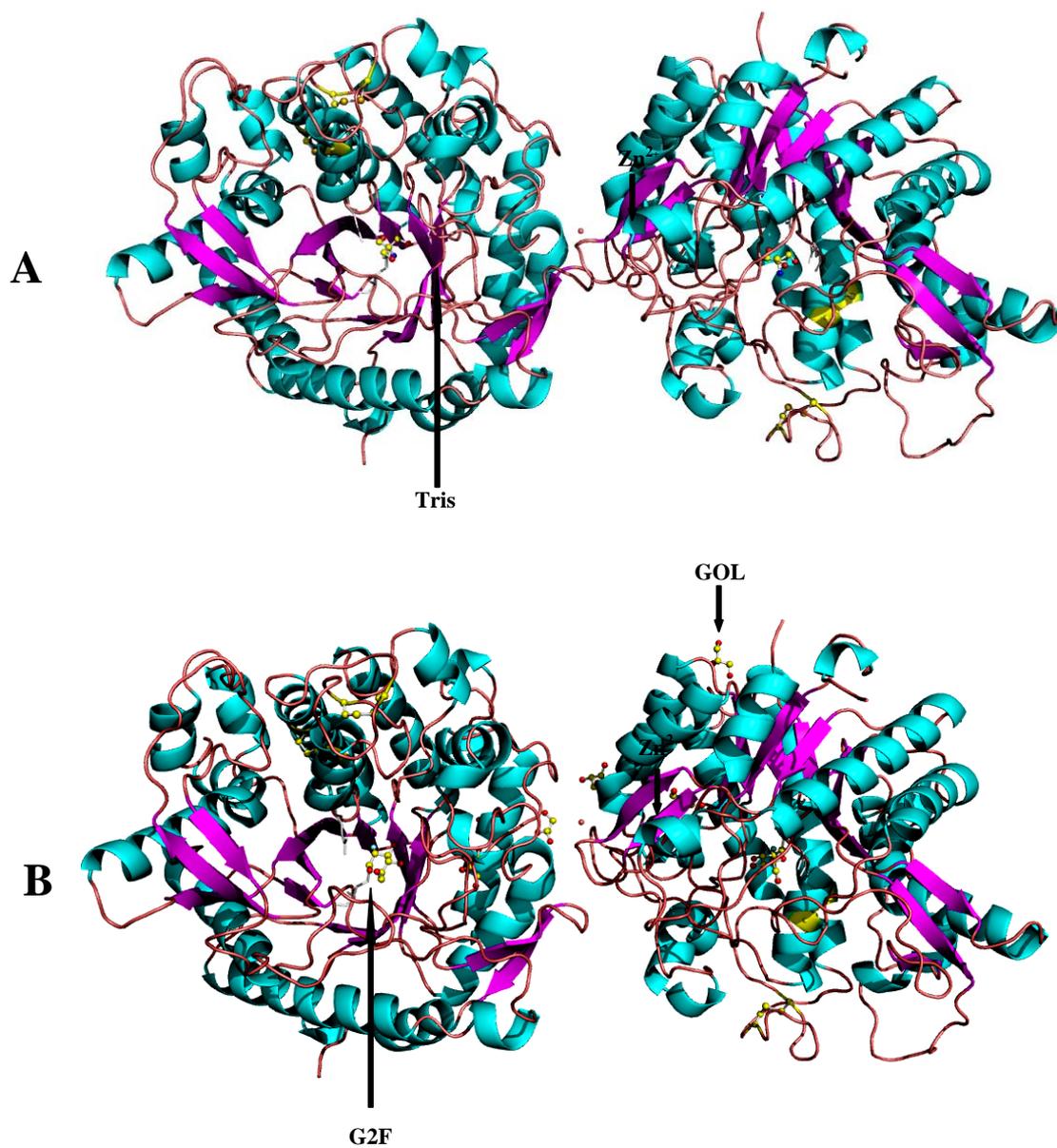


Figure 3.16 Ribbon diagram representation of the dimeric structure of free Os4BGlu12 (A) and G2F complex (B).

The β -strands are colored purple for free Os4BGlu12 structure and the G2F complex structure, α -helices in blue for the structures with and without G2F molecule. The loops are depicted in cyan for free Os4BGlu12 and the G2F complex structure. The

Zn^{2+} ion between two molecules in asymmetric unit is shown in red between A and B protein molecules for free Os4BGlu12 and Os4BGlu12 with G2F. The catalytic residues Glu393 and Glu179 of both structures are shown as sticks colored by white. Two molecules of Tris associated with the A and B molecules in the asymmetric unit of free Os4BGlu12 structure are represented by balls and sticks with carbon in yellow, oxygen in red and nitrogen in blue. The G2F molecule in the complex (B) is represented by balls and sticks with carbon in yellow, oxygen in red, and fluoride in blue. Six glycerol molecules around the surface of the G2F complex structure are colored in yellow (carbons) and red (oxygen). The figure was generated by Pymol (DeLano, 2002).

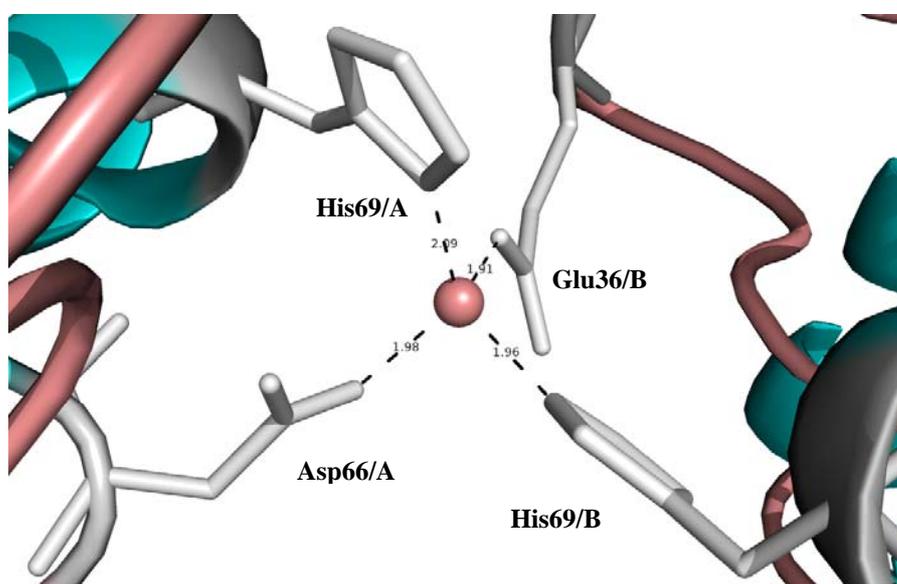


Figure 3.17 Stereo views of the Zn^{2+} ion binding site.

