

**CHARACTERIZATION AND STRUCTURAL STUDIES
OF RICE β -GALACTOSIDASE**

Thipwarin Rimlumduan

**A Thesis Submitted in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy in Biochemistry**

Suranaree University of Technology

Academic Year 2013

การศึกษาสมบัติและโครงสร้างของเอ็นไซม์เบต้ากาแลคโตซิเดสจากข้าว

นางสาวทิพย์วรินทร์ รีมลาควน

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาชีวเคมี
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2556

CHARACTERIZATION AND STRUCTURAL STUDIES OF RICE β -GALACTOSIDASE

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

Thesis Examining Committee

(Assoc. Prof. Dr. Jatuporn Wittayakun)

Chairperson

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

Member (Thesis Advisor)

(Dr. Pakorn Wattana-Amorn)

Member

(Dr. Waraporn Tantanuch)

Member

(Asst.Dr. Jaruwan Siritapetawee)

Member

(Prof. Dr. Sukit Limpijumnong)

Vice Rector for Academic Affairs
and Innovation

(Assoc. Prof. Dr. Prapun Manyum)

Dean of Institute of Science

ทิพย์วรินทร์ ริมล้าควน : การศึกษาสมบัติและโครงสร้างของเอนไซม์เบตากาแลคโตซิเดส จากข้าว (CHARACTERIZATION AND STRUCTURAL STUDIES OF RICE β -GALACTOSIDASE อาจารย์ที่ปรึกษา : ศาสตราจารย์ ดร.เจมส์ เกตุทัต-คาร์นส์, 185 หน้า

เอนไซม์เบตากาแลคโตซิเดสจากพืชถูกจำแนกให้อยู่ในตระกูลของไกลโคไซด์ไฮโดรเลส กลุ่มที่ 35 ซึ่งเอนไซม์นี้พบในพืชหลายชนิดและมีส่วนทางด้านปลายคาร์บอกซิลิกที่มีหน้าที่คล้ายกับเลคตินจากหอยเม่นทะเลที่ทำหน้าที่ในการจับกับน้ำตาลกาแลคโตส แม้ว่าปัจจุบันบทบาทหน้าที่ในการจับกับคาร์โบไฮเดรตของส่วนทางด้านปลายคาร์บอกซิลิกของเอนไซม์กลุ่มนี้ได้มีการศึกษาไปบ้างแล้ว แต่เพื่อเข้าใจบทบาทหน้าที่และโครงสร้างทางด้านปลายคาร์บอกซิลิกของเอนไซม์กลุ่มนี้ที่มาจากข้าว ดีเอ็นเอคู่สมของยีนเบตากาแลคโตซิเดสส่วนทางด้านปลายคาร์บอกซิลิก (OsBGal1 Cter) ถูกเพิ่มจำนวนโดยเทคนิคปฏิกิริยาลูกโซ่พอลิเมอไรเซชันและทำการโคลนเข้าเวกเตอร์ pET32b(+) โปรตีน OsBGal1 Cter ถูกผลิตโดยทั้งแบบไม่ติดฉลาก และแบบติดฉลากไอโซโทป ^{15}N หรือ ^{13}C หรือทั้ง ^{15}N และ ^{13}C โปรตีน OsBGal1 Cter มีน้ำหนักโมเลกุล 33 กิโลดาลตัน ส่วนของโปรตีนไทโอรีดอกซินและบริเวณที่มีกรดอะมิโนฮิสติดีนเรียงต่อกันอยู่ที่ส่วนทางด้านปลายอะมิโนของโปรตีนถูกตัดออกด้วยเอนไซม์ทรอมบิน โปรตีน OsBGal1 Cter ถูกทำให้บริสุทธิ์ด้วยคอลัมน์ IMAC 2 ครั้ง และตามด้วยคอลัมน์เบนซามิไดน จากการตรวจสอบน้ำหนักโมเลกุลของ OsBGal1 Cter โปรตีนที่ถูกทำให้เสียสภาพ มีน้ำหนักประมาณ 13 กิโลดาลตัน และเมื่อตรวจสอบน้ำหนักโมเลกุลแบบธรรมชาติของโปรตีนตัวนี้เท่ากับ 15 กิโลดาลตัน ซึ่งแสดงให้เห็นว่าโปรตีนตัวนี้เป็นโมเลกุลเดี่ยวในสารละลาย ทำการตรวจหาโครงสร้างหลักของโปรตีนโดยวิธี 3D HNCO CBCA(CO)NH และ HNCACB นิวเคลียร์แมกเนติกเรโซแนนซ์ และส่วนของหมู่โซ่ข้างของ OsBGal1 Cter โดยวิธี C(CO)NH และ HCCH-TOCSY นิวเคลียร์แมกเนติกเรโซแนนซ์ โครงสร้างทุติยภูมิของ OsBGal1 Cter ประกอบด้วยแผ่นเบตา 5 แผ่น และเกลียวอัลฟา 1 เกลียว โครงสร้างสามมิติของโปรตีนชนิดนี้คล้ายกับโครงสร้างของส่วนของโปรตีนที่ทำหน้าที่จับกับคาร์โบไฮเดรตจากสัตว์ แต่มีความแตกต่างกันในส่วนของวงรูป โดยที่วงรูปเอ และวงรูปซีของโปรตีน OsBGal1 Cter นั้นมีความยาวกว่าวงรูปของแลโทรฟิลิน-1 จากหนู และส่วนของเลคตินจากปลาแซลมอน นอกจากนี้โครงสร้างของวงรูปเอของ OsBGal1 Cter ไม่สามารถระบุได้แน่ชัด ซึ่งให้เห็นว่าส่วนนี้เป็นบริเวณที่มีความยืดหยุ่นสูง ถึงแม้ว่าโปรตีน OsBGal1 Cter จะถูกทำนายว่าเป็นส่วนของเลคตินที่สามารถจับกับน้ำตาลกาแลคโตส และน้ำตาลแรมโนสได้ แต่ผลจากการทดสอบการจับกับน้ำตาลแรมโนส น้ำตาลกาแลคโตส น้ำตาลกลูโคส และน้ำตาลราฟฟิโนส โดยวิธีเอชเอสคิวซี นิวเคลียร์แมกเนติกเรโซแนนซ์ พบว่า โปรตีน OsBGal1 Cter ไม่สามารถจับกับน้ำตาลที่กล่าวมาได้

ส่วนการทดสอบการจับกันระหว่างโปรตีน OsBGal1 Cter กับน้ำตาลจำพวกโอลิโกแซคคาไรด์ และโพลีแซคคาไรด์ โดยวิธีการโโบไฮเดรตไมโครแอเรย์นั้นพบว่า โปรตีนนี้สามารถจับกับ น้ำตาลอะราบินโนไตรโอส และน้ำตาลกาแลคโตไบโอส ในขณะที่ผลการทดลองจากวิธีเอสทีดี นิวเคลียร์แมกเนติกเรโซแนนซ์ พบว่าโปรตีน OsBGal1 Cter ไม่สามารถจับกับน้ำตาลตัวนี้ได้

ในการศึกษาโครงสร้างสามมิติของเอนไซม์เบตากาแลคโตซิเดสแบบเต็มโมเลกุลนั้น เอนไซม์ตัวนี้ถูกพัฒนาโดยการเปลี่ยนรหัสโคดอนของดีเอ็นเอคู่สมให้เหมาะสมต่อการแสดงออก ของเอนไซม์ตัวนี้ในเชื้อยีสต์ *Pichia pastoris* การแสดงออกของเอนไซม์ตัวนี้ถูกเหนี่ยวนำด้วย เมทานอลความเข้มข้น 1 เปอร์เซ็นต์ ที่อุณหภูมิ 20 องศาเซลเซียส เรียกชื่อใหม่ว่า OsBGal1opt เอนไซม์เบตากาแลคโตซิเดสที่ได้มีขนาดน้ำหนักโมเลกุล 97 กิโลดาลตัน และมีการเติม คาร์โบไฮเดรตที่ตำแหน่งในโครเจน 2 ตำแหน่ง เอนไซม์นี้ทำงานได้ดีที่สุดที่พีเอช 4.5 และอุณหภูมิ ที่เหมาะสมอยู่ที่ 55 องศาเซลเซียส

THIPWARIN RIMLUMDUAN : CHARACTERIZATION AND
STRUCTURAL STUDIES OF RICE β -GALACTOSIDASE. THESIS
ADVISOR : PROF. JAMES R. KETUDAT-CAIRNS, Ph.D. 185 PP.

β -GALATOSIDASE/C-TERMINAL DOMAIN/NMR/RICE/CARBOHYDRATE
BINDING

Plant β -galactosidases are classified in glycoside hydrolase family 35 (GH 35). Many plant BGals have an additional C-terminal domain similar to galactose-binding lectin from sea urchin, although its role in carbohydrate-binding has only been speculated to date. To understand the function and structure of the C-terminal domain from rice β -galactosidase OsBGal1, the cDNA encoding the OsBGal1 C-terminal domain (OsBGal1 Cter) was amplified by PCR and cloned into pET32b(+). The recombinant OsBGal1 Cter was expressed with and without labeling with ^{15}N , ^{13}C or ^{15}N and ^{13}C . The OsBGal1 Cter fusion protein had a denatured molecular weight of approximately 33 kDa. The fusion protein was cleaved with thrombin protease to remove the N-terminal thioredoxin and His tags. The OsBGal1 Cter protein was purified by 2 steps of IMAC and benzamidine column. The free OsBGal1 Cter had a denatured molecular weight of approximately 13 kDa and an apparent native molecular weight of about 15 kDa, indicating that the free OsBGal1 Cter is a monomer in solution. The backbone assignments of OsBGal1 Cter were constructed from 3D HNC(O), CBCA(CO)NH and HNCACB nuclear magnetic resonance (NMR) spectra. Side chain peaks for the OsBGal1 Cter were assigned from C(CO)NH and HCCH-TOCSY spectra. NOESY spectra provided constraints for calculation of the 3-

dimensional structure. The secondary structure of OsBGal1 Cter had 5 β -stands and 1 α -helix. The structure of this domain was similar to carbohydrate binding domains from animals, but showed differences in the loops. Loops A and C of OsBGal1 Cter are longer than the corresponding loops from mouse latrophilin-1 and the chum salmon lectin domain. Loop A of OsBGal1 Cter was not well-defined, suggesting it is flexible. Although OsBGal1Cter was predicted to be a galactose/rhamnose-binding lectin, titration with rhamnose, galactose, glucose and raffinose showed no binding in the HSQC NMR spectra. OsBGal1Cter appeared to bind to α -(1,5)-L-arabinotriose and β -(1,5)-D-galactobiose, as well as several other oligosaccharides and polysaccharides, on a carbohydrate array. The OsBGal1 Cter binding to α -(1,5)-L-arabinotriose was tested by STD NMR, but no signs of binding were observed.

To study the whole β -galactosidase structure and function, rice OsBGal1 β -galactosidase was expressed from a codon-optimized cDNA in *Pichia pastoris*. Protein expression was induced 1% methanol at 20°C to yield the recombinant enzyme, designated OsBGal1opt. This enzyme had an apparent molecular mass of 97 kDa and a slight smearing of the band to upper molecular weight suggested the protein was glycosylated at one or both of the two putative N-glycosylation sites observed in the OsBGal1 sequence. The optimum pH for OsBGal1 expressed in this system was found to be 4.5 and the optimum temperature was at 55°C.

School of Biochemistry

Student's Signature_____

Academic Year 2013

Advisor's Signature_____

ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my advisor Prof. Dr. James R. Ketudat Cairns for providing me the opportunity to study towards my Ph.D. degree in Biochemistry, and his kind advice, continuous guidance, and encouragement. I would also like to express my deepest gratitude to Prof. Dr. Toshiyuki Tanaka at the University of Tsukuba, Ibaraki, Japan for providing me the chance to learn NMR and valuable technical skills, and his precious advice, including solve problems about structure determination. I would like to express special thanks Dr. Yanglin Hua for help and advise on STD and HSQC methods in NMR.

I would like to thank Assoc. Prof. Dr. Jatuporn Wittayakun, Dr. Pakorn Watana-Amorn, Dr. Waraporn Tanthanuch and Asst. Prof. Dr. Jaruwat Siritapetawee for patiently reading this dissertation and providing helpful comments.

Special thanks are extended to all my friends in the School of Biochemistry, Suranaree University of Technology for their help.

Finally, my sincerest gratitude is to my family for their love, encouragement, and support, without which I would not have been able to complete this dissertation and obtain the doctorate degree.

This work was supported by the Commission on Higher Education, Suranaree University of Technology, Thailand and the University of Tsukuba, Japan.

Thipwarin Rimlumduan

CONTENTS

	Page
ABSTRACT IN THAI.....	I
ABSTRACT IN ENGLISH	III
ACKNOWLEDGEMENTS	V
CONTENTS	VI
LIST OF TABLES	XIII
LIST OF FIGURES	XIV
LIST OF ABBREVIATIONS AND SYMBOLS.....	XIX
CHAPTER	
I INTRODUCTION	1
1.1 Glycoside hydrolases.....	1
1.2 Glycosidase mechanism	2
1.2.1 Inverting mechanism.....	2
1.2.2 Retaining mechanism.....	3
1.3 Plant Glycoside Hydrolases	5
1.4 β -galactosidase	6
1.5 Rice β -galactosidases	9
1.6 Cell wall structure and metabolism.....	10
1.7 Recombinant expression of plant β -galactosidases.....	11
1.8 Three-dimensional (3D) structures of GH35 β -galactosidase	13

CONTENTS (Continued)

	Page
1.9 Carbohydrate binding domains	17
1.10 Structure of galactose binding-lectin link domains.....	21
1.10.1 Chum salmon rhamnose binding lectin (RBL) structure	21
1.10.2 The mouse latrophilin RBL structure determined by NMR.....	
investigation by NMR.....	27
1.11.1 The ¹ H- ¹⁵ N Heteronuclear Single-Quantum Correlation	
(HSQC) experiment	27
1.11.2 HNCOC experiment	30
1.11.3 CBCA(CO)NH experiment	31
1.11.4 HNCACB experiment.....	32
1.11.5 C(CO)NH experiment.....	33
1.11.6 HCCH-TOCSY experiment	33
1.11.7 ¹⁵ N edited NOESY.....	34
1.11.8 ¹³ C edited NOESY	35
1.11.9 Three-dimensional model generation	36
1.12 Saturation Transfer Difference (STD) NMR	37
1.13 Carbohydrate microarrays	40
1.14 Objectives	41
II MATERIALS AND METHODS.....	42
2.1 Materials	42
2.1.1 Oligonucleotides primers and optimized OsBGal1 gene.....	42

CONTENTS (Continued)

	Page
2.1.2 Plasmids, bacterial and yeast strains.....	43
2.1.3 Chemicals and reagents.....	47
2.2 General methods for <i>Escherichia coli</i>	48
2.2.1 Amplification of <i>OsBGal1 Cter</i> cDNA	48
2.2.2 Purification of PCR products from agarose gel.....	49
2.2.3 Quantification and expected yield of DNA.....	50
2.2.4 Cloning of DNA fragments into pENTR/D-TOPO vector.....	50
2.2.5 Ligation of DNA fragments into vectors	51
2.2.6 Preparation of competent <i>Escherichia coli</i> cells	51
2.2.7 Transformation and selection of <i>Escherichia coli</i>	52
2.3 Expression of OsBGal1 Cter protein in <i>E. coli</i>	53
2.4 Purification of OsBGal1 Cter protein.....	54
2.4.1 Thioredoxin fusion protein.....	54
2.4.2 Removal of the fusion tag and purification of tag free protein	55
2.5 Protein analysis	56
2.5.1 Sodium Dodecyl Sulfate Poly Acrylamide Gel Electrophoresis (SDS-PAGE).....	56
2.5.2 Bio-Rad protein assay	57
2.6 Determination of the native molecular weight (MW) of the OsBgal1 Cter protein by gel filtration column chromatography.....	57
2.7 Western immunoblot analysis.....	58

CONTENTS (Continued)

	Page
2.8 Protein crystallization.....	59
2.8.1 Preliminary sample preparation.....	59
2.8.2 Initial screening for crystallization conditions	59
2.9 Protein NMR.....	60
2.9.1 Preparation of OsBGal1 Cter sample for NMR	60
2.9.2 NMR data acquisition	60
2.9.3 NMR data processing and analysis.....	61
2.9.4 Structure calculation	61
2.9.5 OsBGal1 Cter titration	61
2.10 General methods for protein expression in yeast.....	62
2.10.1 DNA subcloning of the <i>optimized OsBGal1</i> cDNA.....	62
2.10.2 Preparation of competent <i>Pichia pastoris</i> strain SMD1168H.....	63
2.10.3 DNA preparation for electroporation.....	64
2.10.4 Expression of recombinant OsBGal1 in <i>P. pastoris</i>	65
2.10.5 Purification of recombinant OsBGal1	65
2.10.6 Determination of the optimal pH and temperature for recombinant OsBGal1.....	66
2.10.7 β -Galactosidase activity	66
2.11 Carbohydrate microarray.....	67

CONTENTS (Continued)

	Page
2.12 Detecting binding of OsBGal1 Cter with sugar by Saturation Transfer Difference (STD) method in NMR	71
III RESULTS	72
3.1 Recombinant expression of rice β -galactosidase C-terminal domain (OsBGal1 Cter)	72
3.1.1 Preparation of expression vector	72
3.1.2 Recombinant expression of rice OsBGal1 Cter protein.....	79
3.1.3 Purification of OsBGal1 Cter protein	81
3.2 Expression of OsBGal1 from a new construct with a convenient thrombin cleavage site.....	86
3.2.1 Recombinant plasmid production.....	86
3.2.2 Expression and purification of the new OsBGal1 Cter fusion protein	88
3.3 Determination of molecular weight of native OsBGal1 Cter by gel filtration on a Superdex S75 column.....	94
3.4 Determination structure of OsBGal1 Cter protein by NMR.....	97
3.4.1 Primary structure OsBGal1 Cter	97
3.4.2 Secondary structure of OsBGal1 Cter	98
3.4.3 Tertiary structure of OsBGal1 Cter	101
3.4.4 Overall structure comparison of OsBGal1 Cter with related animal lectins.....	106

CONTENTS (Continued)

	Page
3.5 Investigation of binding of OsBGal1 Cter protein to sugar.....	111
3.5.1 HSQC NMR method.....	111
3.5.1.1 Titration of OsBGal1 Cter with sugars	111
3.5.1.2 Stability of OsBGal1 Cter in HSQC experiments	118
3.5.2 Carbohydrate microarray	121
3.5.3 Saturation transfer difference (STD) method	123
3.6 Expression of OsBGal1 from cDNA optimized for <i>P. pastoris</i>	125
3.7 Protein purification of OsBGal1 from the optimized clone	130
3.8 Effect of pH and temperature on optimized OsBGal1 enzyme activity...	133
IV DISCUSSION	135
4.1 Protein expression and purification of OsBGal1 β -galactosidase C-terminal domain from rice	135
4.2 Protein crystallization.....	137
4.3 NMR structures of OsBGal1 Cter	137
4.3.1 Recombinant protein expression and purification for NMR.....	137
4.3.2 Determination of native molecular weight of OsBGal1 Cter.....	139
4.3.3 Comparison of the sequence of OsBGal1 Cter with related sequences	139
4.3.4 Protein structure determination by NMR.....	140
4.3.5 Relationship of structure to function	141

CONTENTS (Continued)

	Page
4.4 Expression of full length OsBGal1 β -galactosidase in <i>Pichia</i>	142
4.4.1 Sequence analysis of native and optimized OsBGal1 cDNA	142
4.5 Development of an expression system for active OsBGal1	146
4.6 Catalytic properties of OsBGal1	147
V CONCLUSION	149
REFERENCES.....	153
APPENDICES.....	167
APPENDIX A: PLASMID MAPS.....	168
1.1 pET32a(+) and pET32b(+) vector.....	168
1.2 pUC57	170
1.1 pPICZ α B vector.....	171
APPENDIX B NMR CHEMICAL SHIFTS FOR OsBGal1-Cter	172
CURRICULUM VITAE.....	185