Key residues interacting with the zinc ion are labeled and depicted as sticks with carbon, nitrogen and oxygen atoms colored in white. The zinc ion is shown as a gray

sphere, and the distances to the ligands are given in Å. This figure was prepared with the Pymol program (DeLano, 2002).

3.6.3 Comparison of Os4BGlu12 structure with other known GH1 structures

For structural comparison, the known 3D structures of the GH1, cyanogenic β -glucosidase from *Trifolium repens* (1CBG) (Barrett *et al.*, 1995), rice BGlu1 β -glucosidase (2RGM) (Chuenchor *et al.*, 2008), dhurrinase isozyme I from *Sorghum bicolor* (1VO2) (Verdoucq *et al.*, 2004), β -glucosidase from *Triticum aestivum* (2DGA) (Sue *et al.*, 2006), β -glucosidase isozyme I from *Zea mays* (1E1E) (Czizek *et al.*, 2001) and myrosinase from *Sinapis alba* (1MYR) (Burmeister *et al.*, 1997), were superimposed on Os4BGlu12 structure (Figure 3.18). The 1CBG, 2RGM, 1VO2, 2DGA and 1E1E structures represent plant O-glucosidases, and 1MYR is a plant S-glycosidase. These structures share amino acid sequence identity of 44-63%. The C $_{\alpha}$ atoms of molecule A from the 1CBG, 2RGL, 1VO2, 2DGA, 1E1E and 1MYR structures superimposed to rice Os4BGglu12 molecule A with rmsd of 0.85, 0.95, 1.01, 1.03, 1.05 and 1.08 Å, respectively (Table 3.4). The 1CBG had the highest of sequence identity (63%) and the lowest rmsd value of overall structure with rice Os4BGlu12. Whereas, the structure of 1MYR had the largest difference from Os4BGlu12 with the highest rmsd (1.08 Å) and the lowest sequence identity (44%).

The structural superimposition showed that the core (β/α)₈ structures of the GH family 1 are similar in shape and size, but the largest differences were mainly found in loop regions surrounding the active site. The substrate binding pocket of Os4BGlu12 structure is formed primarily by the four extended loops connecting

strands and helices at the C terminal side of the barrel: loop A (Ser25-Asp66, between $\beta 1$ and $\alpha 1$), loop B (Glu179-Arg213, between $\beta 4$ and $\alpha 4$), loop C (Typ321-Pro370, between $\beta 6$ and $\alpha 6$) and loop D (Asn394-Asp412, between $\beta 7$ and $\alpha 7$). Loop A has little difference in length and structural alignment from other plant GH1 structures. This loop contains the residues involved in the dimerization, Asp66 from monomer A and Glu36 from monomer B, which are located at similar positions as the residues at the dimer interfaces of 1CBG (residues Asp66 and His53), 2RGL (Asp65 and His68) and 1MYR (Asp70 and His56). Loop B contains two disulfide bridges, one of which is conserved in plant GH1 structures, and residues in this loop contribute part of the aglycone binding pocket. The largest differences are in loop C as seen in structural comparison (Figure 3.18) and sequence alignment (Figure 3.19). Os4BGlu12 has a longer loop compared to the 1MYR and 2RGL structures, while loop C of 2RGL and Os4BGlu12 have wider cavity than the 1MYR. Loop C forms a part of the aglycone binding pocket and supports the crucial residues defining the binding affinity of the active site. Loop D is the smallest loop and it is not close to the substrate-binding site, although high variation was seen among plant GH1 structures in this loop. This loop in Os4BGlu12 is very similar in shape and size with that of 1CBG.

Table 3.4 Root-mean-square deviation between the structures of Os4BGlu12 and other plant β -glucosidases

PDB code	r.m.s.d	Sequence identity (%)
1CBG	0.85	63
2RGL	0.95	51
1VO2	1.00	48
2DGA	1.03	49
1E1E	1.05	49
1MYR	1.08	44

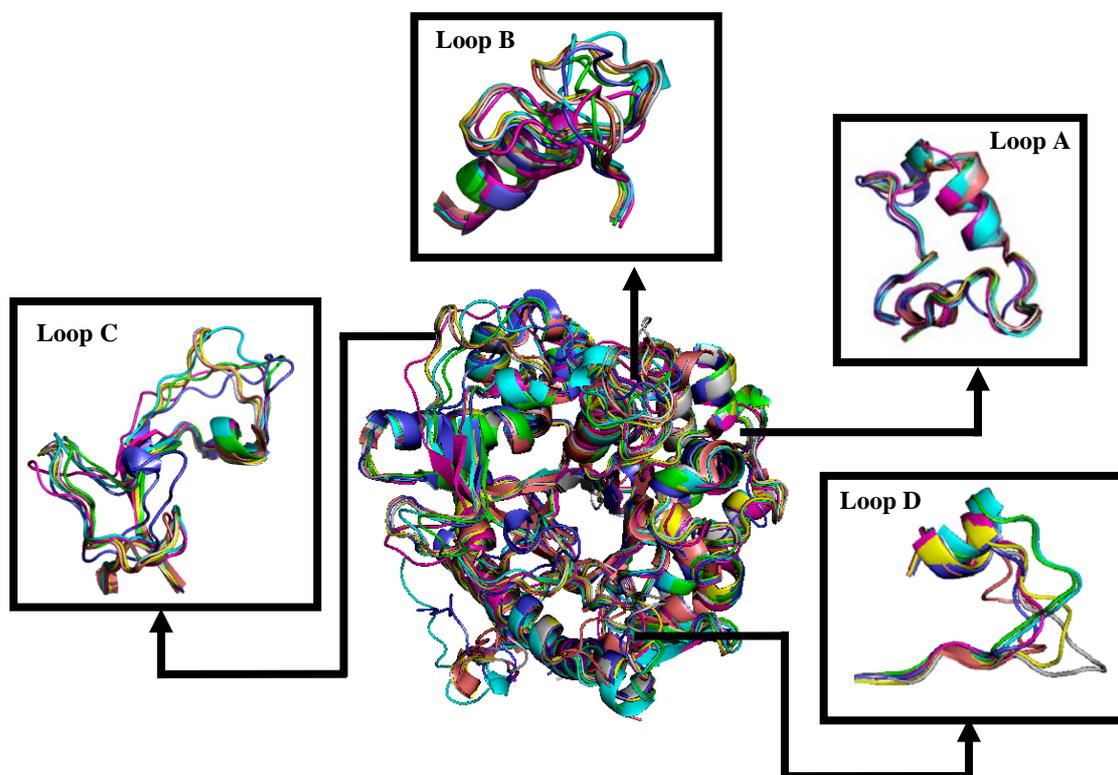


Figure 3.18 Ribbon representation of the superimposition of rice Os4BGlu12 (green), with other GH1 enzymes.