LIST OF TABLES

Table	Page
1.1 CBMs classification based on fold	20
2.1 Primers for OsBGal1 Cter amplification	43
2.2 Recombinant plasmids used for OsBGal1 Cter cloning and expression	44
2.3 Bacterial and yeast strains.....	45
2.4 Cycling parameters for <i>OsBGal1 Cter</i> amplification.....	49
2.5 Primers for optimized OsBGal1 amplification	63
2.6 Cycling parameters for optimized OsBGal1 amplification	63
2.7 List of polysaccharides and oligosaccharides on carbohydrate array	69
3.1 Structural statistics of the 30 structures of the C-terminal domain of OsBGal1.....	104
3.2 Names of sugars to which OsBGal1 Cter appeared to bind in the array	122
4.1 The codon usage of <i>Escherichai coli</i>	144
4.2 The codon usage of <i>Pichai pastoris</i>	145

LIST OF FIGURES

Figure	Page
1.1 Catalytic mechanism of inverting enzyme.....	4
1.2 Enzymatic hydrolysis of β -D-galactosides by β -galactosidase	6
1.3 <i>Penicillium sp.</i> β -galactosidase (PDB ID 1TG7) is composed of five distinct structural domains	16
1.4 Comparison of the native structures of GH35 β -galactosidases. Psp- β -gal (PDB 1TG7), Tr- β -gal (PDB 3OG2) and Btm- β -gal (PDB 3D3A)	17
1.5 CBM distribution in different proteins	19
1.6 CSL3 dimer structure.....	23
1.7 Mouse lactophilin 1 RBL domain structure.....	25
1.8 Rhamnose binding by mouse latrophilin RBL domain.....	26
1.9 ^1H - ^{15}N HSQC spectrum of C-terminal domain of β -galactosidase 1 from rice.....	29
1.10 Schematic of HNCO resonances.....	30
1.11 Schematic of CBCA(CO)NH resonances	31
1.12 Schematic of NHCACB resonances	32
1.13 Schematic of ^{15}N edited NOESY resonances	34
1.14 Schematic of ^{13}C edited NOESY resonances.....	35
1.15 Schematic view of the STD experiment	37

LIST OF FIGURES (Continued)

Figure	Page
1.16 Schematic of the STD-NMR experiment.....	39
2.1 Spots sugar on the carbohydrate array membrane	68
3.1 The full-length cDNA sequence and deduced amino acid sequence of rice OsBGal1	73
3.2 Agarose gel electrophoresis of the cDNA fragment encoding OsBGal1 Cter that was amplified with OsBGal1 Cter_For primer and OsBGal1 Cter_Rev primer.....	75
3.3 Agarose gel electrophoresis of the pENTR/D-TOPO-OsBGal1_Cter plasmid.....	76
3.4 Agarose gel electrophoresis of the pET32a/DEST-OsBGal1_Cter plasmid.....	77
3.5 Agarose gel electrophoresis of pET32a/DEST_OsBGal1 Cter plasmids after digestion with <i>NcoI</i> and <i>EcoRI</i> restriction enzymes.....	78
3.6 Map of the protein sequence encoded by the recombinant pET32a/DEST with the <i>OsBGal1Cter</i> cDNA inserted after the enterokinase cleavage site.....	79
3.7 SDS-PAGE of pET32a/DEST-OsBGal1 Cter protein expressed in <i>E. coli</i> strains Origami(DE3) and Origami B(DE3)	80
3.8 SDS-PAGE of OsBGal1 Cter purified by immobilized Co ²⁺ affinity chromatography (IMAC)	82

LIST OF FIGURES (Continued)

Figure	Page
3.9 SDS-PAGE of OsBGal1 Cter after digestion with thrombin or enterokinase protease	83
3.10 SDS-PAGE of OsBGal1 Cter from the first expression construct after purification by 1 st and 2 nd IMAC steps	84
3.11 Elution profile of the free OsBGal1 1 Cter protein passed through an S75 gel filtration column and SDS-PAGE of fractions at the protein peak....	85
3.12 Agarose gel electrophoresis of OsBGal1 Cter cDNA fragment amplified with OsBGal1_Cter_NdeI_For and OsBGal1_Cter_BamHI_Rev primers.....	86
3.13 Agarose gel electrophoresis of pET32b-OsBGal1_Cter_jp_new plasmid	87
3.14 Map of the protein sequence encoded by the recombinant pET32b-OsBGal1 Cter_jp_new plasmid with the cDNA encoding OsBGal1Cter inserted after a thrombin cleavage site.....	87
3.15 SDS-PAGE of the OsBGal1 Cter_jp_new fusion protein purified by IMAC	89
3.16 The OsBGal1 Cter protein detected by SDS-PAGE and western immunoblot analysis	90
3.17 Time course of thrombin protease digestion of thioredoxin- OsBGal1 Cter fusion protein analyzed by SDS-PAGE.....	91
3.18 17% SDS-PAGE analysis of free OsBGal1 Cter purified by 2 nd immobilized Co ²⁺ affinity chromatography (IMAC) column	92

LIST OF FIGURES (Continued)

Figure	Page
3.19 SDS PAGE analysis of OsBGal1 Cter fractions from the steps of purification for structure determination by NMR.....	93
3.20 Elution profiles for estimation of molecular weight of the native form of the purified OsBGal1 Cter protein by Superdex S75 gel filtration chromatography	95
3.21 Estimation of molecular weight of the native form of the purified OsBGal1 Cter protein from the Superdex S75 gel filtration chromatography standard curve.....	96
3.22 Amino acid sequence comparison of OsBGal1 Cter., with other C-terminal domain from plant.....	99
3.23 Amino acid sequence comparison of OsBGal1 Cter., chum salmon (CSL3), sea urchin (SUEL) and mouse latrophilin 1 (Lphn1).....	100
3.24 The overall fold of the OsBGal1 Cter.....	102
3.25 The disulfide bridges of OsBGal1 Cter	103
3.26 Superimposition of the final 10 lowest energy structures of the OsBGal1 Cter.....	105
3.27 Superimposition of the structures of rice OsBGal1 Cter (blue), with other lectin domain of mouse latrophilin (2JXA, cyan) and chum salmon (2ZX4, green).....	107
3.28 Superimposition of the structures of the lectin domains of mouse latrophilin and chum salmon.....	109

LIST OF FIGURES (Continued)

Figure	Page
3.29 Residues use for binding to galactose from mouse latrophilin	110
3.30 ¹ H- ¹⁵ N HSQC NMR spectrum of OsBGal1 Cter in phosphate buffer in D ₂ O	112
3.31 Test of the binding of OsBGal1 Cter with galactose by HSQC in NMR	113
3.32 Test of the binding of OsBGal1 Cter with rhamnose by HSQC in NMR	114
3.33 Test of the binding of OsBGal1 Cter with glucose by HSQC in NMR.....	115
3.34 Test of the binding of OsBGal1 Cter with raffinose by HSQC in NMR.....	116
3.35 Comparision of sugar binding residues between other lectin domains of mouse latrophilin and chum salmon and OsBGal1 Cter.....	117
3.36 Stability of ¹⁵ N labeled OsBGal1 Cter in NMR.....	119
3.37 Stability of ¹³ C ¹⁵ N labeled OsBGal1 Cter in NMR	120
3.38 OsBGal1 Cter binding to spots on the carbohydrate array membrane	121
3.39 STD NMR of OsBGal1 Cter with 1,5- α -L-arabinotriose	124
3.40 Alignment of the nucleotide sequences of the native and optimized OsBGal1 cDNA	126
3.41 Expression cassette of the pPICZ α BNH ₈ _OsBGal1 plasmid construct.....	129
3.42 Time course of β -glucosidase activity	131
3.43 SDS-PAGE of optimized OsBGal1 after purification via IMAC	132
3.44 Activity versus pH profile of OsBGal1	133
3.45 Activity profile of OsBGal1 over the temperature range 20-75°C	134

LIST OF ABBREVIATIONS

A	absorbance
Å	Ångström(s)
°C	degrees Celsius
bis-acrylamide	N,N-methylene-bis-acrylamide
BSA	bovine serum albumin
CaCl ₂	calcium chloride
cDNA	complementary deoxyribonucleic acid
CV	column volumes
DNA	deoxyribonucleic acid
DNaseI	deoxyribonuclease I
DTT	1,4-dithio-DL-threitol
EDTA	ethylene diamine tetraacetic acid
(n/μ/m)g	(nano, micro, milli) gram(s)
h	hour(s)
IMAC	immobilized metal-affinity chromatography
IPTG	isopropyl β-D-thioglucopyranoside
(k)bp	(kilo) base pair(s)
kDa	kilodalton(s)
(μ/m)L	(micro, milli) liter(s)
(μ/m)M	(micro, milli) molar

LIST OF ABBREVIATIONS (Continued)

<i>Mr</i>	molecular weight
MES	2-(<i>N</i> -morpholino)ethanesulfonic acid
MgCl	magnesium chloride
min	minute(s)
MWCO	molecular weight cut-off
NaCl	sodium chloride
NaOAc	sodium acetate
NaOH	sodium hydroxide
nm	nanometer(s)
OD	optical density
OsBGal1	rice β -galctosidase isoenzyme 1
OsBGal1 Cter	rice β -galctosidase isoenzyme 1 C-terminal domain
PCR	polymerase chain reaction
PMSF	phenylmethylsulfonylfluoride
<i>p</i> NP	<i>para</i> -nitrophenyl
<i>p</i> NPGal	<i>para</i> -nitrophenyl- β -D-galactopyranoside
PAGE	polyacrylamide gel electrophoresis
PEG	polyethyleneglycol
rpm	rotations per minute
s	second(s)
SDS	sodium dodecyl sulfate
S75	superdex 75

LIST OF ABBREVIATIONS (Continued)

T _m	melting temperature
TEMED	N, N, N', N'-Tetramethylethylenediamine
Tris	tris-(hydroxymethyl)-aminomethane
V ₀	initial velocity
v/v	volume per volume
w/v	weight per volume

CHAPTER I

INTRODUCTION

1.1 Glycoside hydrolases

Glycoside hydrolases (GH), also called glycosidases and glycosyl hydrolases (Enzyme Commission, E.C., 3.2.1.-), are a widespread group of enzymes that hydrolyze α or β glycosidic bonds between two carbohydrates or between a carbohydrate and a non-carbohydrate moiety (Durand et al., 1997). The enzymatic hydrolysis usually occurs by one of two major mechanisms leading to overall retention or inversion of anomeric configuration. GH have been divided into families related by amino acid sequences. These classifications have identified over 132 such families at this point in time, as documented in the CAZY (Carbohydrate Active enZYme) website; www.cazy.org/CAZY/ (Coutinho and Henrissat, 1999; Cantarel et al., 2009).

Glycoside hydrolases are found in essentially all domains of life. In bacteria and archaea, they can be found as intracellular and extracellular enzymes that are involved in nutrient acquisition. One of the important types of glycoside hydrolases in bacteria is the enzyme beta-galactosidase (LacZ in *Escherichia coli*), which is expressed in a lactose-dependent manner from the *lac* operon in *E. coli*. In higher organisms, many GH are found in the endoplasmic reticulum and Golgi apparatus.

The GH families can be grouped into clans related by their three dimensional structures and catalytic mechanisms. The GH clan A, for instance, is made up of enzymes with $(\beta/\alpha)_8$ -barrel structures that have the catalytic acid/base and nucleophile

on beta strands 4 and 7, respectively, of the β -barrel. This clan includes glycoside hydrolase families 1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113, and 128, which act with different substrate specificities, but appear to have evolved from a common ancestry (Henrissat et al., 1995; Jenkins et al., 1995; Bolam et al., 1996; Henrissat and Bairoch, 1996). These enzymes hydrolyze the glycosidic bond with retention of the anomeric configuration.

The GH35 enzymes, which are the focus of this thesis, have a range of activities, including β -D-galactosidase (EC 3.2.1.23), exo- β -D-glucosaminidase (EC 3.2.1.165), exo- β -1,4-galactanase (EC 3.2.1.-) (Coutinho and Henrissat, 1999; CAZY:www.cazy.org/CAZY/).

1.2 Glycosidase mechanism

The enzymatic hydrolysis of the glycosidic bond takes place via general acid catalysis that requires two critical residues: a proton donor and a nucleophile/base (Davies and Henrissat, 1995). This hydrolysis occurs via two classes of elementary mechanisms, which are called the inverting and retaining mechanisms (Sinnott, 1990; McCarter and Withers, 1994; Ly and Withers, 1999). Both of these use a pair of carboxylic groups, or occasionally other proton donors/acceptors.

1.2.1 Inverting mechanism

In the proposed mechanism for inverting enzymes, the glycosidic oxygen is initially protonated by a general acid catalyst, followed by a nucleophilic attack on the C1 atom of the sugar by a water molecule, which is activated by extraction of a proton by a catalytic base residue (usually a carboxylate, as shown in Figure 1 A)

leading to inversion of the anomeric conformation. This is known as a single displacement mechanism (Koshland, 1953), bond breaking and bond making both proceed in a single step.

1.2.2 Retaining mechanism

The retaining reactions proceed via a double-displacement mechanism (Koshland, 1953). The first step involves a similar protonation of the glycosidic oxygen by a catalytic acid and attack at the anomeric carbon by an enzyme nucleophile to form a stable intermediate. The enzyme is displaced by a water nucleophile assisted by the base form of the acid catalyst in the second step (Figure 1.1 B, C). Each step inverts the configuration of the anomeric carbon. The displacement steps therefore create an overall retention of the configuration. For retaining enzymes, the intermediate could either be an oxocarbenium ion which is electrostatically stabilized by a carboxylate (Figure 1.1 B) or could involve formation of a covalent bond (Figure 1.1 C), in which one of the catalytic carboxylates is presumed to act as a nucleophile. Transglycosylases (e.g. cyclodextrin glycosyl transferases) employ a reaction mechanism similar to that described for retaining hydrolases. For these enzymes however, the second step of the reaction does not involve a water nucleophile, but the nonreducing end of a saccharide, possibly assisted by the base form of the acid catalyst (Sinnott, 1990).

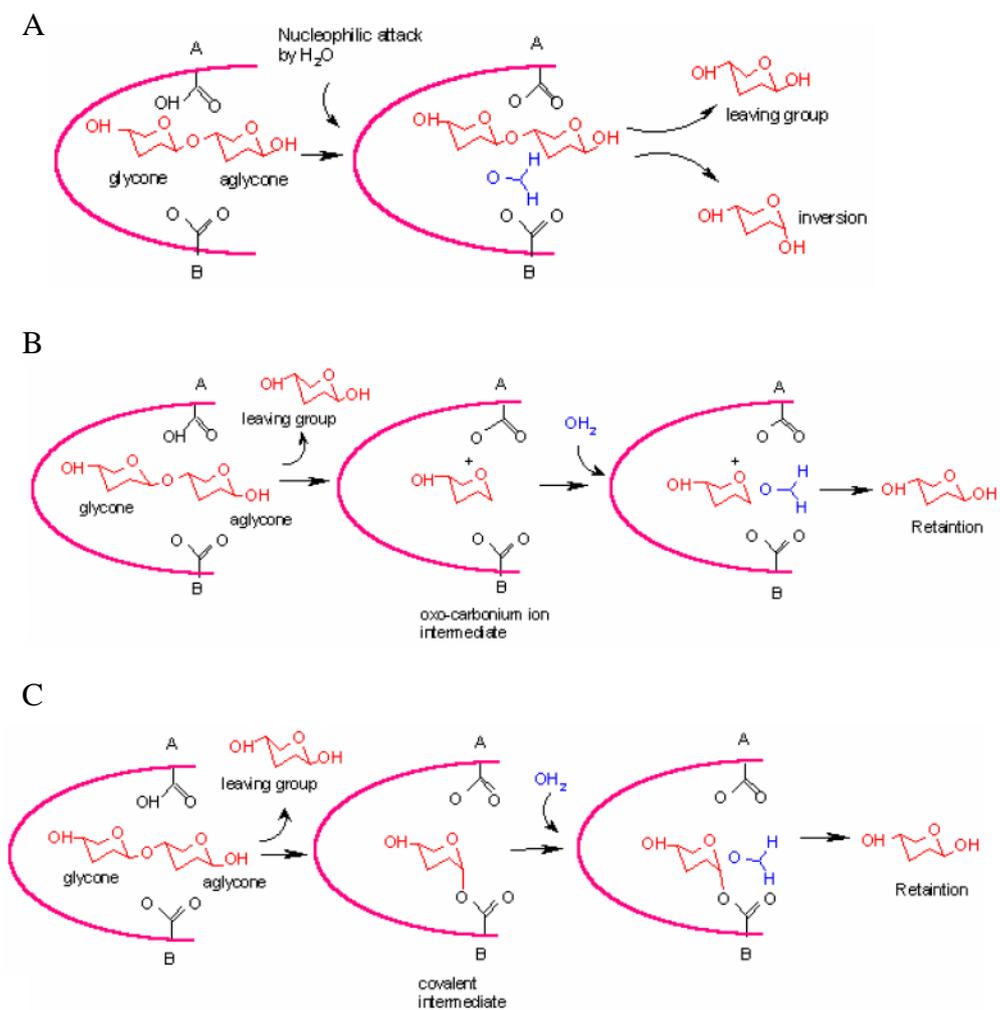


Figure 1.1 Catalytic mechanisms of inverting enzyme (A) and retaining enzyme acting through an oxo-carbonium ion (B) or via a covalent intermediate (C) (Vernon, 1967)

1.3 Plant Glycoside Hydrolases

Glycoside hydrolases are a set of gene families that are of particular interest to investigate in plants. Plants have a wide variety of glycosides and polysaccharides, which carry out many functions. These include cell wall polysaccharides, which form the bulk of the structure and biomass of plants, as well as starch and other forms of storage carbohydrates. Glycosides found in plants include glycolipid components of cellular membranes and pigments, as well as many reactive or bioactive species that are blocked from their activity by sugars for storage or to allow reactions at other positions on the molecule. The compounds with sugar blocking groups include phytohormones, which can be released to active forms by glycoside hydrolases, lignin precursors, which can be transported to the area of lignification, then activated, and defense compounds, which can be activated by mixing of the glycoside hydrolase with the glycoside upon fungal invasion or herbivore breakage of the cells. The glycoside hydrolases may play important roles in plant growth, development and adaptation. In addition, glycoside hydrolases have great potential for biotechnological applications, such as in release of nutrients and flavor compounds in plant derived foods and feeds, conversion of certain sugars, like lactose in milk products to more desirable sugars and in biomass conversion of cellulose and other polysaccharide waste to useful products, including fuels. β -Glycosidases, including β -glucosidases, β -galactosidases and other related enzymes are particularly interesting in this regard.

Halomonas sp. S62 enzyme (Henrissat, 1991, Cantarel et al., 2009). The β -galactosidases have been characterized in plants belong in GH family 35. GH family 35, like other families including β -galactosidases falls into GH Clan A. The catalytic nucleophiles have been defined for the GH family 35 β -galactosidases from *Xanthomonas manihotas* and *Bacillus circulans* by labeling with a 2-deoxy-2-fluorogalactoside (Blanchard et al., 2001). The acid/base and nucleophile could also be putatively identified from the crystal structure of the *Penicillium* sp. β -galactosidases (Rojas et al., 2004).

The functions of β -galactosidases are various. In microorganisms, the first purification of β -galactosidase from *E. coli* was carried out by Cohn and Monod (1951), who found that this enzyme can hydrolyze lactose and other β -galactosides to release monosaccharides (Wallenfels and Well, 1972). Furthermore, the regulation of production of *LacZ* β -galactosidase in *E. coli* helped scientists to understand about gene regulation and made it an important model in the development of molecular biology. Many vectors in current use carry a short segment of *E. coli* DNA that contains the regulatory sequences (the *lac* promoter) and the coding information for the first 146 amino acid of β -galactosidase gene (*LacZ*) (Sambrook et al., 1989). The introduction of such vectors in certain strains of *E. coli* containing the rest of the *LacZ* gene allows recombination to form a functional gene and production of *LacZ* β -galactosidase upon induction with lactose or the non-hydrolyzable β -D-galactoside isopropyl β -D-thiogalactopyranoside (IPTG). However, cloning of a DNA fragment into one of the cloning sites within this *LacZ* sequence will disrupt the protein coding sequence, resulting in a lack of β -galactosidase activity. Thus, cells containing recombinant clones appear white when grown on media containing 5-bromo-4-chloro-

3-indolyl- β -D-galactopyranoside (X-Gal), which forms a blue precipitate when hydrolyzed, while those with non-recombinant plasmid are blue due to the hydrolysis of X-Gal by β -galactosidase.

In plants, the activity of β -galactosidases has mainly been described in the processes of growth, development, senescence and fruit ripening, where they are typically thought to act on cell wall carbohydrates (Konno and Katoh, 1992; Ross et al., 1994; Carey et al., 1995; Li et al., 2001). Though animals and microorganisms have only a few β -galactosidases, mRNA for at least 7 family 35 glycoside hydrolase genes likely to be β -galactosidases have been detected in ripening tomato (Smith and Gross, 2000), while *Arabidopsis thaliana* has 17 GH family 35 genes (Ahn et al., 2007). It appears that at least some of these enzymes are involved in cell wall remodeling, although the large number expressed suggests that they may play a variety of roles.

Most of the studies on plant β -galactosidases have been done on dicotyledons, in which the cell wall contains a relatively large amount of pectic polysaccharides compared to gramineous monocotyledons. It was suggested that β -galactosidases from several plants are involved in the removal of galactose from pectic and hemicellulosic polysaccharides and glycoproteins of cell walls (Kotake et al., 2005). The increase in β -galactosidase in asparagus during harvest-induced senescence would support such a role in a nongramineous monocot (O'Donoghue et al., 1998), which has cell wall structure more similar to dicots than gramineous monocots. Relatively little work has been done on gramineous monocots, in which the role in the cell wall is less obvious.

Among grains, β -galactosidases have been described in barley and rice. Giannakouros et al. (1991) were able to separate 4 isoenzymes of β -galactosidase from germinating barley. A β -galactosidase purified from barley was shown to be a heterodimer with 35 and 45 kDa subunits, which appeared to have originated from the same precursor protein (Triantafillidou and Georgatsos, 2001).

1.5 Rice β -galactosidases

Rice (*Oryza sativa* L.) is one of the world's most important crops. Knowledge of rice molecular genetics and physiology is important for the development of molecular crop improvement techniques (Gurdev and Grary, 1991). Moreover, rice is the primary monocot genome model, with a complete genome sequence available. Therefore, studies in rice allow the use of genome sequence data, as well as cDNA tools, like expressed sequence tags (ESTs) and full length cDNA clones, to aid the investigation.

Fifteen rice BGal genes were identified in the plant genome, one of which encodes a protein similar to animal BGals (OsBGal9), and the remaining 14 fall in a nearly plant-specific subfamily of BGals (Tanthanuch et al., 2008). Gene structure comparison of rice β -galactosidases indicated eight different splicing patterns. The sizes of the majority of the coding exons are conserved, but some appear to have had intron-loss events. The pattern with the highest number of exons, found in OsBGal1, OsBGal3, OsBGal4, OsBGal6 and OsBGal13, contains 19 exons with 18 introns. Moreover, exons 1 through 10 encoding the GH35 catalytic domain and exons 18 and 19 encoding the Gal-lectin-like domain. The proteins from OsBGal2, OsBGal7 and

OsBGal9 lack the Gal-lectin-like domain. All fifteen rice BGal genes were found to be expressed with different but overlapping tissue expression patterns.

In 1993, Konno and Tsumuki reported that a rice β -galactosidase which they purified by DEAE ion exchange CL 6B, Sephacryl S200 gel filtration and affinity chromatography on *p*-aminophenyl- β -D-thiogalactopyranoside-linked Sepharose 4B gave a single protein band with a molecular weight of 42 kDa, similar to the size of a β -galactosidase from radish seeds (45 kDa, Sekimato et al., 1989). Kaneko and Kobayashi (2003) isolated a β -galactosidase from the medium of rice suspension cells, and showed that the purified enzyme contains heterodimeric subunits of approximately 40 and 47 kDa. This enzyme was also found to release galactose from cell wall derived polysaccharides, consistent with a role of rice β -galactosidase in cell wall remodeling. Chantarangsee et al. (2007) cloned cDNA for two rice β -galactosidases and expressed them in *E. coli*, to show the proteins indeed had β -galactosidase activity.

1.6 Cell wall structure and metabolism

An early model of the plant primary cell wall is that it consists of cellulose fibrils, which are coated with hemicellulose and the matrix, composed of pectic polysaccharides (Cosgrove, 2005). Cellulose consists of linear chains of β -(1-4)-linked residues, which are aggregated together with hydrogen bonds to form fibrils. Hemicellulose is composed of a variety of polymers, including xyloglucan, glucomannan, and galactoglucomannans. Pectic polysaccharides are also composed of various monosaccharides, the neutral pectic polysaccharides being arabinan, galactan or arabinogalactans. It is thought that these polymers are linked together in the

three dimensional structure of cell wall by covalent and noncovalent bonds. Cellulose fibrils are held together by hydrogen bonds and similar bonds account for the interaction of cellulose with hemicelluloses. However, it has been postulated that, in addition to simply coating the cellulose fibril, hemicelluloses molecules may cross-link the cellulose fibrils. Seymour et al. (1993) proposed releasing or relaxation of such linkage could be a major cause of cell wall loosening and softening.

Brett and Waldron (1996) noted that the enzymes are located in the cell wall include pectinase, cellulose peroxidase and exoglucosidases, have been reported to be involved in cell wall turnover, including α -galactosidases, β -galactosidases, β -glucosidases and β -xylosidases. Much attention have been given to xyloglucan endotransglycosylase (XET), which is found in a range of walls and may be involved in insertion of xyloglucan into the cell wall and possibly in control of cell wall extensibility. One possible natural substrate of β -galactosidases may be the polymers containing β -D-galactose in plant cell walls, such as arabinogalactan, a polymer of α -L-arabinose and β -D-galactose associated with the wall. In fact, in several cases, β -galactosidases from ripening fruit have been shown to attack pectin polymers (Husain, 2010). These enzymes remove terminal β -D-galactosyl residues, so they can be considered as true β -galactosidases in the cell wall and could be involved in degradation.

1.7 Recombinant expression of plant β -galactosidases

Several β -galactosidases from different plant sources have been expressed in recombinant systems. For example, when Smith and Gross (2000) described a family of at least seven β -galactosidases expressed during tomato fruit development, they

expressed the tomato β -galactosidase isoenzyme 4 (TBG4) in yeast and studied its natural substrate specificity. TBG4 hydrolyzed chelator-soluble pectin, alkali-soluble pectin, and hemicellulose from tomato cell walls, and commercially prepared galactan. *Arabidopsis thaliana* Gal-4 (At5g56870), one of the seventeen putative GH family 35 β -galactosidases in Arabidopsis, was expressed in *Escherichia coli*, *Pichia pastoris*, and insect cells (Ahn et al., 2007). It was shown that in addition to synthetic substrates, Gal-4 hydrolyzes β -(1,3)- and β -(1,4)- linked galacto-oligosaccharides (Gantulga et al., 2008 and 2009; Ishimaru et al., 2009). In 2007, Chantarangsee et al. cloned and expressed OsBGal1 and OsBGal2 in *E. coli*. The recombinant OsBGal1 fusion protein and OsBGal2 were detected with molecular weights of approximately 90 kDa and 55 kDa, respectively. But both enzymes proved difficult to purify because of its apparent susceptibility to degradation. Purified OsBGal1 fusion protein hydrolyzed and transglycosylated *p*-nitrophenyl (*p*NP) β -D-galactopyranoside and hydrolyzed *p*NP β -D-fucopyranoside. Galactose was released by this enzyme from β -(1,3)-, β -(1,4)- and β -(1,6)-linked di- and trisaccharides. Recombinant OsBGal13 was expressed in *E. coli* strain Origami B(DE3) and then purified by IMAC (Tanthanuch et al., 2008).. This enzyme had high activity toward *p*NP- β -D-galactoside and low activity toward *p*NP- β -D-fucoside and *p*NP- β -D-mannoside. Moreover, this enzyme also hydrolyzed β -(1,3)-, β -(1,4)- and β -(1,6)-linked galactobiose and galactotriose, but no release of galactose or arabinose from polysaccharides was detected, except for larchwood arabinogalactan.

1.8 Three-dimensional (3D) structures of GH35 β -galactosidase

Based on the sequence homology criterion, most β -galactosidases belong to GH families 1, 2, 35 and 42, which are all members of the clan called GH-A, as noted in Section 1.4. GH-A enzymes all cleave glycosidic bonds via a retaining mechanism and possess a catalytic domain which is based on a TIM barrel fold. Two glutamic acid residues act as proton donor and nucleophile (Jenkins et al., 1995) and are found at the ends of β -strands 4 and 7 of the barrel, and for this reason this clan is sometimes referred to as the 4/7 superfamily.

The β -galactosidase from *E. coli*, an enzyme which belongs to GH family 2, is one of the most widely studied β -galactosidases. The 3D structure of LacZ β -galactosidase is known and the structural basis for its reaction mechanism has been reported (Jacobson et al., 1994; Juers et al., 1999 and 2000). Another β -galactosidase of known crystal structure is that from *Thermus thermophilus* (A4- β -gal), a member of GH family 42 (Hidaka et al., 2002). LacZ forms a 464 kDa homotetramer in solution, while A4- β -gal is a trimer.

The first 3D-structure of a GH35 enzyme was that of *Penicillium* sp. β -galactosidase (Psp- β -gal). The amino acid sequence translated from the Psp- β -gal gene has 1011 residues. The Psp- β -gal primary structure revealed that the 40 N-terminal amino acid residues are involved in signaling and are cleaved from the mature protein. The crystal structures of Psp- β -gal (PDB accession 1TG7) and its complex with galactose (PDB accession 1XC6) were determined at 1.90 Å and 2.10 Å resolution, respectively (Rojas et al., 2004). The Psp- β -gal catalytic residues Glu200 and Glu299 were identified as the proton donor and the nucleophile of the reaction,

respectively. Its structure can be divided into five domains. The first domain, containing the catalytic site, is a distorted TIM barrel comprising 355 amino acid residues (Leu41-Gly395). The second domain, comprising amino acid residues Tyr396-Tyr576, consists of 16 antiparallel β -strands and an α -helix at its C terminus. The third domain (Trp577-Tyr665) is much smaller than the second and consists of an α -helix at its N-terminus, followed by eight antiparallel β -strands arranged as a Greek key β -sandwich. On exiting the third domain, a short stretch of the polypeptide chain (residues Thr666-Pro687) passes through domain 5, forming a short β -strand prior to entering domain 4 (Figure 1.3). The fourth domain comprises amino acid residues Glu688-Leu861 and is composed of eight β -strands in a β -sandwich, which is best described as a class II right-handed jelly roll (Stirk et al., 1992). The fifth domain is based on a class I jelly roll and consists of a total of eight strands divided into five and three-stranded β -sheets. The first strand of a conventional jelly roll is missing, which is why one of the sheets possesses only three strands. The other sheet includes an additional strand formed by part of the connecting peptide which runs between domains 3 and 4. Of these five domains, only the first is clearly related to plant GH family 35 enzymes based on sequence analysis. An additional GH family 35 structure has been solved for a 625 amino acid residue *Bacteroides thetaiotaomicron* β -galactosidase at 2.15 Å resolution (Palani et al., 2008, PDB accession 3D3A). This smaller structure has only three domains, the GH35 catalytic domain, followed by two β -sandwich domains. In 2010, the crystal structure of a β -galactosidase from *Trichoderma reesei* (*Hypocrea jecorina*) (Tr- β -gal, PDB 3OG2) was reported, together with complex structures with galactose, IPTG and 2-phenylethyl-b-D-

thiogalactoside (PETG) at 1.5, 1.75 and 1.4 Å resolution, respectively (PDB accessions 3OGR, 3OGS, and 3OGV, respectively) (Maksimainen et al., 2010).

The comparison of the native structures of Psp-β-gal, Tr-β-gal and Btmβ-gal (Figure 1.4 A) show that Btm-β-gal consists of three distinct domains, whereas Psp-β-gal and Tr-β-gal consist of five and six domains, respectively. The second and third domains of Btm-β-gal are quite similar with the fourth and fifth domains of Psp-β-gal, and with the fifth and sixth domains of Tr-β-gal. Although the crystal structures of Psp-β-gal and Tr-β-gal are similar, the interpretation of the structure of Tr-β-gal is somewhat different from that presented earlier for Psp-β-gal (Rojas et al., 2004), which considered Psp-β-gal to be composed of five distinct structural domains.

The superimposition of the active sites of the GH35 β-galactosidases shows a remarkable similarity (Figure 1.4 A). In addition to the catalytic residues, the active sites of the GH35 β-galactosidases contain many identical residues (Figure 1.4 B). Based on the galactose-bound crystallographic models of Psp-β-gal and Tr-β-gal, a single galactose molecule is bound to the active site of the GH35 enzyme in the ⁴C₁ chair conformation in the β-anomeric configuration.

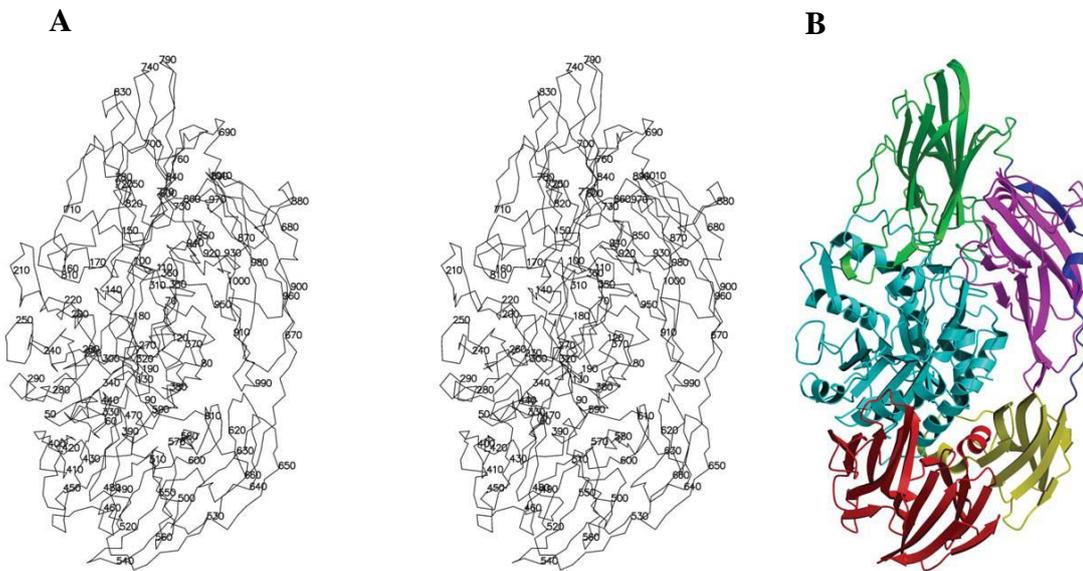


Figure 1.3 *Penicillium* sp. β -galactosidase (PDB ID 1TG7) is composed of five distinct structural domains. The overall structure is built around the first, the catalytic TIM barrel domain. Domain 2 is an all β -sheet domain containing an immunoglobulin-like subdomain, domain 3 is based on a Greek-key β -sandwich and domains 4 and 5 are jelly rolls. (A) Stereo view of the Psp- β -gal C_{α} trace. (B) Ribbon representation of the secondary structure elements of Psp- β -gal, defined by the program PROCHECK. Domains 1–5 are colored in cyan, red, yellow, green and magenta, respectively. The long linker peptide connecting domains 3 and 4 is depicted in blue. The figure is from Rojas et al. (2004).

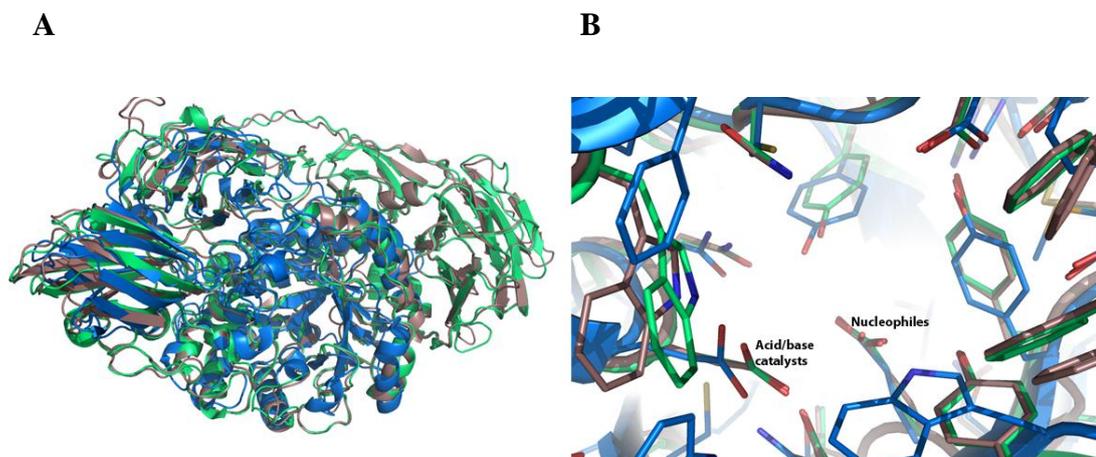


Figure 1.4 **A)** Comparison of the native structures of GH35 β -galactosidases. Psp- β -gal (PDB 1TG7), Tr- β -gal (PDB 3OG2) and Btm- β -gal (PDB 3D3A) are colored in green, brown and blue, respectively. **B)** Comparison of the active sites of the GH35 β -galactosidases. Psp- β -gal, Tr- β -gal and Btm- β -gal are colored as described for part A (Rojas et al., 2004; Palani, et al., 2008 and Maksimainen et al., 2010).

1.9 Carbohydrate binding domains

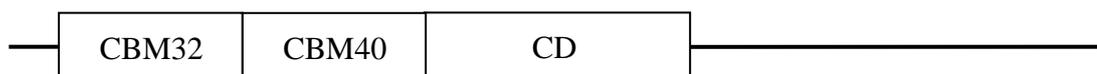
Several carbohydrate-active proteins have obtained noncatalytic modules that interact very specifically with mono, oligo, and polysaccharides. In general, these carbohydrate-binding modules (CBMs) bind heterogeneous and complex carbohydrates (Boraston et al., 2004). CBMs may be found in any domain of life. CBMs were previously classified as cellulose-binding domains (CBDs), based on the initial discovery of several modules that bound cellulose (Tomme et al., 1988; Gilkes et al., 1988). Additional CBMs are continually being found that bind chitin, β -glucans, starch, glycogen, inulin, xylan, arabinofuranose, mannan, fucose, lactose, and galactose, while some CBMs display lectin-like specificity and bind to a variety of cell-surface glycans. The primary function of CBMs is to increase the catalytic

efficiency of the enzymes against soluble and/or insoluble substrates by bringing the catalytic module into prolonged close contact with substrates (Tomme et al., 1995; Bolam et al., 1998).

Carbohydrate-active enzymes (CAZymes) can include one or more CBMs. Furthermore, the same or different combinations of CBMs may be found in different CAZymes. CBMs can be localized at the N- or C-terminal end of these proteins (Abe et al., 2004; Juge et al., 2002). For example, the *Paenibacillus polymyxa* alpha,beta-amylase contains two CBMs between two catalytic modules (Kawazu et al., 1987). Another example, *Clostridium perfringens* exo-alpha-sialidase contains two CBMs, one of which can bind galactose or N-acetylgalactosamine, while the other binds to sialic acid (Boraston et al., 2007). *Thermotoga maritima* endo-1,4-beta-xylanase contains two CBMs, one at the C-terminus from family 22 and a family 9 CBM at the N-terminus. The family 22 CBM binds xylan, whereas the family 9 CBM binds cellulose (Ito et al., 2003; Figure 1.5).