Other structures shown are rice BGlu1 β -glucosidase (2RGM, pink), *Trifolium repens* (1CBG, blue), *Sinapis alba* (1MYR, purple), β -glucosidase isozyme I from *Zea mays* (1E1E, orange), dhurrinase isozyme I from *Sorghum bicolor* (1VO2, yellow) and β -glucosidase from *Triticum aestivum* (2DGA, white). Loops A-D, which constitute the doorway to the active site are expanded to the side, as indicated by the arrows, to show the differences in loop structures. The figures were produced with the Pymol program (DeLano, 2002).

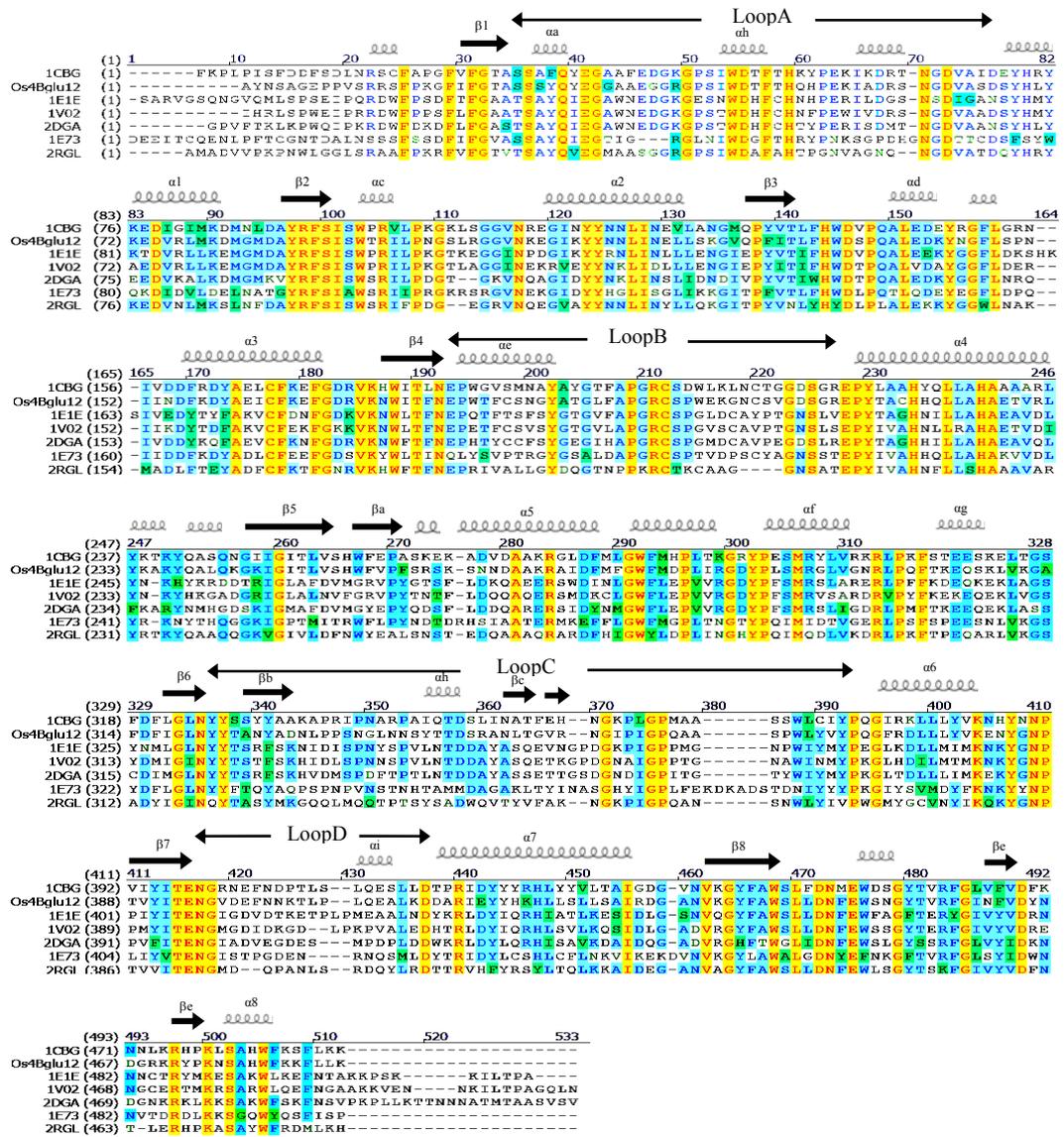


Figure 3.19 Sequence alignment of rice Os4BGlu12 with other GH1 enzymes.

The GH1 enzymes in this alignment include rice BGlu1 β -glucosidase (2RGM), *Trifolium repens* (1CBG), *Sinapis alba* (1MYR), β -glucosidase isozyme I from *Zea mays* (1E1E), dhurrinase isozyme I from *Sorghum bicolor* (1VO2) and β -glucosidase from *Triticum aestivum* (2DGA). The alignment was performed with ClustalW

(Thompson *et al.*, 1994) and the figure was prepared with ESPript (Gouet *et al.*, 1999).

3.6.4 Enzyme active site

The active site of Os4BGlu12 is located at the bottom of an approximately 20 Å deep pocket, and the distance between two walls forming the slot-like pocket is approximately of 8 Å at its narrowest point. It can be divided in two parts, the recognition site for the glucose moiety located at the deepest part of the active site and the binding site for the aglycone at the entrance of the pocket. The hydrophobic environment is present on the aglycone part of the active site, whereas hydrophobic, polar and charged residues are present in the glycone binding site (Figure 3.20).

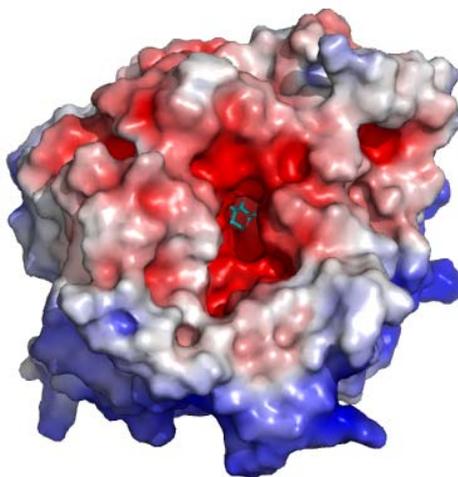


Figure 3.20 Electrostatic surface representation of Os4BGlu12 with G2F located deep in the active site.

The surface showing positively charged region in blue, negatively charged region in red and neutral region in white. G2F is represented by sticks with carbons in blue and oxygen in red. The figure was generated by Pymol (DeLano, 2002).