The most important conformational element of most CBMs is the beta-sheet. The folds and architecture displayed by these beta-sheets have been classified into seven families (Boraston et al., 2004; Hashimoto, 2006: Table 1). The topology of CBM–ligand binding sites, which is related to the class of ligand they recognize, has been used to classify them into three types (Boraston et al., 2004). Type A CBMs have a flat or platform-like hydrophobic surface composed of aromatic residues. The planar conformation of the type A binding site interacts with the flat surfaces of crystalline polysaccharides such as cellulose or chitin (Tormo et al., 1996; McLean et al., 2000). The type A CBMs are found in a range of enzymes, including cellulases, mannanases, xylanases and pectinases. Type B CBMs display a cleft arrangement, in

which aromatic residues interact with free single polysaccharide chains. The type B modules displaying affinity for cellulose, xylan, β -1,3-glucans, β -1,3-,1,4-mixed-linkage glucans, starch, glucomannan and mannan. The final group, type C CBMs are lectin-like and contain smaller binding pockets that only bind mono-, di-, or trisaccharides due to steric restriction in the binding site (Fujimoto et al., 2002).



Clostridium perfringens exo- α -sialidase (Q8XMY5)



Paenibacillus polymyxa alpha, beta-amylase (P21543)



Thermotoga maritima endo- β -1,4-xylanase (Q60037)

Figure 1.5 CBM distribution in different proteins. CD symbolizes the position of the catalytic domain in glycoside hydrolases (Guillén et al., 2009). The CBM are designated by their CBM family number.

Table 1.1 CBMs classification based on fold (Boraston et al. 2004 and Hashimoto 2006).

Fold family	Fold	CBM families
1	β -Sandwich	2, 3, 4, 6, 9, 11, 15, 16, 17, 20, 21,22, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 40, 41, 42, 44, 47, 48, 51
2	β -Trefoil	13, 42
3	Cysteine knot	1
4	Unique	5, 12
5	OB fold	10
6	Hevein fold	18
7	Unique; contains hevein-like fold	14

1.10 Structure of galactose binding-lectin like domains

1.10.1 Chum salmon rhamnose binding lectin (RBL) structure

The C-terminal domains found in many plant β -galactosidases are homologous to a family of domains first described as a galactose-binding lectin in sea urchin eggs (Trainotti et al., 2001; Ozeki et al., 1991). Because several of these animal carbohydrate binding modules were found to bind more tightly to L-rhamnose, the family has been designated L-rhamnose binding lectins (RBLs). As noted above, the RBL is a domain first characterized from sea urchin eggs, where it was abbreviated SUEL (Ozeki et al., 1991). RBLs have since been found in over 25 species of fish (Hosono et al., 1993; Tateno et al., 2002; Terada et al., 2007; Watanabe et al., 2008), oyster (Naganuma et al., 2006) and ascidian (Gasparini et al., 2008). The RBLs bind specifically to L-rhamnose or D-galactose (Kilpatrick et al., 2002; Natta et al., 2007). Moreover, RBL domains do not require cofactors, such as metal ions, for carbohydrate recognition (Ozeki et al., 1991).

The structures of the RBL carbohydrate-binding modules from animals have been investigated. In 2009, the crystal structure of the RBL from eggs of chum salmon (CSL3) was determined. The homodimer of CSL3 has two lobes that are interconnected through linkers (Figure 1.7 A). Each lobe consists of N and C-terminal RBL domains from different subunits. The N- and C-terminal domains show 73% identity in amino acid sequence. Each domain has two anti-parallel β -sheets, one comprising β -strands β 2 and β 4 and the other β -strands β 1, β 3, and β 5, and three helices (α 1–3). A total of four disulfide bonds connect the backbone within each domain (Figure 1.7 B) (Tsuyoshi et al., 2009). The structures were also solved for the

complexes of CSL3 bound with L-rhamnose, melibiose (Gal- α -1-6-Gal), and globotriose ceramide (Gb3).

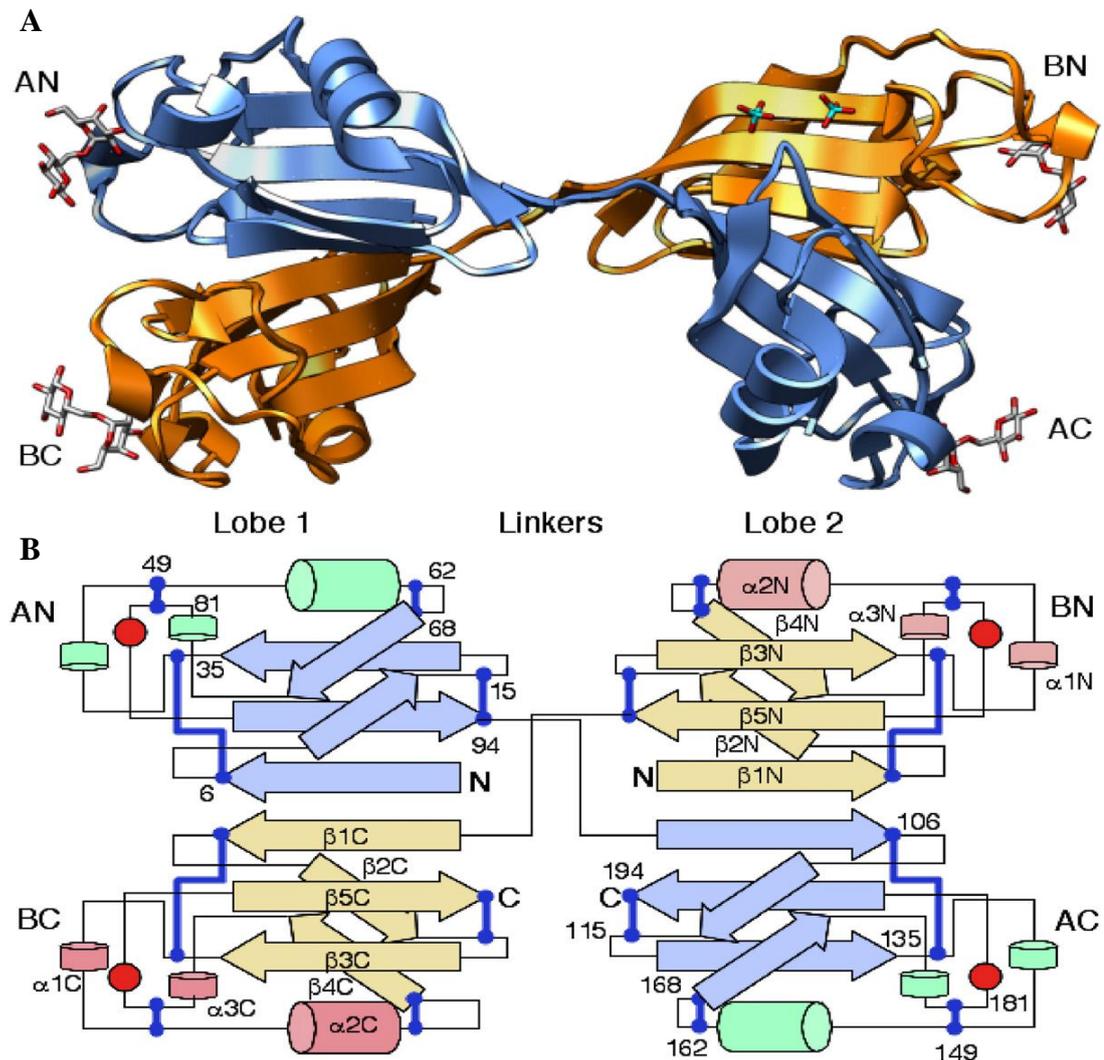


Figure 1.6 CSL3 dimer structure. (A) Subunits A and B of CSL3 are shown in blue and orange ribbon models, respectively. The bound melibiose molecules and phosphate ions are shown as stick models. The N- and C-terminal RBL domains of chains A and B are named AN, AC, BN, and BC, respectively. The domains AN-BC and AC-BN form lobes 1 and 2, which are connected with two linkers. (B) Diagram of the secondary structure arrangement. The helices and strands of CSL3 are shown in cylinders and arrows, respectively, and are labeled only for subunit B. The disulfide bonds are shown in thick lines, and labeled only for subunit A. Red circles indicate the positions of the primary binding sites (Shirai et al., 2009).

1.10.2 The mouse latrophilin RBL structure determined by NMR

Latrophilins are putative class G protein-coupled receptors (Bjarnadottir et al., 2007) widely expressed in the brain (Matsushita et al., 1999). Their physiological function is unknown. Mouse latrophilin-1 (Lat-1) has the carbohydrate binding properties of the rhamnose binding lectin (RBL) domain. The structure of the RBL domain from mouse latrophilin-1 was determined by NMR (Vakonakis et al., 2008). The RBL domain from mouse latrophilin-1 contains eight cysteines that are necessary for folding. This protein is a monomer in solution. The structure of the latrophilin RBL domain has five β strands (residues 36-41, 45-49, 54-65, 101-105, and 121-130), a single, long α -helix (residues 86-96), and single-turn α -helix at residues 75-78 and single-turn 3_{10} -helix at residues 108-110. The overall fold is that of a β sandwich with two antiparallel sheets. The two sheets diverge between β 4 and β 3, and α 2 caps the exposed side of the sandwich. The RBL domain of mouse latrophilin 1 has 4 disulfide bridges. The RBL domain includes two long loops connecting β 3 and α 2 (residues 66-85, loop 1) and β 4 with β 5 (residues 106-120, loop 2). Loop 1 is longer than loop 2 (Figure 1.9 A).

When the protein was titrated with sugars and the NMR spectrum changes observed, the RBL domain bound tightly with L-rhamnose ($K_d = 1.8$ mM) and also bound to D-galactose, D-fucose, and L-arabinose, in decreasing order of affinity.

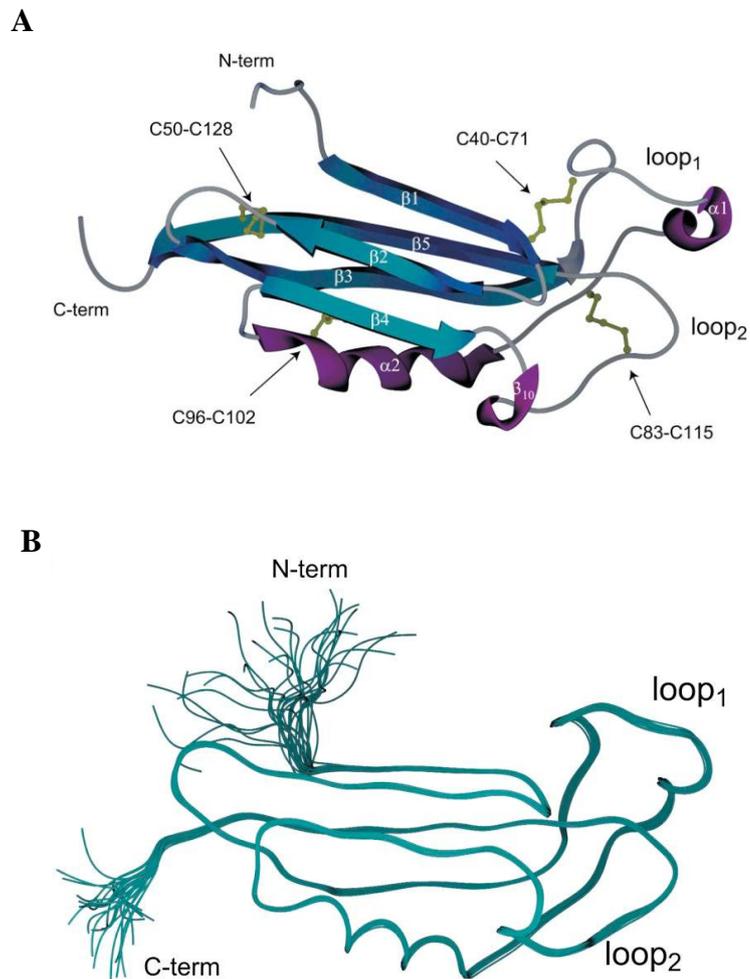


Figure 1.7 Mouse lactophilin 1 RBL domain structure. (A) RBL structure of mouse lactophilin 1. The secondary structure elements and loop 1 and loop 2 are indicated. Disulfide bridges are shown in gold. (B) The 25 structure ensemble of RBL (Vokonakis et al., 2008).

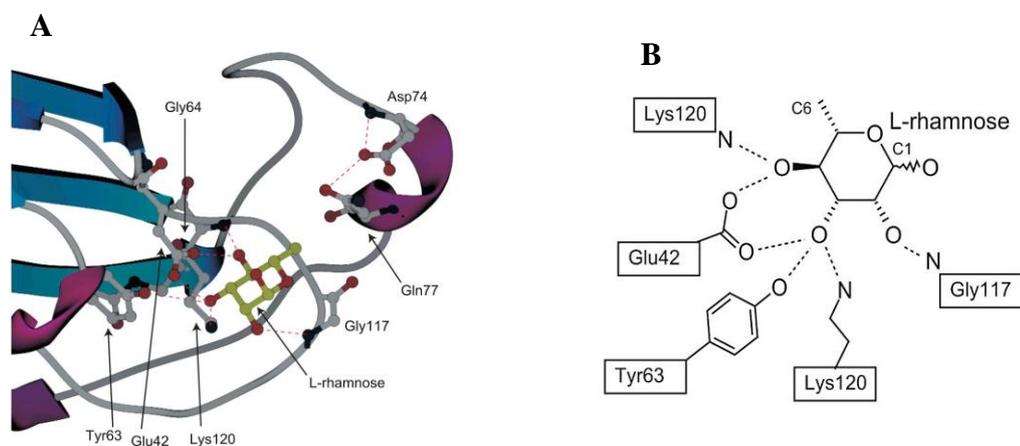


Figure 1.8 Rhamnose binding by mouse latrophilin RBL domain. (A) Detail of the structure of the complex of the latrophilin RBL domain with rhamnose (gold) with interfacial residues indicated. The hydrogen bonding network formed is denoted by dashed lines. (B) Coordination of rhamnose binding. The C1 and C6 positions are noted for clarity (Vokanakis et al., 2008)

1.11 Protein structure investigation by NMR

NMR was first used to determine protein structure in the 1980s. Structural determination of protein by NMR requires the assignment of the NMR signals from the many nuclei in the protein (Wüthrich, 2001). This technique is generally limited to small proteins, because large proteins have too many overlapping peaks in their NMR spectra. Many experiments that use different pulse sequences to link spectral peaks associated with different neighboring atoms are used to annotate the peaks to specific atoms in the primary structure in order to determine the structure of protein, including HSQC, HNCQ, CBCACONH, HNCACB and CCONH. Certain bond information from these spectra is then combined with through space interactions determined by Nuclear Overhauser Effect Spectroscopy (NOESY) in a set of constraints that define the 3-D structure. The structure of a protein can be calculated by optimizing models to match the constraints and the known properties of protein structure, such as bond lengths, bond angles, etc.

1.11.1 The ^1H - ^{15}N Heteronuclear Single-Quantum Correlation (HSQC) experiment

The ^{15}N HSQC is normally the first spectrum obtained for the assignment of resonances where each amide peak is assigned to a particular residue in the protein. This experiment is one of the most frequently recorded in protein NMR. The HSQC experiment can be performed with isotope labeled proteins in NMR. Such proteins are produced by expressing the protein in cells grown in ^{15}N -labelled media. Peaks can be observed in the spectra for each residue, with the exception of proline, which lacks an amide proton attached to a nitrogen in the peptide bond, and normally

the N-terminal residue (which has an NH_3^+ group) cannot readily be observed due to exchange of protons with the solvent.

If the protein is folded, the peaks are well-dispersed, and most of the individual peaks can be distinguished. In a HSQC spectrum, the NH_2 peaks from the side chain of asparagine and glutamine appear as doublets on the top right corner (Figure 1.9). The side chain amine peaks from tryptophan are shifted down and appear near the bottom left corner. Moreover the backbone amide peaks of glycine normally appear near the top of the spectrum (Figure 1.9). A large cluster of overlapped peaks around the middle of the spectrum would indicate the presence of significant unstructured elements in the protein. In such cases, there are critical overlaps of resonances the assignment is difficult.

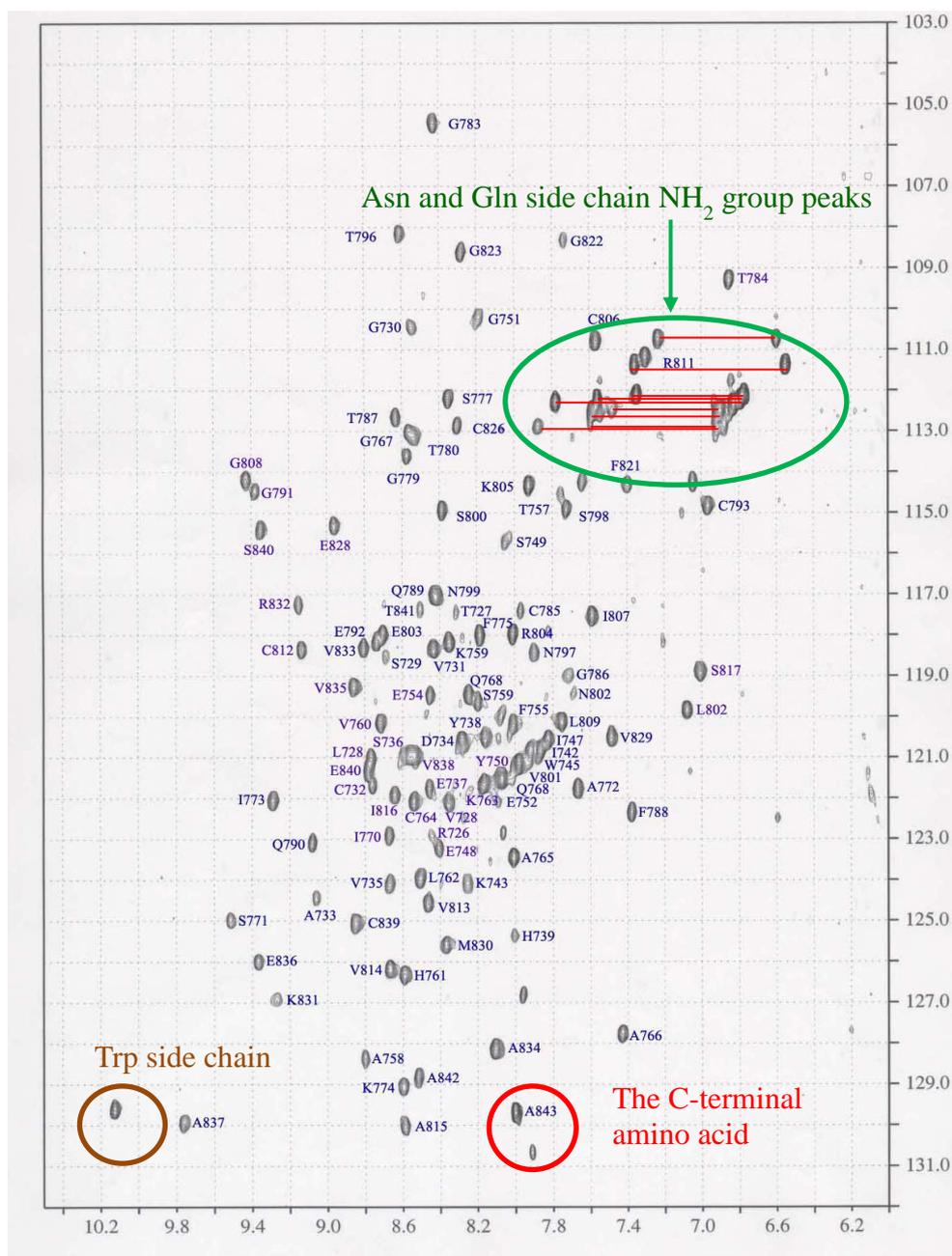


Figure 1.9 ^1H - ^{15}N HSQC spectrum of C-terminal domain of β -galactosidase 1 from rice. Each peak in the spectrum represents a bonded N-H pair, with its two coordinates corresponding to the chemical shifts of each of the H and N atoms.