The overall conformation of molecules A and B are very similar with an rmsd of 0.06 Å, and the electrostatic surface of A and B molecules was nearly the same, like the overall conformations of both molecules. A tris molecule was observed at the bottom of active site of free Os4BGlu12 structure. Tris molecules with different conformations are present in molecules A and B in the asymmetric unit. Following refinement of two conformations of Tris, the $F_{\text{obs}}-F_{\text{calc}}$ electron density map was relatively clean in molecule B, whereas it showed less density but significant features in molecule A (Figure 3.21 A, B), as was also seen in the $2F_{\text{obs}}-F_{\text{calc}}$ electron density map (Figure 3.21 C, D). At the bottom of the slot-like active site of the free Os4BGlu12 structure, a cluster of residues form hydrogen bonds with O-1, O-2, O-3 and N-1 of Tris. Several hydrogen-bonding interactions with the Tris in molecules A and B are different (Figure 3.22 A, B, C, D). For the binding mode of Tris in molecule A, O3 and N1 of Tris are hydrogen bonded to Glu393 O ϵ 1 at distances of 2.75 Å and 3.09 Å, respectively, and O3 also forms a hydrogen bond with Glu179 O ϵ 2 (2.73 Å) (Figure 3.22 A). The hydroxyl group of Tyr322 forms a hydrogen bond with N1 of Tris (3.01 Å). O2 of Tris interacts with Glu449 O ϵ 1 (2.58 Å), O ϵ 2 (3.44 Å) and Gln29 N ϵ 2 (3.15 Å), while O1 of Tris forms a hydrogen bond with Glu449 O ϵ 1 (2.71 Å). There are four hydrogen bonds formed between Tris and amino acid of molecule B (figure 3.22 B). O1 and O2 of Tris form hydrogen bonds with Glu179 O ϵ 2 (2.97 Å) and Gln29 O ϵ 1 (3.28 Å), respectively, while O3 and N1 interact with Glu449 O ϵ 2 at distances of 3.09 Å and 3.10 Å, respectively (Table 3.4).

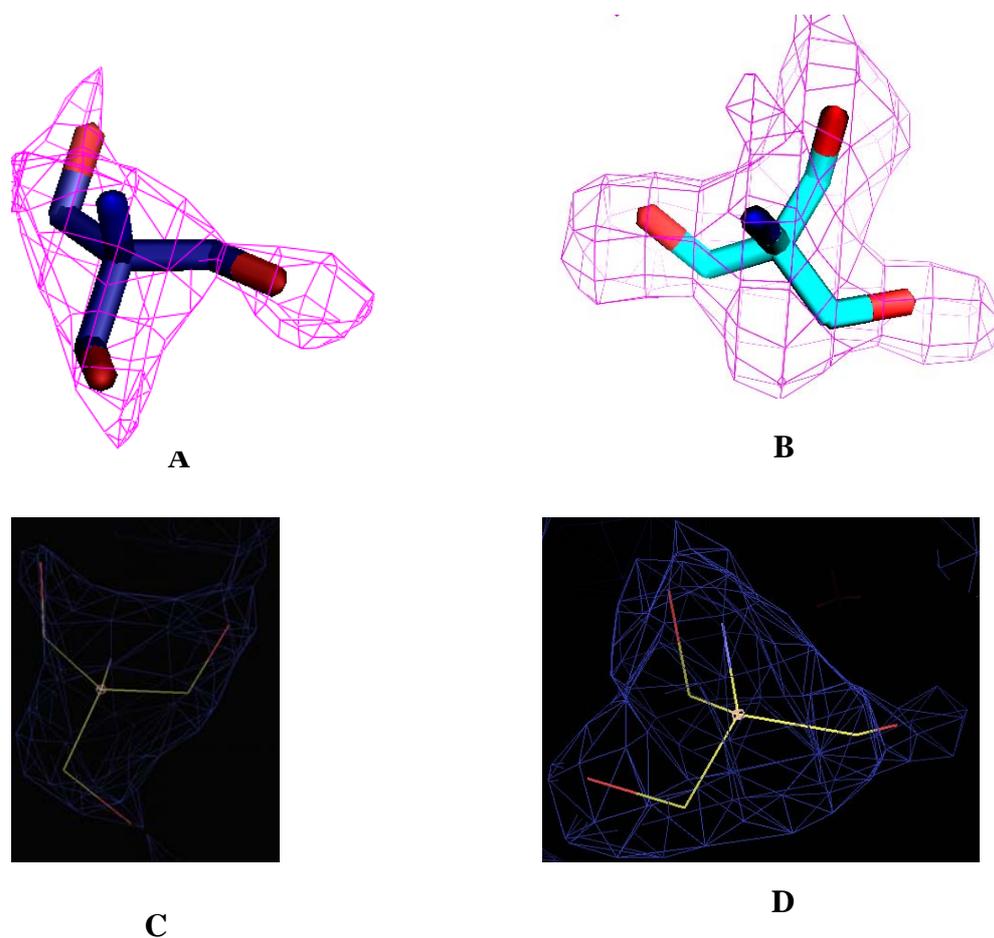
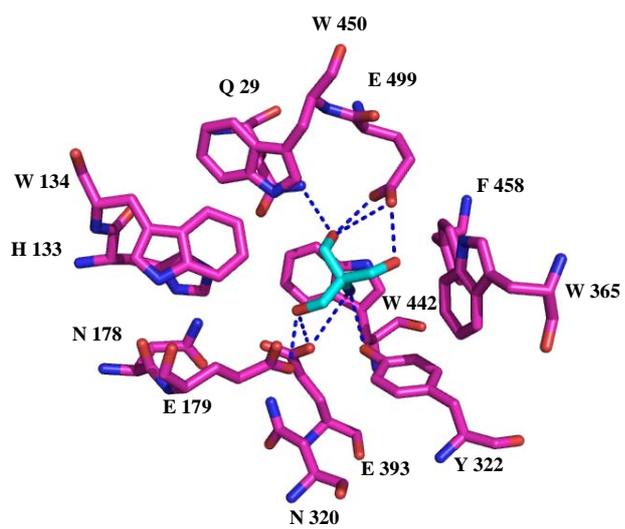
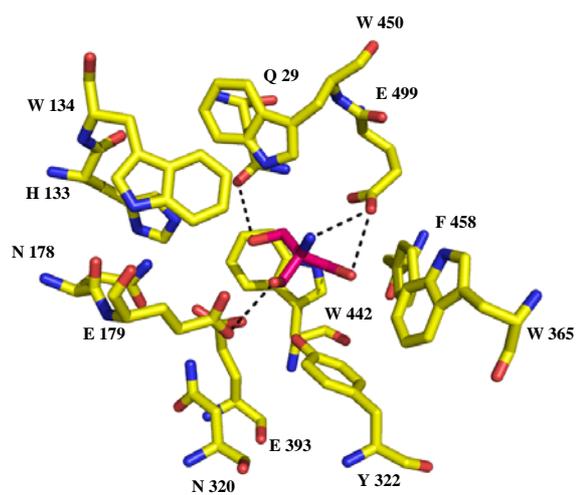


Figure 3.21 Two Tris molecules from A and B monomer with the surrounding electron density maps.