1.11.2 HNC(O) experiment

The HNC(O) experiment is a three dimensional NMR method for determining structure of protein. This experiment is designed to correlate critical sequence backbone connectivity between the ^1H - ^{15}N pair of one residue with the carbonyl (^{13}CO) resonance of the previous residue. The side chains carboxamides of asparagine and glutamine are also visible in this experiment.

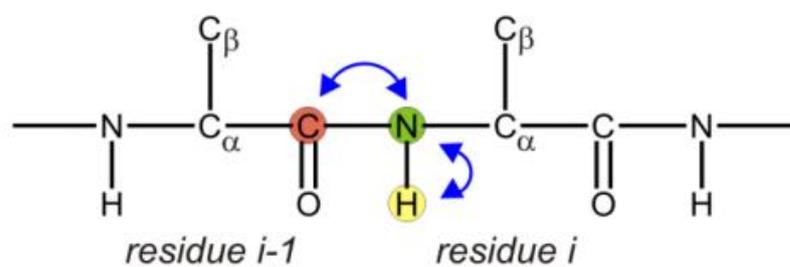


Figure 1.10 Schematic of HNC(O) resonances.

1.11.3 CBCA(CO)NH experiment

The CBCA(CO)NH experiment is similarly used in determining the three-dimensional structure of proteins. This experiment is designed to correlate the ^1H and ^{15}N amide resonances of one residue with both $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ resonances of its preceding residue via the intervening ^{13}CO spin by means of the $^1\text{J}(\text{NH})$, $^1\text{J}(\text{N},\text{CO})$, $^1\text{J}(\text{C}\alpha,\text{CO})$ and optional $^1\text{J}(\text{C}\alpha,\text{C}\beta)$ coupling constants.

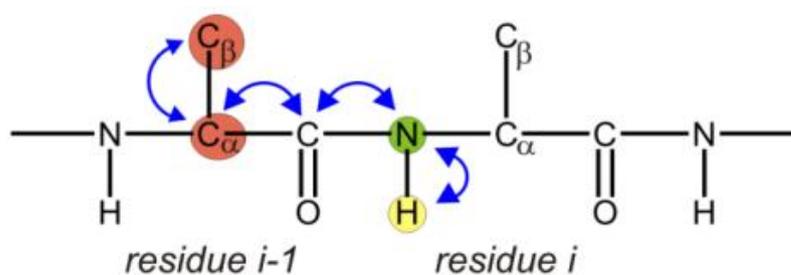


Figure 1.11 Schematic of CBCA(CO)NH resonances.

1.11.4 HNCACB experiment

The HNCACB spectrum resolves amide proton/nitrogen correlations in the same fashion as the 2D ^1H - ^{15}N HSQC. The third dimension of the spectrum contains the ^{13}C chemical shifts of the $\text{C}\alpha$ and $\text{C}\beta$ resonances of a given residue and of the residue before it in the sequence.

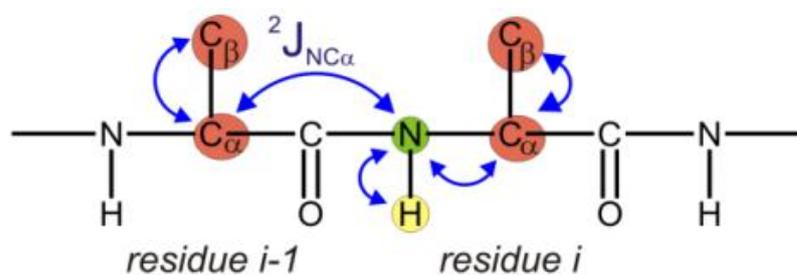


Figure 1.12 Schematic of HNCACB resonances.

1.11.5 C(CO)NH experiment

The C(CO)NH experiment is designed to correlate the ^1H and ^{15}N amide resonances of one residue with $^{13}\text{C}\alpha$ and all other ^{13}C side-chain resonances of its preceding residue via the intervening ^{13}CO spin by means of the $^1\text{J}(\text{NH})$, $^1\text{J}(\text{N,CO})$, $^1\text{J}(\text{C}\alpha,\text{CO})$ and $^1\text{J}(\text{C,C})$ coupling constants.

1.11.6 HCCH-TOCSY experiment

This spectrum is used for side chain assignment. The HCCH-TOCSY experiment is specifically designed to correlate side-chain aliphatic proton and ^{13}C resonances via $^1\text{J}(\text{CH})$ and $^1\text{J}(\text{CC})$ coupling constants. The experiment provides nearly complete assignments of all aliphatic ^1H and ^{13}C resonances, with the exception of some resonances of the long aliphatic side chains, such as Lys or Arg, for which substantial overlap remains.

1.11.7 ^{15}N edit NOESY

Magnetisation is exchanged between all hydrogens by the Nuclear Overhauser Effect. The magnetisation is transferred to neighbouring ^{15}N nuclei and back to ^1H for detection (Figure 1.13). This spectrum can be used to obtain restraints for structure calculations. It can also be used to help assignment for small to medium-sized proteins.

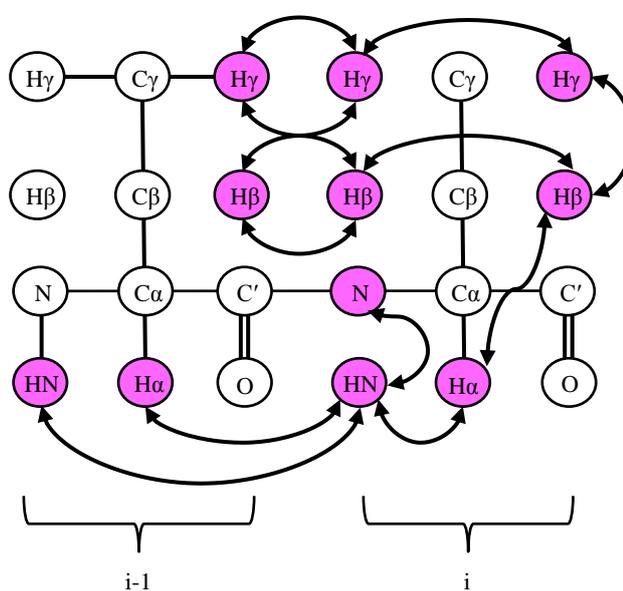


Figure 1.13 Schematic of ^{15}N edit NOESY resonances.

1.11.9 Three-dimensional model generation

Once the peaks in the NMR spectra have been annotated to specific atoms in the primary structure and their NOE and J-coupling interaction used to generate a list of constraints on the distances and bond angles between the atoms, a set of 3-dimensional spectra can be calculated (Wilder, 2000). Energy equations are set-up to include all the constraints, as well as restraints on the bond angles and length based on known values, and randomly generated 3D structure are reconfigured by various methods in order to minimize the energy of these equations. Constraints that consistently have high energy mismatch to the best models may be eliminated or modified and new constraints may be obtained by comparison of the model to the spectra, allowing models to be calculated. Generally, once the models cannot be improved further, an ensemble of the structures that have the lowest energies is reported, and an average model, in which the atoms are in their average position is generated then the energy minimized, may also be reported.

1.12 Saturation Transfer Difference NMR (STD)

The STD method is a one technique that permits detection of ligand binding to proteins (Viegas, 2001). This method does not require isotope labeling of the protein or the ligand (Figure 1.15). The STD method requires small quantities of protein (<20 μM) and the ligand is used in 50-100-fold molar excess over the protein.

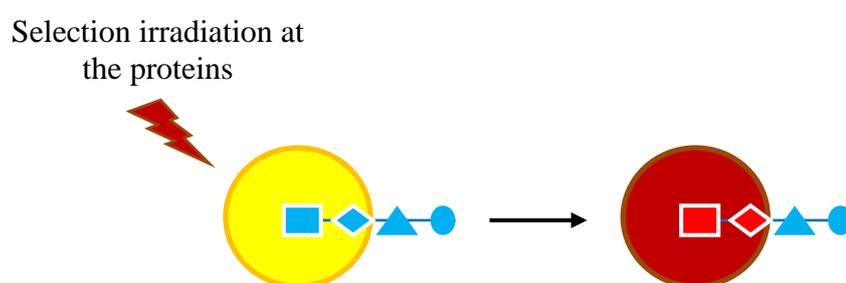


Figure 1.15 Schematic view of the STD experiment. The protein protons are selectively saturated at a specific frequency and the effect of the saturation is transmitted through the whole polypeptide. Any ligand interacting with the protein will also be affected by the saturation. During the irradiation time, the saturation is transferred to the bound ligand - first to the protons belonging to the ligand epitope, then to the rest of the ligand (Calle et al., 2011).

Basically, STD experiment involves subtracting the spectrum in which the protein was selectively saturated with signal intensities I_{SAT} (the *on-resonance* spectrum obtained by irradiation at a region of the spectrum that contains only resonance of protein/receptor) from one without protein saturation (the *off-resonance* spectrum), with signal intensities I_0 . In the difference spectrum ($I_{STD} = I_0 - I_{SAT}$), only signals from ligands that received saturation transfer from the protein will remain. Other compounds that are present but do not bind to the receptor will not receive saturation transfer. It means their signal of the *on-resonance* and *off-resonance* spectra are equal, so that after subtraction no signals will appear, which suggests that the protein cannot bind to these ligands. If the signal from the ligand after subtraction of the *on-resonance* and *off-resonance* spectra indicates a difference, it means protein can bind to the ligands (Figure 1.16). For a molecule that binds to the receptor, only the signals of the hydrogen that are in close contact to the protein and receive magnetization transfer will appear in the difference spectrum and from those, the ones that are closer to the protein will have more intense signals, owing to a more efficient saturation transfer.

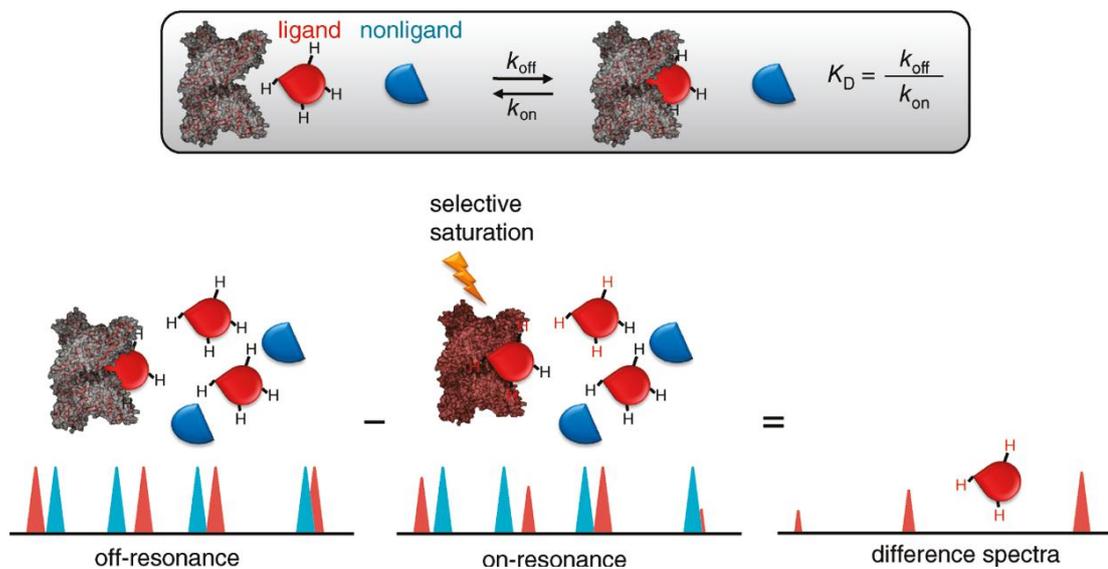


Figure 1.16 Scheme of the STD-NMR experiment. The exchange between free and bound ligand allows intermolecular transfer of magnetization from the receptor to the bound small molecule (Viegas et al., 2011).

1.13 Carbohydrate microarrays

A microarray is a 2D array on a solid substrate, such as a glass slide or silicon thin-film cell (Chang, 1983). In 1983, microarrays were first introduced and demonstrated in antibody microarrays by Chang (1983). He investigated specific cell surface antigens and incubated with arranged antibody spots on a solid surface. In the 1990's gene chip were introduced in the biotechnology industry and their use started and grow. Nowadays, microarrays are widely used, including DNA microarrays, antibody microarrays, protein microarrays and carbohydrate arrays (glycoarrays) (Wang et al., 2002).

In the last decade, carbohydrate microarrays have had a wide range of applications in biological and biomedical research (Park et al., 2012). These arrays consist of carbohydrate samples coated on a solid surface in a microarray format. This technique is used for rapid analysis of the glycan binding properties of lectins and antibodies, the quantitative measurements of glycan-protein interactions, detection of cells and pathogens, and identification of disease-related anti-glycan antibodies for diagnosis.

1.14 Objectives

In order to improve on the current understanding of structure function relationships in plant β -galactosidases, the objective of this thesis include:

- 1) To construct improved expression constructs for rice β -galactosidase 1 (OsBGal1) and β -galactosidase 1 C-terminal domain (OsBGal1Cter).
- 2) To produce milligram quantities of highly purified OsBgal1 and OsBGal1 Cter from a recombinant *Escherichia coli* or yeast system for structural studies.
- 3) To determine the structure of OsBGal1Cter and/or OsBGal1 by protein crystallography or NMR.
- 4) To determine whether OsBGal1 Cter binds to carbohydrates, such as simple sugars or cell wall components.

CHAPTER II

MATERIALS AND METHODS

2.1 Materials

2.1.1 Oligonucleotides primers and optimized OsBGal1 gene.

All oligonucleotide primers used to amplify the C-terminal domain of rice β -galactosidase 1 (OsBGal1) are shown in Table 2.1. OsBGal1Cter_For, OsBGal1 Cter Rev primers were ordered from Life Technologies, Japan, Ltd. OsBGal1Cter For_2211 and OsBGal1 Cter_STOP_Rev primers were ordered from Proligo Singapore Pty. Ltd. (Singapore).

The *Optimized OsBGal1* gene, which was optimized for expression in *Pichia pastoris*, was ordered from Genscript Corp. and provided in the PUC57 plasmid. For further cloning, this gene was amplified by *Pfu* DNA polymerase with the M13 Forward and M13 Reverse primers.

Table 2.1 Primers for OsBGal1 Cter amplification.

Primer	Primer sequence (5'->3')	T _m (°C)
OsBGal1Cter For_ 2211	CACCCGGACAGTCTCAGGTGTC	60
OsBGal1 Cter_STOP_Rev	ATCAAGCACAAGCCCACC	64
OsBGal1Cter_For	GGGCATATGCGGACAGTCTCAGGTGT	72
OsBGal1 Cter_Rev	GCCGGATCCATCAAGCACAAGCCCACC	88

2.1.2 Plasmids, bacterial and yeast strains

Plasmids for this work included pENTRTM/D-TOPO from Invitrogen (Life Technologies, Carlsbad, CA, USA), pET15b(+) and pET32b(+) from Novagen (Merck Biosciences, Madison, WI, USA), pET32a/DEST (Opassiri et al., 2006), pUC57 (GenScript, Piscataway, NJ, USA), and pPICZ α BNH8 (Toonkool et al., 2006). Their antibiotic resistance and sizes are shown in Table 2.2.

The bacteria used for DNA cloning were *Escherichia coli* strains DH5 α and TOP10 (Invitrogen), while strain Origami(DE3) and Origami B(DE3) was used for recombinant protein expression. The *Pichia pastoris* host strain used for recombinant protein expression in yeast was SMD1168H. Some properties of these strains are shown in Table 2.3.

Table 2.2 Recombinant plasmids used for OsBGal1 Cter cloning and expression.

Recombinant Plasmid DNA	Antibiotic resistance	Total size (kb)
pENTR TM /D-TOPO- <i>OsBGal1 Cter</i>	Kanamycin (30 µg/ml)	~3.0
pET32a/DEST- <i>OsBGal1 Cter</i>	Ampicillin (50 µg/ml)	~6.2
pET15b(+)- <i>OsBGal1 Cter</i>	Ampicillin (50 µg/ml)	~6.1
pET32b(+)- <i>OsBGal1 Cter</i>	Ampicillin (50 µg/ml)	~6.3
pCU57- <i>OsBGal1opt</i>	Ampicillin (50 µg/ml)	~5.0
pPICZαB NH8- <i>OsBGal1opt</i>	Zeocin (100 µg/ml)	~6.0

Table 2.3 Bacterial and yeast strains.

strain	Antibiotic resistance	Genotype	Features
DH5 α	None	F^- <i>endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG</i> 80 <i>dlacZ</i> Δ M15 Δ (<i>lacZYA-argF</i>)U169, <i>hsdR17</i> (rK ⁻ mK ⁺) λ^-	A high copy number strain used for DNA manipulation
TOP10	None	F^- <i>mcrA</i> Δ (<i>mrr-hsdRMS-mcrBC</i>) Φ 80 <i>lacZ</i> Δ M15 Δ <i>lacX74 recA1 araD139</i> Δ (<i>araleu</i>)7697 <i>galU galK rpsL</i> (StrR) <i>endA1 nupG</i> λ^-	stable replication of high-copy number plasmids
Origami B(DE3)	Tetracyclin (12.5 μ g/ml) Kanamycin (30 μ g/ml)	F^- <i>ompT hsdS_B</i> (r _B ⁻ m _B ⁻) <i>gal dcm lacY1 aphC</i> (DE3) <i>gor522::Tn10 trxB</i> (Kan ^R , Tet ^R)	Enhances the formation of disulfide bonds in the cytoplasm for greater yield of active protein

Table 2.3 Bacterial and yeast strains (Continued).

strain	Antibiotic resistance	Genotype	Features
Origami (DE3)	Kanamycin (30 µg/ml), Tetracyclin (12.5 µg/ml)	$\Delta(ara-leu)7697$ $\Delta lacX74 \Delta phoA$ <i>PvuII phoR araD139</i> <i>ahpC galE</i> <i>galK rpsL F'[lac+</i> <i>lacIq pro]</i> (DE3) <i>gor522::Tn10 trxB</i> pLacI (CamR, KanR, StrR, TerR)	Enhances the formation of disulfide bonds in the cytoplasm for greater yield of active protein
SMD1168H	None	<i>pep4::URA3 his4</i> <i>ura3</i>	Lacking protease A activity to reduce proteolysis of recombinant protein.

2.1.3 Chemicals and reagents

Ammonium persulfate, calcium chloride, cobalt(II) chloride, disodium ethylenediamine tetraacetate (EDTA), glucose, sodium acetate, sodium carbonate, methanol, sodium dodecyl sulfate (SDS), sodium chloride (NaCl), and sodium hydroxide were purchased from Carlo ERBA (Rodano, Milano, Italy). Acrylamide, N,N-methylene-bis-acrylamide, ampicillin, kanamycin were purchased from BIO BASIC INC (Marg, Mumbai, India). Agar, peptone, yeast extract and yeast nitrogen base w/o amino acid and ammonium sulphate were purchased from HIMEDIA Laboratories Pvt. Ltd (Markham, Ontario, Canada). DNaseI, lysozyme, bovine serum albumin, phenylmethylsulfonylfluoride (PMSF), deuterium oxide (D₂O), and para-nitrophenyl (pNP)- β -D-galactoside, were purchased from Sigma (St. Louis, MO, USA). *Taq* DNA polymerase, *Pfu* DNA polymerase, agarose and deoxynucleoside triphosphates (dNTPs) were purchased from Promega (Madison, WI, USA).

IMAC sepharoseTM 6 Fast Flow resin was purchased from GE Healthcare (Little Chalfont, UK). Ultrafiltration membranes 30 kDa MW, 10 kDa MW, 5 kDa MW cutoff, Ultrafree MC 0.22 μ m and 0.45 μ m filters were from Millipore (Bedford, MA, USA) Some of the kits used were Bio-RAD protein assay kit (Bio-RAD Corp., Hercules, CA, USA), QIA prep spin miniprep plasmid extraction kit (QIAGEN) and Perfectprep Gel Cleanup kit (QIAGEN, Hilden, Germany). Other chemicals and molecular reagents used but not listed here were purchased from a variety of suppliers. Liquid nitrogen was obtained from the Center for Scientific and Technological Equipment, Suranaree University of Technology.

2.2 General Methods for *Escherichia coli*

2.2.1 Amplification of *OsBGal1 Cter* cDNA

The cDNA fragments encoding the C-terminal domain of rice β -galactosidase 1 protein (*OsBGal1 Cter*) was generated by PCR with the *OsBGal1 Cter For_2211* primer and *OsBGal1 Cter_STOP_Rev* primer (Table 2.1). The amplification was done with 30 cycles (Table 2.4). The fifty microliter PCR reaction was composed of 0.2 mM dNTP, 1X *Pfu* buffer (Promega, Madison, WI, USA), 0.8 μ M *OsBGal1 Cter_For_211* primer, 0.8 μ M *OsBGal1 Cter stop_Rev* primer, 100 ng *OsBGal1* cDNA, 0.05 U *Pfu* polymerase (Promega) and sterile distilled water up to 50 μ L total.

To produce a new construct of *OsBGal1 Cter* with a thrombin site after the fusion tag (*OsBGal1_Cter_tt*), the cDNA fragments encoding the C-terminal domain of the *OsBGal1* β -galactosidase protein was generated by PCR with the *OsBGal1 Cter_For* primer, which inserted an *NdeI* restriction site, and *OsBgal1 Cter_Rev* primer, which inserted a *BamHI* restriction site. The amplification was achieved by 30 cycles (Table 2.4). The fifty microliter PCR reaction was composed of 0.2 mM dNTP, 1X *ExTaq* buffer (TAKARA, Shiga, Japan), 0.8 μ M *OsBGal1_Cter_forward* primer, 0.8 μ M *OsBGal1_Cter_reverse* primer, 100 ng *OsBGal1* cDNA, 0.4 U *ExTaq* polymerase (TAKARA, Japan) and sterile distilled water up to 50 μ L total.

Table 2.4 Cycling parameters for *OsBGall Cter* amplification.

Segment	Cycles	Temperature	Time
1	1	95°C	5 min
		98°C	5 min
2	30	95°C	45 s for <i>Pfu</i> polymerase (Promega)
		98°C	10 s for Ex <i>Taq</i> polymerase (TAKARA)
		58°C	30 s
		72°C	1 min
3	1	72°C	10 min

2.2.2 Purification of PCR products from agarose gel

To purify PCR products, the DNA band cut from an agarose gel was extracted with the QIAQuick gel purification kit (QIAGEN) by the vendor's recommended protocol. The PCR products from 100 μ L reactions were separated on 1% agarose gels and stained with 10 mg/mL ethidium bromide. The DNA band was excised and placed in a 1.5 mL microtube. Three volumes of QG buffer were added to 1 volume of gel (estimated by weight) and the tube was incubated at 50°C with shaking for 10 min or until the gel was completely dissolved. The sample was applied to the QIAQuick column, which was centrifuged for 1 min at 15,000 rpm, and the flow through was discarded. Five hundred microliters of QG buffer was added to the column and it was centrifuged at 15,000 rpm for 1 min. Seven hundred fifty microliters of PE buffer was added to the column and centrifuged at 15,000 rpm for 1 min. The flow-through solution was discarded and the column was centrifuged for

an additional 1 min. The column was placed into a new 1.5 mL microtube, and 50 μ L of distilled water was added, incubated for 1 min and centrifuged at 15,000 rpm for 1 min. The eluted DNA product was kept at -30°C.

2.2.3 Quantification and expected yield of DNA

Two microliters of DNA solution were mixed with distilled water to give a final volume of 100 μ L. The absorbance at 260 (A_{260}) and 280 nm (A_{280}) of the 100 μ L dilution solution was measured with a DU-7400 UV/VIS Spectrophotometer (Beckman, USA). The A_{260}/A_{280} ratio of 1.6-1.8 indicated good purity of DNA. One unit of absorbance at A_{260} nm is equivalent to 50 μ g/mL of pure DNA (Sambrook et al., 1989). The DNA concentration was calculated with the following equation:

$$\text{Concentration } (\mu\text{g/mL}) \text{ of DNA} = (A_{260} \times \text{dilution factor} \times 50 \mu\text{g/mL})$$

2.2.4 Cloning of DNA fragments into pENTR/D-TOPO vector

The cloning reaction was done by added 1 μ L of freshly purified PCR product, 1 μ L of salt solution (1.2 M NaCl and 0.06 M MgCl_2), 1 μ L of pENTR/D-TOPO vector (Invitrogen) and sterile double distilled water up to 6 μ L. Then, the reaction was mixed gently and incubated for 10 min at 23°C. After that, the reaction was placed on ice and used to transform One Shot® TOP10 chemically competent *E. coli* (Invitrogen).

2.2.5 Ligation of DNA fragments into vectors

The OsBGal1 Cter PCR products were cloned into the *Nde*I and *Bam*HI sites of pET15b(+) and then, the recombinant DNA was subcloned into the *Nco*I and *Bam*HI sites of pET32b(+) and transformed into DH5 α *E. coli*. The gel purified inserts for *OsBGal1 Cter* were ligated into the *Nde*I and *Bam*HI sites of pET15b(+) and *Nco*I and *Bam*HI site of pET32b(+). The reaction mixture (20 μ L) was composed of 4.4 μ L of linearized pET-15b vector (0.03 pmole), 2.8 μ L of eluted DNA fragment (0.15 pmole), 10 μ L of 2X ligation mix (TAKARA: 132 mM Tris-HCl, pH 7.6, 13.2 mM MgCl₂, 20 mM DTT and 0.2 mM ATP) and dH₂O up to 20 μ L total. The reaction solution was incubated at 16°C for 15 min. The ligation was used to transform bacteria immediately or kept at -30°C.

2.2.6 Preparation of competent *Escherichia coli* cells

Competent *E. coli* strain DH5 α or Origami B(DE3) cells were prepared by the CaCl₂ method. A single colony of *E. coli* was inoculated into 5 mL of LB broth and grown at 37°C overnight with shaking at 200 rpm. Fifty microliters of inoculum culture was inoculated into a 250 mL flask containing 50 mL of LB broth and grown at 37°C with 250 rpm shaking until the optical density (OD) at 600 nm (OD₆₀₀) was about 0.4. The culture was transferred into a pre-chilled sterile polypropylene tube, chilled on ice for 5 min and centrifuged at 4,000 rpm, 4°C for 10 min to collect cell pellets. The cell pellets were resuspended with 10 mL of ice-cold 0.1 M MgCl₂. The resuspended cells were pelleted by centrifugation at 4,000 rpm, 4°C for 10 min. The pellets were resuspended with 20 mL of ice-cold 0.1 M CaCl₂. The suspension was kept on ice for at least 20 min. The resuspended cells were pelleted by

centrifugation at 4,000 rpm, 4°C for 10 min and the supernatant was decanted. Then, the cells were resuspended with 1 mL of ice cold 85 mM CaCl₂ containing 15% glycerol. The competent cells were aliquotted into 1.5 mL microcentrifuge tubes (100 µL/tube) and used immediately or kept at -80°C.

2.2.7 Transformation and selection of *Escherichia coli*

For transformation, an aliquot of frozen competent cells was thawed 20 min on ice. The ligation reaction or plasmid (10-200 ng) was added to fresh or thawed competent cells, mixed by pipetting up and down, and incubated on ice for 30 min. The cells were transformed with plasmid by heat shocking them at 42°C for 45-60 s and quickly chilling on ice for 2-5 min. One hundred microliters of LB broth were added into the transformed competent cells and they were incubated at 37°C with or without shaking at 200 rpm for 1 h. The cells were spread on an agar plate containing the appropriate antibiotic(s) as shown in Table 2.2. For ligations into pET-15b(+) and pET-32b(+) vectors, the transformed DH5α cells were spread on LB plates containing 50 µg/mL ampicillin and incubated at 37°C overnight to select for recombinant colonies. For OrigamiB(DE3) cells transformed with pET32b-*OsBGall_Cter* plasmids, the cells were spread on LB plates containing 30 µg/mL kanamycin, 12.5 µg/mL tetracycline and 50 µg/mL ampicillin.

2.3 Expression of OsBGal1 Cter protein in *E. coli*

For expression in LB broth, a single colony was inoculated into 5 mL of LB broth containing 15 µg/mL kanamycin, 12.5 µg/mL tetracycline, and 50 µg/mL ampicillin, and the culture was incubated at 37°C in an incubator shaker overnight. A 1% overnight culture of starter culture was inoculated in LB broth containing 30 µg/mL kanamycin, 12.5 µg/mL tetracycline, and 50 µg/mL ampicillin, and incubated for 4-5 h. at 37°C, 200 rpm until the OD₆₀₀ reached 0.5-0.6. Then, IPTG was added to a final concentration of 0.4 mM for induction of protein expression, and the culture was grown for a further 18 h at 20°C. The bacterial cells were precooled on ice 20 min and collected by centrifugation at 5,000 rpm for 15 min., and then the cell pellets were kept at -80°C. Frozen pellets were thawed and extracted on ice with lysis buffer (20 mM Tris-HCl, pH 8.0, 0.2 mg/mL lysozyme, 2 mM PMSF and 1% Triton X-100). The cell suspension was then incubated at room temperature for 30 min. Cells were completely broken using a BANDELIN Sonopuls HD2200 Ultrasonic homogenizer (United Instrument Co. Ltd., Berlin, Germany). Sonication was done on ice for 3 times with 10% power output, for 30 s with cooling for 1 min in between. The soluble protein was harvested by centrifugation at 12,000 g for 15 min at 4°C.

For expression in M9 media, a single freshly transformed colony of *E. coli* strain Origami B(DE3) containing pET32b(+)/OsBGal1Cterm was picked and inoculated into 3 mL LB containing 50 µg/mL ampicillin, 30 µg/mL kanamycin and 12.5 µg/mL tetracycline and cultured overnight. The culture mixed with 60 mL LB broth containing the same antibiotics and incubated with shaking at 240 rpm, at 37°C,

until the OD₆₀₀ reached 1.0. The cultured cells were collected by centrifugation at 2,500 rpm at room temperature for 10 min. The pellets were resuspended with 1 L M9 media. The cells were cultured until the OD₆₀₀ reached 0.6-0.8. Then, IPTG (1 M) was added to 0.4 mM final concentration in the 1 L M9 media containing 50 µg/mL ampicillin, 30 µg/mL kanamycin and 12.5 µg/mL tetracycline, and the incubation continued for 18 h at 20°C. Induced cultures were harvested by centrifugation at 6,000 x g at 4°C for 15 min. The pellets were kept at -80°C. The cell pellets were resuspended in 5 mL/gram freshly prepared extraction buffer and incubated at room temperature for 30 min. Then, the cells were sonicated for 1 min 30 s 2 times, and incubated on ice for 5 min and soluble and insoluble fractions were separated by ultracentrifugation at 20,000 rpm at 4°C for 20 min.

2.4 Purification of OsBGal1 Cter protein

2.4.1 Thioredoxin fusion protein

The fusion protein was purified by Immobilized Metal-Affinity Chromatography (IMAC) on Talon cobalt resin (Clontech, TAKARA) at 4°C. To bind the soluble protein, one milliliter of cobalt resin was added to ten milliliters of soluble protein extract, and then gently shaken upside down (60 rpm) for 30 min. The resin bound protein was put into a column, and then, washed with 10 column volumes (CV) of equilibration buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl), then washed with 5 CV of 10 mM imidazole in equilibration buffer, and 5 CV of 20 mM imidazole in equilibration buffer. Finally, the bound protein was eluted with 250 mM imidazole in equilibration buffer. The protein band patterns of the purification fractions were checked by SDS-PAGE. Then, the fractions containing OsBGal1 Cter were pooled,

imidazole was removed and the buffer changed to 20 mM Tris-HCl, pH 8.0, in a 10 kDa molecular weight cut-off (MWCO) Centricon centrifugal filter.

2.4.2 Removal of the fusion tag and purification of tag free protein

Since thioredoxin fusion tag might interfere with crystallization and complicate the Nuclear Magnetic Resonance (NMR) spectra, only the non-fusion protein was used for crystallization and NMR. To remove the N-terminal fusion tag, OsBGal1 Cter fusion protein was cleaved with 1 μ g thrombin protease (Sigma) per milligram of protein in 20 mM Tris-HCl, pH 8.0 at 4°C for 2 h. The tag free protein was removed from the fusion tag and uncleaved fusion protein by loading the protein onto the Talon cobalt resin (Clontech, Mountain View, CA, USA.). The solution of unbound protein was collected and the resin was washed with 10 and 20 mM imidazole in equilibration buffer (20 mM Tris HCl, pH 8.0, 300 mM NaCl) and the fusion tag and any uncleaved protein was eluted from the resin with 250 mM imidazole in equilibration buffer. The protein contents of all fractions were assessed by SDS-PAGE. The fractions containing free OsBGal1 Cter were combined, and the protein was concentrated and the buffer changed to 20 mM phosphate buffer, pH 7.4, containing 0.1 M NaCl, in a 5 kDa MWCO Centricon centrifugal filter to achieve a final volume of 1.0 mL. The protein was then passed through a Superdex 75 gel-filtration chromatography column (GE Healthcare) equilibrated and eluted with 0.1 M NaCl in 20 mM phosphate buffer, pH 7.4, at a flow rate of 0.5 mL/min. The peak fractions containing the purified protein were pooled and concentrated, and the buffer changed to 20 mM phosphate buffer, pH 7.4, containing 0.5 M NaCl with a 5 kDa MWCO Centricon centrifugal filter. The protease was removed from protein by

adsorbing it to benzamidine resin (GE Healthcare) equilibrated in 0.5 M NaCl, 20 mM phosphate buffer, pH 7.4. The protein band pattern of the unbound fraction was checked on SDS-PAGE. The purified protein was concentrated and the buffer changed to 20 mM sodium phosphate buffer, pH 7.4, containing 0.1 M NaCl with a 5 kDa MWCO Centricon centrifugal filter. The purified protein was used for NMR.

2.5 Protein analysis

2.5.1 Sodium Dodecyl Sulfate Poly Acrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE was done by the method of Laemmli (Laemmli, 1970). The 15%-17% separating gel consisted of 15%-17% (w/v) acrylamide, 0.375 M Tris-HCl, pH 8.8, 0.1% SDS, 0.1% ammonium persulfate (APS) and 0.05% *N,N',N'',N'''*-tetramethylethylenediamine (TEMED). The 4% stacking gel consisted of 4% (w/v) acrylamide, 0.126 M Tris-HCl, pH 6.8, 0.1% SDS, 0.1% APS and 0.05% TEMED. Each protein sample was mixed with 4 volumes of 5X SDS sample buffer (0.05 M Tris-HCl, pH 6.8, 10% SDS, 40% (v/v) glycerol, 0.01% bromophenol blue, and 20% (v/v) β -mercaptoethanol), and boiled for 5 min before loading to the gel. The upper and lower reservoirs of the electrophoresis apparatus were filled with electrophoresis buffer (0.025 M Tris, 0.129 M glycine, and 0.1% SDS, pH 8.3). Electrophoresis was carried out at a constant voltage of 160 V until the tracking dye reached the bottom of gel. Protein bands were visualized by staining in staining solution containing 0.1% (w/v) Coomassie brilliant blue R-250, 40% (v/v) methanol and 10% (v/v) acetic acid for 30 min to 1 h, followed by several washes with destaining solution containing 40% (v/v) methanol and 10% (v/v) acetic acid. The relative molecular weight (M_r) of

the protein was estimated by comparison to a series of molecular weight markers (GE Healthcare): phosphorylase b (97 kDa), bovine serum albumin (66 kDa), ovalbumin (45 kDa), glyceraldehyde-3-P-dehydrogenase (36 kDa), bovine carbonic anhydrase (29 kDa), bovine pancreatic trypsinogen (24 kDa), soy bean trypsin inhibitor (21 kDa), and bovine milk α -lactalbumin (14 kDa).

2.5.2 Bio-Rad protein assay

The protein concentrations were estimated according to the Bradford method (Bradford, 1976) with a Bio-Rad kit (Bio-Rad Corp., Hercules, CA, USA) and bovine serum albumin as a standard (0-10 μ g). The reaction mixture contained suitably diluted protein and distilled water in a total volume of 800 μ L. To start the reaction, 200 μ L of concentrated Bio-Rad Bradford reagent was added and the tubes were vigorously mixed. The reaction was incubated at room temperature for 10 min, and the absorbance at 595 nm measured with a Genesys 10 UV spectrophotometer (Rochester, NY, USA).

2.6 Determination of the native molecular weight (MW) of the OsBGal1 Cter protein by gel filtration column chromatography

The OsBGal1 Cter native MW was determined by gel filtration on a Superdex S75 chromatography column (GE Healthcare). The column was equilibrated with 3 volumes of 20 mM sodium phosphate buffer, pH 7.4, containing 0.1 M NaCl at 4°C, at a flow rate of 0.5 mL/min. A 1 mL aliquot of a mixture of 5 standard protein gel filtration (15 mg/mL blue dextran, 6 mg/ml BSA, 5 mg/mL ovalbumin, 5 mg/mL α -chymotrypsinogen A, 5 mg/mL ribonuclease A and 4 mg/mL aprotinin) in elution

buffer was loaded onto the column and eluted with 120 mL of the buffer at 4°C. Fractions (1 mL) were collected using the fraction collector and the column was washed with the buffer. The OsBGal1 Cter protein was applied to the column with the same procedure used for the standards. The absorbance at 280 nm of all fractions was measured.

For the molecular weight determination, the standard curve obtained from log MW versus K_{av} of gel filtration standard was then constructed. The K_{av} was calculated with the following equation:

$$K_{av} = \frac{\text{Elution volume (Ve)} - \text{Void volume (Vo)}}{\text{Total bed volume (Vt)} - \text{Void volume (Vo)}}$$

2.7 Western immunoblot analysis

The proteins were separated by SDS-PAGE, as described in Section 2.3.1. The protein in the gel was transferred by electroblotting onto a nitrocellulose membrane with transfer buffer (25 mM Tris-HCl pH 8.0, 190 mM glycine, and 20% (v/v) methanol) at 100 volts, on ice for 60 min. The membrane was blocked with 5% skim milk in phosphate buffered saline (PBS; 20 mM phosphate buffer, pH 7.4, 150 mM NaCl) at room temperature for at least 60 min or at 4°C overnight. The membrane was washed 3 times for 5 min each with wash buffer (PBST; PBS containing 0.05% Tween 20). A mouse monoclonal anti-poly Histidine IgG2a isotype antibody (Sigma) or polyclonal antibodies with specific peptide derived from OsGBal1 Cter (N'-QIESYGEPEFHTAKC-C') produced by GenScript Corporation (Piscataway, NJ, USA) was used as the primary antibody at a 1:2000 dilution in PBST, and incubated with the membrane for 2 h. The membrane was washed 3 times

for 5 min each with wash buffer. The membrane was incubated with a 1:2000 dilution of peroxidase conjugated goat anti-mouse IgG (Sigma) or with a 1:2000 dilution of peroxidase conjugated goat anti-rabbit IgG (Sigma) as the secondary antibody in wash buffer for 2 h and then was washed 3 times for 5 min each with wash buffer. Finally, the membrane was developed with aminoethyl carbazole substrate kit (Sigma) for 30 min or until a band was observed on the membrane.

2.8 Protein crystallization

2.8.1 Preliminary sample preparation

The purified protein was concentrated using a 5 kDa MWCO Centricon centrifugal filter at 4000 *g*, approximately to 10 mg/mL with 20 mM Tris-HCl, pH 8.0, and kept at 4°C until crystallization. Before crystallization, the protein solution was filtered through an Ultrafree-MC 0.22 μm filter (Millipore) at 4000 *g* for 5 min. to eliminate microbial contamination, dust, micro-crystals, and precipitated protein.