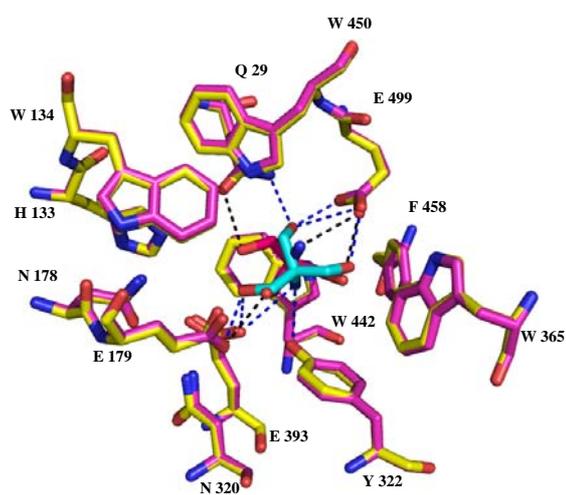
The corresponding sites in the $F_{\text{obs}}-F_{\text{calc}}$ OMIT maps are shown as the pink mesh, which indicates the 2σ height cutoff of both molecules represented in (A) and (B). (C) and (D), $2F_{\text{obs}}-F_{\text{calc}}$ residual electron density maps contoured at 1σ of molecule A and B, respectively. (A) and (B) were generated with Pymol (DeLano, 2002) and C and D with the Coot program (Emsley P. and Cowtan K, 2004).



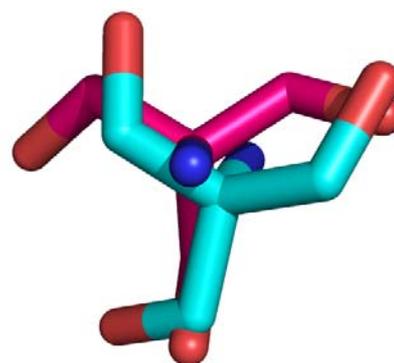
A



B



C



D

Figure 3.22 Stereo view of active site region of free Os4BGlu12 with Tris bound. Tris molecules are shown in the active site of A and B. (A) and (B) show the active site-Tris binding interactions. (C) Superimposition of freeOs4BGlu12 of the molecule A in pink and B in yellow. (D), superimposition of Tris from the A and B monomers and comparison of the conformation of Tris. The amino acids surrounding Tris are represented by sticks with carbons of monomer A in pink and those of B in yellow, nitrogens in blue and oxygens in red (of both molecules). The Tris ligand in each conformation is represented by sticks with carbons of molecule A in pink and those of molecule B in yellow, oxygen in red and nitrogen in blue of both molecules. Hydrogen bonding interactions between Tris and amino acids are shown as blue (A molecule) and black (B molecule) dashed lines. The figure was generated by Pymol (DeLano, 2002).

Table 3.4 Hydrogen bonding interaction between two conformations of Tris molecules and amino acid residues in the active site of enzyme.

Residues	Tris(A)	H bondong distance (Å)	Residues	Tris (B)	H bondong distance (Å)
Glu179 Oε2	O3	2.73	Glu179 Oε2	O1	2.97
	O3	2.75		O3	3.09
Glu393 Oε2	N1	3.09	Glu449 Oε2	N1	3.10
Glu449 Oε1	O2	2.58	Gln29 Oε1	O2	3.28
Glu449 Oε2	O2	3.48			

Glu449 Oε1	O1	2.71
Gln29 Nε2	O2	3.15
Tyr322 OH	N1	3.01

The structure of the crystal soaked in a solution containing G2F clearly revealed the presence of a covalent glucosidic bond between the anomeric carbon of G2F and the Glu393 nucleophile residue (Figure 3.23). Besides the two catalytic residues Glu179 and Glu393, the amino acids forming the glucose binding site (subsite -1) of Os4BGlu12, including Gln29, His133, Trp134, Asn178, Asn320, Tyr322, Trp365, Trp450, Phe458, Gln447 and Glu499, are highly conserved in GH1 enzymes. The recognition of the ⁴C₁ chair conformation of the glucose ring of G2F at subsite -1 is mediated by seven hydrogen bonds (Figure 3.24). O3 of G2F form hydrogen bonds with Trp450 Nε1 at a distance of 3.10 Å. Nε1 of His133 is hydrogen bonded to O3 of G2F at a distance of 2.81 Å. O3 of G2F also interacts with Gln29 Oε1 (2.65 Å), while O4 forms hydrogen bond with Gln29 Nε2 (2.91 Å). Oε1 and Oε2 of Glu449 form hydrogen bonds with O4 (2.57 Å) and O6 (2.90 Å) of G2F, respectively (Table 3.5). A water molecule (wat320) has a hydrogen bond to Glu179 Oε1 (3.01 Å), in this complex structure, while Glu179 of the free enzyme structure formed a hydrogen bond with Tris molecule. The distances (1.36 Å) and angle (146.16°) between Glc1 of sugar anomeric carbon and Glu393 Oε1, is similar to *S*-glycosidase *Sinapis alba* myrosidase (1E73) (Burmeister *et al.*, 2000).

A hydrophobic environment is present in the aglycone part of the active site. This hydrophobic pocket is formed by residues Phe193, Trp186, Trp181, Trp365 and Trp367. The polar residues found in the aglycone part include Asn186,

Asn325, His252 and Thr183. Trp358 in rice BGlu1, which corresponds to Trp365 in Os4BGlu12, had sticking interactions with Glc2 and Glc3 of celooligosaccharides at subsites +1 and +2 (Chuenchor, 2007). His252 and Asn186 in Os4BGlu12 correspond to Asn254 and Leu183 in rice BGlu1.

Table 3.5 Hydrogen bonding interaction between G2F and amino acid residues in the active site of enzyme.

Residues	G2F	H bonding distance (Å)
Glu449 Oε1	O4	2.57
Glu449 Oε2	O6	2.90
Gln29 Oε1	O3	2.65
Gln29 Nε2	O4	2.90
His133 Nε1	O3	2.81
Trp450 Nε1	O3	3.10

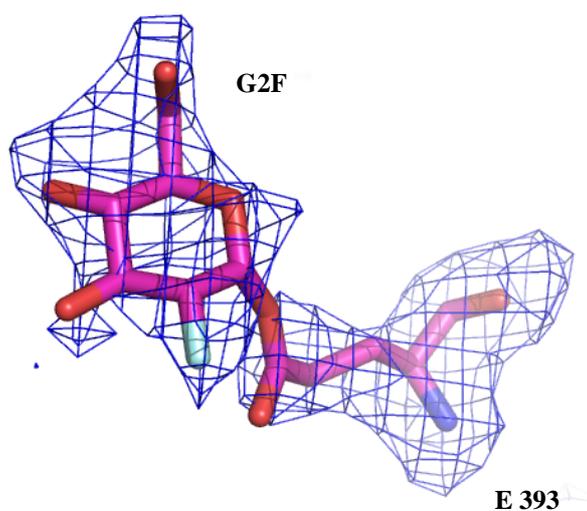


Figure 3.23 Structure of covalent glycosyl-enzyme intermediate.