2.8.2 Initial screening for crystallization conditions

Screening for crystallization conditions was done with the Crystal Screen High Throughput HR2-130 (Hampton Research, CA, USA) kits, Wizard I & II kits (Emerald BioSystems, Inc., Seattle, WA, USA) and the JY screen of polyethylene glycol in divalent salts (Assoc. Prof. Dr. Jirandon Yuvaniyama, Mahidol University, Bangkok, Thailand). Microbatch screening was performed with 96 well plastic plates (Hampton Research) or 60 well plates (Nunc, Roskilde, Denmark). Ten microliters of 100% paraffin oil (Paraffin, highly liquid, Merck) was first pipetted into each well, then 1 μL of precipitant solution was added into the cone shaped depression well.

Then, 1 μL of 10 mg/mL protein was mixed by pipetting under an oil layer. To ensure a single drop of protein and precipitant was formed, each well was carefully checked under a Zeiss Stemi 2000-C stereo microscope (Zeiss Corp, NJ, USA). If a single drop was not obtained, a cat whisker was used to push the separated drops together under oil. The crystallization plate was covered with the plate cover to prevent dust and debris from outside, placed on a moist sponge in a plastic box, to maintain humidity, and incubated at 15°C.

2.9 Protein NMR

2.9.1 Preparation of OsBGal1 Cter sample for NMR

Unlabeled or ^{15}N - or $^{15}\text{N}/^{13}\text{C}$ -labeled OsBGal1 Cter protein was prepared by expressing the protein in modified minimal M9 medium (Section 2.3) supplemented with ^{15}N ammonium chloride or ^{15}N ammonium chloride/ ^{13}C D-glucose (Isotec Inc., Miamisburg, OH, USA) as a nitrogen and carbon sources for the labeled samples. NMR samples were prepared in the range of 0.25 to 1.0 mM protein in 20 mM sodium phosphate, pH 8.0, and 100 mM NaCl in D_2O .

2.9.2 NMR data acquisition

The ^{15}N -HSQC, $^{13}\text{C}/^{15}\text{N}$ -HSQC, HNCOC, CBCA(CO)NH, HNCACB, C(CO)NH and HCCH-TOCSY NMR spectra were acquired at 25°C on a Varian UNITY INOVA 500, while ^{15}N -NOESY spectra was obtained from a Bruker AVANCE DRX 800 spectrometer equipped with a triple resonance pulse field gradient probe.

2.9.3 NMR data processing and analysis

All spectra were processed and displayed using NMRPipe and NMRDraw (Delaglio et al., 1995) and analyzed by Pipp (Garrett et al., 1991). The secondary structure of OsBGal1 Cter was characterized from the chemical shift indexes (CSI) based on C α and C β chemical shifts (Metzler et al., 1993; Wishart et al., 1995) and NOE connectivity patterns.

2.9.4 Structure calculation

Structure calculations for OsBGal1 Cter were performed using the YASAP protocol (Nilges et al., 1988) within the X-PLOR software package (Brünger, 1992), as described by Bagby et al., 1994.

2.9.5 OsBGal1 Cter titration.

OsBGal1 Cter (0.15 mM) uniformly labeled with ^{15}N was titrated with galactose at 1:1 to a 1:10 molar ratio of protein:sugar. For rhamnose, glucose and raffinose, the molar ratio of 1:10 protein:sugar was used. The ^1H - ^{15}N HSQC spectra were collected in the presence and absence of the sugars and observed for differences. Between runs, the sugar was washed out with D $_2$ O buffer in a centricon centrifugal filter, and a new control spectrum was collected before adding the next sugar.

2.10 General methods for protein expression in yeast

2.10.1 DNA subcloning of the *optimized OsBGal1* cDNA

When the OsBGal1 protein from rice was initially expressed in *Pichia pastoris* from its native cDNA, expression was low (Mallika Changtarangsee,

unpublished data), because the *OsBGall* cDNA codon usage does not match the codon bias of the host cells. So, an optimized *OsBGall* cDNA suitable for expression in the yeast was synthesized by Genscript Corporation (Piscataway, NJ, USA). The cDNA sequence was optimized by GenScript with their propriety software to maximize codon frequency in *Pichia pastoris*, while minimizing mRNA structures and repetitive sequences. The optimized *OsBGall* cDNA was inserted into the pUC57 vector by GenScript Corporation. The optimized *OsBGall* cDNA was designed with the cloning restriction sites of *Pst*I and *Xba*I at the 5' and 3' ends, respectively. The gene was amplified by *Pfu* DNA polymerase with the M13 forward primer and M13 reverse primers (Table 2.5) according to the conditions in Table 2.6. The PCR product (~2.5 kb) and pPICZ α BNH8 expression vector were digested with *Pst*I and *Xba*I restriction endonucleases at 37°C overnight. The digested fragments and vector were analyzed by 1% agarose gel electrophoresis and purified from the gel by the Perfectprep® Gel Cleanup (Eppendorf, Hamburg, Germany). The *optimized OsBGall* cDNA and pPICZ α BNH8 vector were ligated together with T4 ligase (Promega) and a ratio of 3:1 PCR product per plasmid, and the reaction was incubated at 14°C overnight. The reaction product was transformed into DH5 α competent cells and selected on a 25 μ g/mL zeocin LB plate. The recombinant expression vector was cut checked with *Pst*I and *Xba*I restriction endonucleases to confirm the insert, and the DNA was sequenced at Macrogen (Seoul, Rep. of Korea).

Table 2.5 Primers for optimized OsBGal1 amplification.

Primer	sequence (5'→3')	T _m (°C)
M13 forward	GTTTCCCAGTCACGAC	52
M13 reverse	CAGGAAACAGCTATGAC	52

Table 2.6 Cycling parameters for optimized OsBGal1 amplification.

Segment	Cycles	Temperature	Time
1	1	95°C	5 min
2	30	95°C	1 min
		52°C	1 min
		72°C	5 min 30 s
3	1	72°C	5 min

2.10.2 Preparation of competent *Pichia pastoris* strain SMD1168H

A glycerol stock of *P. pastoris* strain SMD1168H was streaked on a yeast extract peptone dextrose (YPD) plate without antibiotic, which was then incubated at 28°C for 2-3 days. A single colony was inoculated into 500 mL YPD broth in a 2 liter flask and grown at 28°C with 220 rpm shaking overnight until the OD₆₀₀ reached 1.3-1.5. The cells were collected by centrifugation at 1,500 x g for 5 min at 4°C. The pellet was resuspended 2 times in 500 mL and 250 mL of ice-cold sterile water and collected by centrifugation at 1,500 x g for 5 min at 4°C each time. Next, the pellet was resuspended with 20 mL of ice-cold 1 M sorbitol and centrifuged

at 1,500 x g for 5 min at 4°C. Finally, the pellet was resuspended and kept in 1 mL of ice-cold 1 M sorbitol.

2.10.3 DNA preparation for electroporation

Seven to ten micrograms of circular recombinant pPICZ α BNH₈-*OsBGallOpt* plasmid was linearized with 1.6 U *Sac*I (New England Bio Labs) in 50 μ L of reaction that was incubated at 37°C overnight. Linearization of the plasmid was checked by electrophoresis of 2 μ L of reaction on a 1% agarose gel. The restriction endonuclease was inactivated by heating at 65°C for 10 min. Linear DNA was precipitated by adding 0.1 volume of 3 M sodium acetate, pH 5.2, and 2.5 volumes of 100% ethanol, and then incubating at -20°C for 30 min. Precipitated DNA was collected by centrifugation at 12,000 x g for 10 min. The DNA pellet was washed with 500 mL of 70% ethanol and centrifuged at 12,000 x g for 10 min. All ethanol was removed by inverting the tube on tissue paper for 10 min. The DNA pellet was dissolved in 5-10 μ L of sterile deionized water.

Linearized pPICZ α BNH₈-*OsBGallOpt* (7 μ g) was added to a microcentrifuge tube containing 80 μ L of *P. pastoris* competent cells and mixed gently by pipetting. The cell mixture was transferred to a pre-cooled 0.2 cm electroporation cuvette. The cuvette with the cells was incubated on ice for 5 min. The linearized vector was transformed into the *P. pastoris* by GenePulser electroporator (Bio-Rad) with the parameters of 1.5 kV, 25 μ F and 400 Ω . After that, 1 mL of 1 M sorbitol was immediately added to the electroporated cells and they were incubated at 30°C for 1 h without shaking. Two hundred microliters of cell solution

was spread on a YPDS plate containing 100 µg/mL zeocin antibiotic. The YPDS plate was incubated at 28°C for 3-5 days.

2.10.4 Expression of recombinant OsBGal1 in *P. pastoris*

A single colony of pPICZαBNH₈-*OsBGal1Opt* that had been selected on a 500 µg/mL zeocin YPD plate was inoculated into 500 mL of BMGY medium containing 100 µg/mL zeocin and grown in a shaking incubator (220 rpm) at 28°C until the cell culture OD₆₀₀ reached 2-3. Cells were harvested by centrifugation at 3000 x g for 5 min at 20°C and resuspended in 1000 mL BMMY medium at the final OD₆₀₀ of 1. Protein expression was induced by adding methanol to 1% (v/v) final concentration every 24 h for 7 days.

2.10.5 Purification of recombinant OsBGal1

The culture broth with secreted OsBGal1 was supplemented with PMSF to 1 mM final concentration and the pH was adjusted to 7.5 with K₂HPO₄. The adjusted media was mixed with Co²⁺-bound IMAC resin at room temperature for 30 min, the resin was poured into a column, washed with 10 CV of equilibration buffer, and the OsBGal1 protein eluted with 250 mM imidazole in the equilibration buffer. The active fractions were pooled and the buffer changed to 20 mM Tris-HCl, pH 8.0, by 30 kDa MWCO Centricon centrifugal filter at 3000 x g. The purified OsBGal1 was used to study the pH and temperature optima.

2.10.6 Determination of the optimal pH and temperature for recombinant OsBGal1

The optimum pH was determined by incubating 35 μg OsBGal1 enzyme in a reaction volume of 140 μL containing 100 mM universal buffer (citric acid, boric acid and trisodium phosphate), pH 3-12 at 0.5 pH unit increments, with 10 mM *para*-nitrophenyl- β -D-galactopyranoside (*p*NPGal) for 20 minutes at 55°C. The reaction was stopped with 100 μL of 2 M sodium carbonate. The absorbance of the *p*NP released was measured at 405 nm, which was compared to that of a *p*NP standard curve in the same solution to determine the amount released.

To determine the temperature optimum, 35 μg enzyme was incubated in 100 mM sodium acetate, pH 5.0, over a temperature range of 20 to 75°C in 5° increments for 10 min and then *p*NPGal was added at 10 mM final concentration to the reaction and incubated for 20 min. The amount of *p*NP released was determined as described above.

2.10.7 β -Galactosidase activity

The β -D-galactosidase activity of OsBGal1 was determined by mixing with 1 mM *p*NPGal in 50 mM sodium acetate, pH 5.0, in a reaction volume of 140 μL . The reaction was incubated at 55°C for 20 min and stopped with 70 μL of 2 M sodium carbonate. The *p*-nitrophenol (*p*NP) released was quantified by measuring the absorbance of *p*-nitrophenolate at 405 nm in an iEMS reader MF microtiterplate photometer (Labsystem iMES MF, Finland) and the absorbance of a control of reaction without enzyme was subtracted as the blank. The *p*NP released was

determined by comparing the absorbance with that of a *p*NP standard curve in the same buffer.

2.11 Carbohydrate microarray

The microarray with oligosaccharides and polysaccharides coated on a membrane was provided by Prof. William G. T. Willats of the University of Copenhagen, Denmark. The membrane was incubated with blocking solution (5% w/v skim milk powder in PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.7 mM KH₂PO₄, pH 7.5) 1 h, shaking at room temperature. Then, the membrane was incubated with 10 µg/mL purified OsBGal1 Cter in blocking solution at 4°C overnight with shaking (~100 rpm). Next, the membrane was washed with PBS, 1 time 10 min, at room temperature. The membrane was incubated with 1:500 diluted rabbit anti-OsBGal1 Cter antibody in blocking solution, 2 h at room temperature and washed with PBS 1 time, 10 min at room temperature. The membrane was incubated with a 1:2000 dilution of peroxidase-conjugated goat anti-rabbit IgG in blocking solution, as the secondary antibody, 2 h at room temperature. Finally, the membrane was washed with PBS, 1 time 10 min, at room and then detected using the aminoethyl carbazole substrate kit according to the supplier's directions.

1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12		
Ink	Ink	Ink	Ink	1																					
1	1*	1**	2	2*	2**	3	3*	3**	4	4*	4**	1	1*	1**	2	2*	2**	3	3*	3**	4	4*	4**	2	
5	5*	5**	6	6*	6**	7	7*	7**	8	8*	8**	5	5*	5**	6	6*	6**	7	7*	7**	8	8*	8**	3	
9	9*	9**	10	10*	10**	11	11*	11**	12	12*	12**	9	9*	9**	10	10*	10**	11	11*	11**	12	12*	12**	4	
13	13*	13**	14	14*	14**	15	15*	15**	16	16*	16**	13	13*	13**	14	14*	14**	15	15*	15**	16	16*	16**	5	
17	17*	17**	18	18*	18**	19	19*	19**	20	20*	20**	17	17*	17**	18	18*	18**	19	19*	19**	20	20*	20**	6	
21	21*	21**	22	22*	22**	23	23*	23**	24	24*	24**	21	21*	21**	22	22*	22**	23	23*	23**	24	24*	24**	7	
25	25*	25**	26	26*	26**	27	27*	27**	28	28*	28**	25	25*	25**	26	26*	26**	27	27*	27**	28	28*	28**	8	
29	29*	29**	30	30*	30**	31	31*	31**	32	32*	32**	29	29*	29**	30	30*	30**	31	31*	31**	32	32*	32**	9	
33	33*	33**	34	34*	34**	35	35*	35**	36	36*	36**	33	33*	33**	34	34*	34**	35	35*	35**	36	36*	36**	10	
37	37*	37**	38	38*	38**	39	39*	39**	40	40*	40**	37	37*	37**	38	38*	38**	39	39*	39**	40	40*	40**	11	
41	41*	41**	42	42*	42**	43	43*	43**	44	44*	44**	41	41*	41**	42	42*	42**	43	43*	43**	44	44*	44**	12	
45	45*	45**	46	46*	46**	47	47*	47**	48	48*	48**	45	45*	45**	46	46*	46**	47	47*	47**	48	48*	48**	13	
49	49*	49**	50	50*	50**	51	51*	51**	52	52*	52**	49	49*	49**	50	50*	50**	51	51*	51**	52	52*	52**	14	
53	53*	53**	54	54*	54**	55	55*	55**	56	56*	56**	53	53*	53**	54	54*	54**	55	55*	55**	56	56*	56**	15	
Ink	Ink	Ink	Ink	16																					
Ink	Ink	Ink	Ink	17																					
B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	18
B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	19
B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	20
Ink	Ink	Ink	Ink	21																					
57	57*	57**	58	58*	58**	59	59*	59**	60	60*	60**	57	57*	57**	58	58*	58**	59	59*	59**	60	60*	60**	22	
61	61*	61**	62	62*	62**	63	63*	63**	64	64*	64**	61	61*	61**	62	62*	62**	63	63*	63**	64	64*	64**	23	
65	65*	65**	66	66*	66**	67	67*	67**	68	68*	68**	65	65*	65**	66	66*	66**	67	67*	67**	68	68*	68**	24	
69	69*	69**	70	70*	70**	71	71*	71**	72	72*	72**	69	69*	69**	70	70*	70**	71	71*	71**	72	72*	72**	25	
73	73*	73**	74	74*	74**	75	75*	75**	76	76*	76**	73	73*	73**	74	74*	74**	75	75*	75**	76	76*	76**	26	
77	77*	77**	78	78*	78**	79	79*	79**	80	80*	80**	77	77*	77**	78	78*	78**	79	79*	79**	80	80*	80**	27	
81	81*	81**	82	82*	82**	83	83*	83**	84	84*	84**	81	81*	81**	82	82*	82**	83	83*	83**	84	84*	84**	28	
85	85*	85**	86	86*	86**	87	87*	87**	88	88*	88**	85	85*	85**	86	86*	86**	87	87*	87**	88	88*	88**	29	
89	89*	89**	90	90*	90**	91	91*	91**	92	92*	92**	89	89*	89**	90	90*	90**	91	91*	91**	92	92*	92**	30	
93	93*	93**	94	94*	94**	95	95*	95**	96	96*	96**	93	93*	93**	94	94*	94**	95	95*	95**	96	96*	96**	31	
97	97*	97**	98	98*	98**	99	99*	99**	100	100*	100**	97	97*	97**	98	98*	98**	99	99*	99**	100	100*	100**	32	
101	101*	101**	102	102*	102**	103	103*	103**	104	104*	104**	101	101*	101**	102	102*	102**	103	103*	103**	104	104*	104**	33	
105	105*	105**	106	106*	106**	107	107*	107**	108	108*	108**	105	105*	105**	106	106*	106**	107	107*	107**	108	108*	108**	34	
109	109*	109**	110	110*	110**	111	111*	111**	112	112*	112**	109	109*	109**	110	110*	110**	111	111*	111**	112	112*	112**	35	
113	113*	113**	114	114*	114**	115	115*	115**	116	116*	116**	113	113*	113**	114	114*	114**	115	115*	115**	116	116*	116**	36	
117	117*	117**	118	118*	118**	119	119*	119**	120	120*	120**	117	117*	117**	118	118*	118**	119	119*	119**	120	120*	120**	37	
121	121*	121**	122	122*	122**	123	123*	123**	124	124*	124**	121	121*	121**	122	122*	122**	123	123*	123**	124	124*	124**	38	
Ink	Ink	Ink	Ink	39																					
Ink	Ink	Ink	Ink	40																					
Ink	Ink	Ink	Ink	41																					

Figure 2.1 Spots sugar on the carbohydrate array membrane. Concentrations (star indicates lower dilution) samples 1-56 polysaccharides samples: 1 mg/mL sample, 0.2 mg/mL samples (*), and 0.04 mg/mL samples (**). Samples 57-123: oligosaccharides sample: 2 mg/mL samples, 0.4 mg/mL samples, and 0.08 mg/mL samples (**).

Table 2.7 List of polysaccharides and oligosaccharides on carbohydrate array (from Prof. W.G.T. Willats).