The electron density of the $F_{\text{obs}}-F_{\text{calc}}$ OMIT map is contoured at the 3σ level. The sugar ring is shown as sticks with carbon in pink, oxygen in red, and fluorine in bright green. The nucleophile (Glu393) is represented by sticks with carbon in pink, nitrogen in dark blue, and oxygen in red. The figure was generated by Pymol (DeLano, 2002).

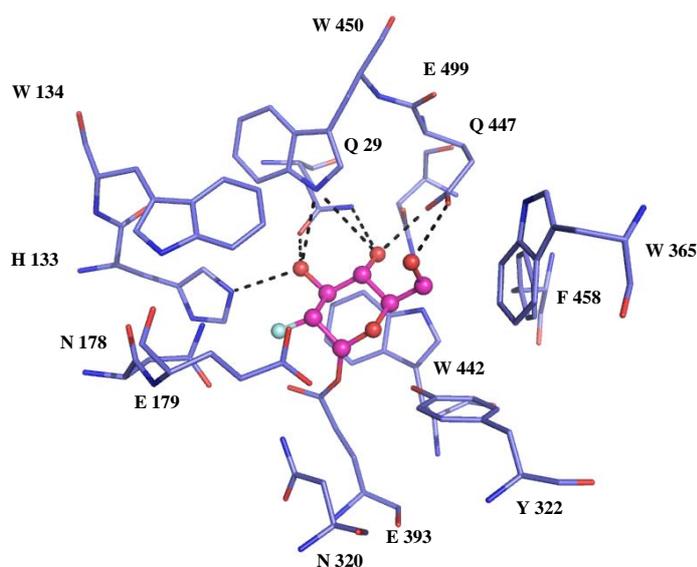


Figure 3.24 Stereo view of protein-ligand interaction in the active site of the Os4BGlu12-G2F complex.

The amino acids surrounding the -1 subsite are represented by sticks with carbon in blue, nitrogen in dark blue and oxygen in red. The covalent G2F complex is represented by balls and sticks with carbon in pink, oxygen in red and fluorine in blue. Hydrogen bonding interactions between G2F and amino acids are shown as black dotted lines. The figure was generated by Pymol (DeLano, 2002).

- Barrett, T., Suresh, C.G., Tolley, S.P., Dodson, E.J., and Hughes, M.A. (1995). The crystal structure of a cyanogenic β -glucosidase from white clover, a family 1 glycosyl hydrolase. **Structure**. 3: 951-960.
- Bergfors, T.M. (1999). Buffers and pH. Protein crystallization: **Techniques, strategies, and tips: A laboratory manual**. pp 54-61.
- Bolam, D.N., Hughes, N., Virden, R., Lakey, J.H., Hazlewood, G.P., Henrissat, B., Braithwaite, K.L., and Gilbert, H.J. (1996). Mannanase A from *Pseudomonas fluorescens* subsp. *cellulosa* is a glycosyl hydrolase in which E212 and E320 are the putative catalytic residues. **Biochemistry**. 35: 16195-16204.
- Bradford, M.M. (1976). A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. **Anal. Biochem.** 72: 264-273.
- Brzobohaty, B., Moore, I., Kristoffersen, P., Bako, L., Campos, N., Schell, J., and Palme, K. (1993). Release of active cytokinin by a β -glucosidase localized to the maize root meristem. **Science**. 262: 1051-1054.
- Burmeister, W.P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S., and Henrissat, B. (1997). The crystal structures of *Sinapsis alba* myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase. **Structure**. 5: 663-675.
- Burmeister, W.P., Cottaz, S., Rollin, P., Vasella, A., and Henrissat, B. (2000). High resolution X-ray crystallography shows that ascorbate is a cofactor for myrosinase and substitutes for the function of catalytic base. **J. Biol. Chem.** 275: 30385-39393.

- Chayen, N.E., Shaw Stewart, P.D., and Blow, D.M. (1992). Microbatch Crystallization Under Oil - a New Technique Allowing Many Small-volume Crystallization Trials. *Journal of Crystal Growth*. **J. Cryst. Growth**. 122: 76-180.
- Chi, Y.I., Martinez-Cruz, L.A., Jancarik, J., Swanson, R.V., Robertson, D.E., and Kim, S.H. (1999). Crystal structure of the β -glucosidase from hyperthermophile *Thermosphaera aggregans*: insights into its activity and thermostability. **FEBS Lett.** 445: 375-383.
- Chuenchor, W. (2007). **Structure-function Relationships on Rice BGlu1 β -glucosidase**. Ph.D. Dissertation, Suranaree University of Technology, Nakhon Ratchasima, Thailand.
- Chuenchor, W., Pengthaisong, S., Robinson, R.C., Yuvaniyama, J., Oonant, W., Bevan, D.R., Esen, A., Chen, C.J., Opassiri, R., Svasti, J., and Ketudat-Cairns, JR. (2008). Structural insights into rice BGlu1 β -glucosidase oligosaccharide hydrolysis and transglycosylation. **J. Mol. Biol.** 377: 1200-1215.
- Coutinho, P.M. and Henrissat, B. (1999). **Carbohydrate active enzymes: an integrated database approach. In Recent Advances in Carbohydrate Bioengineering** (Gilbert, H.J., Davies, G., Henrissat, B. and Svensson, B., eds.), pp. 3-12, The Royal Society of Chemistry, Cambridge.
- Czjzek, M., Cicek, M., Zamboni, V., Burmeister, W.P., Bevan, D.R., Henrissat, B., and Esen, A. (2000). The mechanism of substrate (aglycone) specificity in β -glucosidases is revealed by crystal structures of mutant maize

- β -glucosidase-DIMBOA, -DIMBOAGlc, and -dhurrin complexes. **Proc. Natl. Acad. Sci. USA.** 97: 13555-13560.
- Czjzek, M., Cicek, M., Zamboni, V., Burmeister, W.P., Bevan, D.R., Henrissat, B., and Esen, A. (2001). Crystal structure of a monocotyledon (maize ZMGluc1) β -glucosidase and a model of its complex with *p*-nitrophenyl β -D-thioglucoside. **Biochem. J.** 354: 37-46.
- Delano, W.L. (2002). **The Pymol Molecular Graphics System.** (Online). Available: <http://www.pymol.org.html>.
- Dharmawardhana, D.P., Ellis, B.E., and Carlson, J.E. (1995). A β -glucosidase from lodgepole pine specific for the lignin precursor coniferin. **Plant Physiol.** 107: 331-339.
- Dietz, K.J., Sauter, A., Wichert, K., Messdaghi, D., and Hartung, W. (2000). Extracellular β -glucosidase activity in barley involved in the hydrolysis of ABA glucose conjugate in leaves. **J. Exp. Biol.** 51: 937-944.
- Duroux, L., Delmotte, F.M., Lancelin, J.M., Keravis, G., and Jay-Alleand, C. (1998). Insight into naphthoquinone metabolism: β -glucosidase-catalysed hydrolysis of hydrojuglone β -D-glucopyranoside. **Biochem. J.** 333: 275-283.
- Emsley, P. and Cowtan, K. (2004). Coot-Model-building tools for molecular graphics. **Acta. Cryst. D60:** 2126-2132.
- Esen, A. (1992), Purification and partial characterization of maize (*Zea mays* L.) β -glucosidase. **Plant Physiology.** 98: 174-188.
- Esen, A. (1993). B-Glucosidase: Overview in *β -Glucosidases: Biochemistry and Molecular Biology* (Esen, A., ed) pp.1-26, **American Chemical Society**, Washington, DC.

- Guasch, A., Vallmitjana, M., Perez, R., Querol, E., Perez-Pons, J.A., and Coll, M. (1999). Cloning, overexpression, crystallization and preliminary X-ray analysis of a family 1 beta-glucosidase from *Streptomyces*. **Acta Crystallog. sect. D**, 55: 679-682.
- Gouet, P., Courcelle, E., Stuart, D.I. and Metz, F. (1999). **Bioinformatics**. 15: 305-308.
- Hakulinen, N., Paavilainen, S., Korpela, T., and Rouvinen, J. (2000). The crystal structure of β -glucosidase from *Bacillus circulans* sp. *alkalophilus*: ability to form long polymeric assemblies. **J. Struct. Biol.** 129: 69-79.
- Henrissat, B. and Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. **Biochem. J.** 293: 781-788.
- Henrissat, B. and Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. **Biochem. J.** 316: 695-696.
- Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J-P., and Davies, G. (1995). Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. **Proc. Natl. Acad. Sci. USA.** 92: 7090-7094.
- Henrissat, B. and Davies, G. T. (1997). Structural and sequence-based classification of glycoside hydrolases. **Curr. Op. Struct. Biol.** 7: 637-644.
- Hrmova, M., Harvey, A.J., Wang, J., Shirley, N.J., Jones, G.P., Stone, B.A., Hoj, P.B., and Fincher, G.B. (1996). Barley β -D-glucan exohydrolases with β -D-glucosidase activity. Purification, characterization, and determination of primary structure from a cDNA clone. **J. Biol. Chem.** 271: 5277-5286.

- Husebye, H., Arzt, S., Burmeister, W.P., Haertel, F.V., Brandt, A., Rossiter, J. T., and Bones, A.M. (2005). Crystal structure at 1.1 Å resolution of an insect myrosinase from *Brevicoryne brassicae* shows its close relationship to beta-glucosidases. **Insect. Biochem. Mol. Biol.** 35: 1311- 1320.
- Isorna, P., Polaina, J., Latorre-García, L., Cañada, F.J., González, B., and Sanz-Aparicio, J. (2007). Crystal structures of *Paenibacillus polymyxa* β-glucosidase B complexes reveal the molecular basis of substrate specificity and give new insights into the catalytic machinery of family 1 glycosidases. **J. Mol. Biol.** 371: 1204-1218.
- Iwami, K. and Yasumoto, K. (1986). Synthesis of pyridoxine β-glucosidase and its *in situ* absorption in rat small intestine. **Nutr. Res.** 6: 407-414.
- Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G., and Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. **Trends Biochem. Sci.** 23: 403-405.
- Jenkins, J., Leggio, L.L., Harris, G., and Pickersgill, R. (1995). β-Glucosidase, β-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold β/α architecture and with two conserved glutamates near the carboxy-terminal ends of β-strands four and seven. **FEBS Lett.** 362: 281-285.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. **Nature.** 227: 680-685.
- Leah, R., Kigel, J., Svedsen, I., and Mundy, J. (1995). Biochemical and molecular characterization of a barley seed β-glucosidase. **J. Biol. Chem.** 270: 15789-15797.

- Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. **J. Appl. Cryst.** 26: 283-291.
- Maneesan, J. (2007). **Study of the natural substrate structure of rice β -glucosidases** Master's degree Dissertation, Suranaree University of Technology, Nakhon Ratchasima, Thailand.
- Marana, S.R. (2006). Molecular basis of substrate specificity in family 1 glycoside hydrolases. **IUBMB Life.** 58: 63-73.
- McCarter, J. and Withers, S.G. (1994). Mechanisms of enzymatic glycoside hydrolysis. **Curr. Opin. Struc. Bio.** 4: 885-892.
- Menegus, F., Cattaruzza, L., Ragg, E., and Scaglione, L. (1995). R(-)-pantoyllactone- β -D-glucopyranoside: characterization of a metabolite from rice seedlings. **Phytochemistry.** 40: 1617-1621.
- Mizutani, M., Nakanishi, H., Ema, J., Ma, S., Noguchi, E., Inohara-Ochiai, M., Fukuchi-Mizutani, M., Nakao, M., and Sakata, K. (2002). Cloning of β -primeverosidase from tea leaves, a key enzyme in tea aroma formation. **J. Plant Physiol.** 130: 2164-2176.
- Nijikken, Y., Tsukada, T., Igarashi, K., Samejima, M., Wakagi, T., Shoun, H., and Fushinobu, S. (2007). Crystal structure of intracellular family 1 β -glucosidase BGL1A from the basidiomycete *Phanerochaete chrysosporium*. **FEBS. Lett.** 581: 1514-1520.
- Noguchi, J., Hayashi, Y., Baba, Y., Zkino, O., and Kimura, M. (2008). Crystal structure of the covalent intermediate of human cytosolic β -glucosidase. **Biochem. Biophys. Res. Comm.** 374: 549-552.

- Opassiri, R., Hua, Y., Wara-Aswapati, O., Akiyama, T., Svasti, J., Esen, A., and Ketudat-Cairns, J. R. (2004). β -glucosidase, exo- β -glucanase and pyridoxine transglucosylase activities of rice BGlu1. **Biochem. J.** 379: 125-131.
- Opassiri, R., Ketudat-Cairns, J. R., Akiyama, T., Wara-Aswapati, O., Svasti, J., and Esen, A. (2003). Characterization of a rice β -glucosidase genes highly expressed in flower and germinating shoot. **Plant Sci.** 165: 627-638.
- Opassiri, R., Pomthong, B., Onkoksoong, T., Akiyama, T., Esen, A., and Ketudat Cairns, J. R. (2006). Analysis of rice glycosyl hydrolase family 1 and expression of Os4bglu12 β -glucosidase. **BMC Plant Biol.** 6: 33.
- Otwinowski, Z. and Minor, W., (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode, **Methods in Enzymology**. Academic Press, New York Volume 276: 307-326.
- Pomthong, B. (2007). **Recombinant protein expression and functional characterization of β -glucosidase and β -glucanase from rice.** Ph.D. Dissertation, Suranaree University of Technology, Nakhon Ratchasima, Thailand.
- Poulton J.E. (1990). Cyanogenesis in plants. **Plant Physiol.** 94: 401-405.
- Rajan, S.S., Yang, X., Collart, F., Yip, V.L., Withers, S.G., Varrot, A., Thompson, J., Davies, G.J., and Anderson, W.F. (2004). Novel catalytic mechanism of glycoside hydrolysis based on the structure of an $\text{NAD}^+/\text{Mn}^{2+}$ -dependent phospho-alpha-glucosidase from *Bacillus subtilis*. **Structure.** 12: 1619-1629.
- Reese, E.T. (1977). Degradation of polymeric carbohydrates by microbial enzymes. **Recent Advances Phytochemistry.** 11: 311-364. Quoted in Cicek, M., and

- Sanz-Aparicio, J., Hermoso, J.A., Martinez-Ripoll, M., Lequerica, J.L., and Polaina, J. (1998). Crystal structure of β -glucosidase A from *Bacillus polymyxa*: insights into the catalytic activity in family 1 glycosyl hydrolases. **J. Mol. Bio.** 275: 491-502.
- Schliemann, W. (1984). Hydrolysis of conjugated gibberellins by β -glucosidases from dwarf rice (*Oryza sativa* L. cv. Tan-ginbozu). **J. Plant Physiol.** 116: 123-132.
- Seshadri, S. (2009). **Functional and structural studies of rice β -glucosidase**. Ph.D. Dissertation, Suranaree University of Technology, Nakhon Ratchasima, Thailand.
- Smith, A.R. and Van Staden, J. (1978). Changes in endogenous cytokine levels in kernels of *Zea mays* L. during inhibition and germination. **J. Exp. Bot.** 29: 1067-1073.
- Sue, M., Yamazaki, K., Yajima, S., Nomura, T., Matsukawa, T., Iwamura, H., and Miyamoto, T. (2006). Molecular and structural characterization of hexameric β -D-glucosidases in wheat and rye. **Plant. Physiol.** 141: 1237-1247.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. **Nucleic Acids Res.** 22: 4673-4680.
- Tribolo, S., Berrin, J.G., Kroon, P.A., Czjzek, M., and Juge, N. (2007). The Crystal structure of human cytosolic β -glucosidase unravels the substrate aglycone specificity of a family 1 glycoside hydrolase. **J. Mol. Biol.** 370: 964.
- Vagin, A. and Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. **J. Appl. Cryst.** 30: 1022-1025.

- Varghese, J.N., Garret, T.P.J., Column, P.M., Chen, L., Hoj P.B., and Fincher, G. B. (1994). The three-dimensional structure of two plant- β -glycan endohydrolases with distinct substrate specificities. **Proceedings of the National Academy of Sciences, U.S.A.** 91: 2785-2789.
- Verdoucq, L., Czjzek, M., Moriniere, J., Bevan, D.R., and Esen, A. (2003). Mutation and structural analysis of aglycone specificity in maize and sorghum β -glucosidases. **J. Biol. Chem.** 278: 25055-25062.
- Verdoucq, L., Morinière, J., Bevan, D.R., Esen, A., Vasella, A., Henrissat, B., and Czjzek, M. (2004). Structural determinants of substrate specificity in family 1 β -glucosidases: Novel insights from the crystal structure of sorghum dhurrinase-1, a plant β -glucosidase with strict specificity, in complex with its natural substrate. **J. Biol. Chem.** 279: 31796-31803.
- Wang, Q., Graham, R.W., Trimbur, D., Warren, R.A.J., and Withers, S.G. (1994). Changing enzymatic reaction mechanisms by mutagenesis: Conversion of a retaining glucosidase to an inverting enzyme. **J. Am. Chem. Soc.** 116:11594-1595.
- Warren, R.A.J. (2000) Engineering β -glucoside hydrolase. In Alberghina L. (ed.). **Protein engineering in industrial biotechnology.** (pp 169-188). Harwood Academic Publishers.
- Weber, P.C. (1997). **Methods Enzymol.** 276: 13-23.
- Wiesmann, C., Hengstenberg W., and Schuz, G.E. (1997). Crystal structure and mechanism of 6-phospho- β -galactosidase from *Lacococcus latis*. **J. Mol. Biol.** 269: 815-860.

- Withers, S.G. (2001). Mechanisms of glycosyl transferases and hydrolases. **Carbohydr. Pol.** 44: 325-337.
- Withers, S.G. and Street, I.P. (1988). Identification of a covalent α -D-glucopyranosyl enzyme intermediate formed on a β -glucosidase. **J. Am Chem. Soc.** 110: 8551-8553.
- Withers, S.G., Warren, R.A.J., Street, I.P., Rupitz, K., Kempton, J.B., and Aebersold, R. (1990). Unequivocal demonstration of the involvement of a glutamate residue as a nucleophile in the mechanism of a "retaining" glycosidase. **J. Am Chem. Soc.** 112: 5887-5889.
- Zechel, D.L. and Withers, S.G. (2000). Glycosidase mechanisms: anatomy of a finely tuned catalyst. **Acc. Chem. Res.** 33: 11-18.
- Zechel, D.L., Boraston, A.B., Gloster, T., Boraston, C.M., Macdonald, J.M., Tilbrook, D.M., Stick, R.V., and Davies, G.J. (2003). Iminosugar glycosidase inhibitors: structural and thermodynamic dissection of the binding of isofagomine and 1-deoxynojirimycin to β -glucosidases. **J. Am. Chem. Soc.** 125: 14313-14323.
- Zouhar, J., Nanak, E., and Brzobohatý, B. (1999). Expression, single-step purification, and matrix-assisted refolding of a maize cytokinin glucoside-specific β -glucosidase. **Prot. Expr. Purif.** 17: 153-162.
- Zouhar, J., Vevodova, J., Marek, J., Damborsky, J., Su, X.D., and Brzobohaty, B. (2001). Insights into the functional architecture of the catalytic center of a maize β -glucosidase Zm-p60.1 **Plant Physiol.** 127: 973-985.

CURRICULUM VITAE

NAME Mr Sompong Sansenya

DATE OF BIRTH 21 July 1982

PLACE OF BIRTH Roi-Et, Thailand

INSTITUTION ATTENDED

Maejo University, Chiang Mai, Thailand, 2001-2005,

Bachelor of Education, Chemistry,