Name of polysaccharide	Name of oligosaccharide
1. Mannan (ivory nut)	65. β -(1 \rightarrow 4)-D-Galactobiose
2. Galactomannan (crob)	66. β -(1 \rightarrow 4)-D-Galactopentaose
3. Glucomannan (konjac)	67. 6 ¹ - α -D-Galactosyl- β -(1 \rightarrow 4)-D-mannobiose
4. Xylan (birch)	68. 6 ¹ - α -D-Galactosyl- β -(1 \rightarrow 4)-D-mannotriose
5. Arabinoxylan (wheat)	69. 6 ¹ - α -D-Galactosyl-(1-4)- β -D-Mannobiose/ Manotriose
6. Xyloglucan Tamarind seed	70. (1 \rightarrow 6 ³ ,6 ⁴)- α -D-digalactosyl- β -(1 \rightarrow 4) -D-Manopentaose
7. MLG Lichenan, β -glucan (1 \rightarrow 3),(1 \rightarrow 4)- β -D-gucan	71. D-Mannose
8. β -Glucan (Yeast), (1 \rightarrow 6),(1 \rightarrow 3)- β -D-glucan	72. (1 \rightarrow 4)- β -D-Mannobiose
9. β -Glucan (Oat), (1 \rightarrow 3),(1 \rightarrow 4)- β -D-glucan	73. (1 \rightarrow 4)- β -D-Mannotriose
10. β -Glucan (Barley flour), (1 \rightarrow 3),(1 \rightarrow 4)- β -D-glucan	74. (1 \rightarrow 4)- β -D-Mannotetraose
11. β -Glucan (<i>Euglena gracilis</i>) (1 \rightarrow 3),(1 \rightarrow 4)- β -D-glucan	75. (1 \rightarrow 4)- β -D-Mannopentase
12. Carboxymethyl cellulose (CMC 4M) (1 \rightarrow 4)- β -D-glucan w. Me-C(O)-substituents	76. (1 \rightarrow 4)- β -D-Mannoheptaose
13. Hydroxymethyl Cellulose	77. Ioprimeverose, α -D-Xylopyranosyl (1 \rightarrow 6)Glucose
14. Hydroxyethyl Cellulose	78. Xyloglucan heptamer, XXX~OH XGO7)
15. Hydroxypropyl Celulose	79. Xyloglucan namer, XLLG~OH (XGO9)
16. Ethyl cellulose	80. XG-oligosaccharide (XG14)
17. 2-Hydroxyethyl cellulose	81. (1 \rightarrow 4)- β -D-xylobiose
18. Methyl cellulose	82. (1 \rightarrow 4)- β -D-xylotriose
19. Pachyman (1 \rightarrow 3)- β -D-glucan	83. (1 \rightarrow 4)- β -D-xylotetraose
20. Pullulan (1 \rightarrow 6),(1 \rightarrow 4)- α -D-glucan	84. (1 \rightarrow 4)- β -D-xylopentaose
21. Laminarin	85. (1 \rightarrow 4)- β -D-xylohexaose
22. Arabinogalactan, type II (AGP)	86. Aldouronic acids ⁴ - α -D- glucuronosyl- β -(1 \rightarrow 4)-D-xylotetraose
23. Locust bean gum, Galactomannan rich gum	87. Glucoronoxylanoligo (XU ^{4m2} XX)
24. Gum Guar	88. Glucoronoxylanoligo (U ^{4m2} XX)
25. Gum karaya	89. Cellotriose- β -(1 \rightarrow 4)-D-glucotriose
26. Gum tragacanth	90. Cellotetraose - β -(1 \rightarrow 4)-D-glucotetraose
27. Gum Ghatti (Indian gum)	91. Cellopentaose - β -(1 \rightarrow 4)-D-glucopentaose
28. Xanthane gum Rhodigel, 80	92. Cellohexaose - β -(1 \rightarrow 4)-D-glucohexose
29. Xanthane gum Rhodigel, TSC	93. (1 \rightarrow 3),(1 \rightarrow 4)- β -D-Glucotriose (Mlg3a)
30. Gum arabic	94. (1 \rightarrow 3),(1 \rightarrow 4)- β -D-Glucotiose (Mlg3b)
31. Lime pectin DE: 81% (E81)	95. (1 \rightarrow 3),(1 \rightarrow 4)- β -D-Glucotetraose (Mlg4a)
32. Lime pectin DE: 15% (B15)	
33. Lime pectin DE: 43% (B43)	

Table 2.7 List of polysaccharides and oligosaccharides on carbohydrate array (Continued).

Name of polysaccharide	Name of oligosaccharide
34. Lime pectin DE: 64% (B64)	96. (1→3),(1→4)-β-D-Glucotetraose (Mlg4b)
35. Lime pectin DE: 71% (B71)	97. (1→3),(1→4)-β-D-Glucotetraose (Mlg4c)
36. Lime pectin DE: 11% (F11)	98. Laminaribiose (1→3)-β-D-glucobiose
37. Lime pectin DE: 31% (F31)	99. Laminaritriose (1→3)-β-D-glucotriose
38. Lime pectin DE: 58% (F58)	100. Laminaritetraose (1→3)-β-D-glucotetraose
39. Lime pectin DE: 76% (F76)	101. Laminaripentaose (1→3)-β-D-glucopeptaose
40. Lime pectin DE: 16% (P16)	102. Laminarihexaose (1→3)-β-D-glucohexaose
41. Lime pectin DE: 32% (P32)	103. Maltose (1→4)-α-D-glucobiose
42. Lime pectin DE: 46% (P46)	104. Maltotriose, α-(1→4)-D-Glucotetraose
43. Lime pectin DE: 60% (P60)	105. Maltohexaose (1→4)-α-D-Glucopeptaose
44. Lime pectin DE: 66% (P66)	106. Maltotetraose (1→6),(1→4)-α-D-glucotetraose
45. Lime pectin DE: 76% (P76)	107. Maltoheptaose, α-(1→6), (1→4)-D-Glucoheptaose
46. Sugar beet pectin with DE 62% & DA 30%	108. N-Acetyl-2-deoxy-glucos-2-amine(N-acetyl-2-deoxy-2-amino-D-glucose)
47. Sugar beet Arabinan	109. <i>Di</i> acetyl-Chitobiose
48. Linear Arabinan	110. <i>Tri</i> acetyl-Chitotriose
49. Pectic galactan, (1→4)-β-D-galactose polymer	111. <i>Tetra</i> acetyl-Chitotetraose
50. RGI (soy bean)	112. <i>Penta</i> acetyl-Chitopentaose
51. RGI (potato)	113. <i>Hexa</i> acetyl-chitopentaose
52. Lime pectin DE: 0% (E0)	114. Lactose, D-galactosyl-β-1→4-D-glucose
53. Lemon pectin	115. D-Glucose
54. Apple pectin	116. 4 ² ,6 ² -D- <i>di</i> Galactosyl-(1→4)-β-D-Galactobiose
55. CP Kelco Pectin	117. 6 ² -β-D-Galactosyl-(1→4)-β-D-Galactotriose
56. Sigma esterified citrus pectin	118. 6 ² -α-D-Galactosyl-(1→4)-β-D-Galactotriose
57. α-(1→5)-L-Arabinobiose	119. α-(1→5)-L-Arabinobiose, feruloylated
58. α-(1→5)-L-Arabinotriose	120. α-(1→5)-L-Arabinotriose, feruloylated
59. α-(1→5)-L-Arabinotetraose	121. β-(1→4)-D-Galactobiose, feruloylated
60. α-(1→5)-L-Arabinopentaose	122. RGI backbone (chem. Synthesised Rha-GalA-Rha-GalA-Rha-GalA)
61. α-(1→5)-L-Arabinohexaose	123. BSA
62. α-(1→5)-L-Arabinohexaose	
63. α-(1→5)-L-Arabinooctaose	
64. D-Galactose	

2.12 Detecting binding of OsBGal1 Cter with sugar by Saturation Transfer Difference (STD) method in NMR.

STD NMR is applied to investigate ligand binding to protein receptors (Canales et al., 2008) including antigen-antibody recognition events (Oberli et al., 2010). After the last step of purification, the OsBGal1 Cter buffer was changed to 10 mM sodium phosphate buffer, pH 7.5, that was made with 99.9% D₂O (Sigma). Moreover, the sugar ligands were dissolved in the same buffer. The OsBGal1 Cter (30 μM) protein was mixed with a 50-fold molar excess (1.5 mM) of 1,5-α-L-arabinotriose in an NMR tube and the proton spectra determined in Bruker AVANCE DRX 500 spectrometer equipped with a Bruker Prodigy cryoprobe. The protein-saturating irradiation was varied at 7.25 ppm (3629 Hz), 7.04 ppm (3524 Hz), 6.67 ppm (3436 Hz), 6.27 ppm (3139 Hz), 3.05 ppm (1528 Hz), 2.60 ppm (1300 Hz), 1.47 ppm (1011 Hz), (735 Hz) and 0.74 ppm (372 Hz), respectively. The difference STD spectra were generated by subtracting the spectra in which the protein was selectively saturated by irradiation at a region of the spectrum that contains only resonance of protein (at the frequencies given above) from the spectrum of the sugar-protein mixture without protein saturation.

CHAPTER III

RESULTS

3.1 Recombinant expression of rice β -galactosidase C-terminal domain (OsBGal1 Cter)

3.1.1 Preparation of expression vector

The full-length cDNA sequence of OsBGal1 consists of 2565 nucleotides, including a 2529 nucleotide open reading frame that encodes a precursor protein of 843 amino acid residues (Chantarangsee et al., 2007), as shown in Figure 3.1. The cDNA encoding the OsBGal1 C-terminal domain (OsBGal1 Cter), including the stop codon, was amplified by PCR (Figure 3.2), then cloned into the pENTR/D-TOPO® vector (Figure 3.3). The cDNA encoding OsBGal1 Cter was subcloned into the pET32a/DEST (Figure 3.4) and digestion of the recombinant plasmid with *NcoI* and *EcoRI* restriction endonucleases showed the expected size band of 420 bp (Figure 3.5).

```

OsBGal1  ATTCACTAGTGATTTGAGGAGAATGGGTGAGTGATGGGGAGGGGGTGCCTGGCTGCGCTG  60
                                                M G R G C L A A L
OsBGal1  CTCGGCGGCGCGGTGGCGGTGGCCGTGCTGGTTCGCCGTGCTCCACTGCGCGGTGACGTAC  120
L G G A V A V A V L V A V V H C A V T Y
OsBGal1  GACAAGAAGGCGGTGCTCGTCGACGGCCAGAGGAGGATTCTTCTCCGGATCCATACAT  180
D K K A V L V D G Q R R I L F S G S I H
OsBGal1  TACCCGAGGAGCACACCCGAAATGTGGGACGGGCTAATTGAGAAGGCTAAAGATGGAGGC  240
Y P R S T P E M W D G L I E K A K D G G
OsBGal1  TTGGATGTGATCCAGACCTATGTCTTTTGAATGGGCATGAACCACTCCGGAAATTAC  300
L D V I Q T Y V F W N G H E P T P G N Y
OsBGal1  AATTTTGAAGGGAGGTACGATCTGGTCAGGTTTCATCAAGACTGTCCAGAAGGCTGGCATG  360
N F E G R Y D L V R F I K T V Q K A G M
OsBGal1  TTTGTTTCATCTCCGCATCGGTCCCTACATTTGTGGAGAGTGAATTTTGGGGGATTTCCA  420
F V H L R I G P Y I C G E W N F G G F P
OsBGal1  GTTTGGTTGAAGTATGTACCAGGCATCAGTTTCAGGACGGACAATGAACCTTTCAAGAAT  480
V W L K Y V P G I S F R T D N E P F K N
OsBGal1  GCAATGCAGGGGTTACAGAGAAAATTTGGGCATGATGAAGAGTGAACCTCTTTGCT  540
A M Q G F T E K I V G M M K S E N L F A
OsBGal1  TCACAAGGCGGTCTATTATCCTCTCTCAGATTGAGAACGAGTATGGGCCAGAAGGTA  600
S Q G G P I I L S Q I E N E Y G P E G K
OsBGal1  GAGTTTGGGGCTGCCGGCAAGGCATATATCAACTGGGCGGCAAAGATGGCCGTGGGATTG  660
E F G A A G K A Y I N W A A K M A V G L
OsBGal1  GACACCGGTGTGCCGTGGGTGATGTGCAAGGAGGATGACGCACCTGACCCAGTGATCAAT  720
D T G V P W V M C K E D D A P D P V I N
OsBGal1  GCATGCAATGGTTTCTATTGTGACACATTTTCTCCTAACAAGCCTTACAAGCCTACGATG  780
A C N G F Y C D T F S P N K P Y K P T M
OsBGal1  TGGACTGAAGCTTGGAGTGGATGGTTTACTGAATTCGGAGGAACCATCCGTCAACGACCA  840
W T E A W S G W F T E F G G T I R Q R P
OsBGal1  GTTGAAGATCTCGCATTTGGTGTGCTCGCTTCGTACAGAAGGGTGGTTCTTTTATCAAC  900
V E D L A F G V A R F V Q K G G S F I N
OsBGal1  TACTACATGTATCATGGAGGAACGAATTTTGGTTCGCACGGCTGGAGGTCCCTTTATCACC  960
Y Y M Y H G G T N F G R T A G G P F I T
OsBGal1  ACGAGCTATGATTATGATGCTCCACTTGATGAATATGGTCTTGCAAGGGAACCAAAGTTT  1020
T S Y D Y D A P L D E Y G L A R E P K F
OsBGal1  GGGCACCTTAAAGAADCCATAGAGCTTAAAGTTATGTGAGCAGCCTTTGGTTTCTGFC  1080
G H L K E L H R A V K L C E Q P L V S A
OsBGal1  GATCCAACCTGTGACTACCCCTTGAAGTATGCAAGAGGCCCATGTGTTCCGATCTTCTCT  1140
D P T V T T L G S M Q E A H V F R S S S
OsBGal1  GGCTGTGAGCTTTCCTTGCAAACCTACAATCTAACTCGTATGCCAAAGTTATCTTCAAC  1200
G C A A F L A N Y N S N S Y A K V I F N
OsBGal1  AATGAAAATTACAGCCTTCCACCTTGGTCAATCAGCATCCTTCCTGATTGCAAAAATGTT  1260
N E N Y S L P P W S I S I L P D C K N V

```

Figure 3.1 The full-length cDNA sequence and deduced amino acid sequence of rice OsBGal1. An arrow shows the putative signal peptide cleavage site predicted by the SignalP program (Nielsen et al., 1997). The C-terminal galactose-binding-lectin-like domain or OsBGal1 Cter is shown with a grey background. Underlined nucleotide letters indicate the positions of PCR primers used for cloning of the cDNA encoding OsBGal1 Cter.

```

OsBGal1  GTTTTTAACACTGCAACAGTTGGTGTTTCAGACAAATCAAATGCAAATGTGGGCAGACGGG 1320
          V F N T A T V G V Q T N Q M Q M W A D G
OsBGal1  GCTTCTCAATGATGTGGGAGAAGTATGATGAGGAGGTTGATTCATTGGCAGCTGCTCCA 1380
          A S S M M W E K Y D E E V D S L A A A P
OsBGal1  TTGCTCACGTCAACTGGTCTACTTGAGCAGCTTAATGTCACAAGAGACACCAGTGATTAC 1440
          L L T S T G L L E Q L N V T R D T S D Y
OsBGal1  CTCTGGTACATTACGAGGGTGGAGGTAGACCCATCTGAGAAGTTTCTACAAGGTGGCAGC 1500
          L W Y I T R V E V D P S E K F L Q G G T
OsBGal1  CCTCTGTCACCTCACTGTGCAGTCTGCTGGCCATGCGCTGCATGTCTTCATCAATGGGCAA 1560
          P L S L T V Q S A G H A L H V F I N G Q
OsBGal1  CTCCAAGTTCTGCCTATGGAACCAGGGAAGATCGGAAAATCTCATATAGTGGCAATGCT 1620
          L Q G S A Y G T R E D R K I S Y S G N A
OsBGal1  AACCTTCGTGCTGGTACAAACAAAGTTGCACTGTTGAGTGTGCTTGTGGACTGCCGAAT 1680
          N L R A G T N K V A L L S V A C G L P N
OsBGal1  GTCGGAGTGCATTATGAGACGTGGAACACTGGTGTGTTGGTCTGTTGTGATTACGCGG 1740
          V G V H Y E T W N T G V V G P V V I H G
OsBGal1  TTGGACGAAGGTTCAAGAGATCTGACTTGGCAGACTTGGTCTATCAGTTCAGGTTGGC 1800
          L D E G S R D L T W Q T W S Y Q F Q V G
OsBGal1  CTGAAAGGTGAACAGATGAATCTAAACTCCTTAGAAGGCTCAGGCTCAGTTGAATGGATG 1860
          L K G E Q M N L N S L E G S G S V E W M
OsBGal1  CAAGGATCATTGGTAGCACAAAACCAACACCGTTGGCATGGTATAGGGCATACTTTGAT 1920
          Q G S L V A Q N Q Q P L A W Y R A Y F D
OsBGal1  ACTCCCAGTGGTGCAGGAGCCACTGGCTCTGGATATGGGCAGCATGGGTAAGGTCAAATA 1980
          T P S G D E P L A L D M G S M G K G Q I
OsBGal1  TGGATAAATGGGCAAAGCATTGGACGGTACTGGACAGCATATGCGGAAGGTGACTGCAAA 2040
          W I N G Q S I G R Y W T A Y A E G D C K
OsBGal1  GGTGGCCATTACACTGGGTATACAGGGCACCCAAGTGTGAGGCAGGTTGTGGTTCAGCCT 2100
          G C H Y T G S Y R A P K C Q A G C G Q P
OsBGal1  ACACAGCCTGGTATCATGTGCCAAGATCTGGTTGCAACCAACTAGAAAATCTGCTAGTG 2160
          T Q R W Y H V P R S W L Q P T R N L L V
OsBGal1  GTTTTTGAGGAAGTGGCGGTGATTCTTCAAAGATTGCCCTTGCGAAGCGGACAGTCTCA 2220
          V F E E L G G D S S K I A L A K R T V S
OsBGal1  GGTGCTGTGCTGATGTATCTGAATATCATCCAAATATCAAGAAGTGGCAGATCGAGAGC 2280
          G V C A D V S E Y H P N I K N W Q I E S
OsBGal1  TATGGGAACAGAGTTCACACGGCAAAGGTGCATTTAAAATGTGCACCTGGGCAGACC 2340
          Y G E P E F H T A K V H L K C A P G Q T
OsBGal1  ATTTCTGCAATCAAATTTGCTAGCTTTGGGACACCTCTTGGAACTTGCAGAACATTCCAG 2400
          I S A I K F A S F G T P L G T C G T F Q
OsBGal1  CAAGGGGAGTGCCATTCAATTAACCTCAAACTCTGTTCTTAAAAGGAAATGCATTGGACTA 2460
          Q G E C H S I N S N S V L E R K C I G L
OsBGal1  GAAAGATGTGTCGTCGCAATCTCTCCAGCAACTTTGGTGGAGATCCCTGCCCGGAGGTG 2520
          E R C V V A I S P S N F G G D P C P E V
OsBGal1  ATGAAAAGGGTGGCGGTTGAGGCGGTATGCTCTACCGCTGCATAGGGCAATTTCTTTGCC 2580
          M K R V A V E A V C S T A A *
OsBGal1  GAGACGAATTTAGAGATGGCACCAAACAGGTGGGCTTGTGCTTGA 2625

```

Figure 3.1 (Continued) The full-length cDNA sequence and deduced amino acid sequence of rice OsBGal1. An arrow shows the putative signal peptide cleavage site predicted by the SignalP program (Nielsen et al., 1997). The C-terminal galactose-binding-lectin-like domain or OsBGal1 Cter is shown with a grey background. Underlined nucleotide letters indicate the positions of PCR primers used for cloning of the cDNA encoding OsBGal1 Cter.

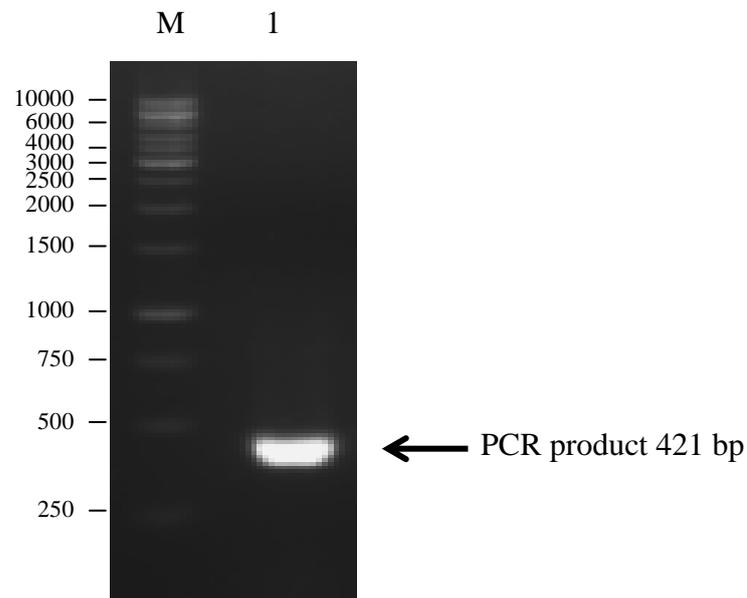


Figure 3.2 Agarose gel electrophoresis of the cDNA fragment encoding OsBGal1 Cter that was amplified with OsBal1 Cter_For primer and OsBgal1 Cter_Rev primer. The PCR product was amplified with panicle rice cDNA as a template. Lane M, 1 kbp DNA marker (SibEnzyme); lane 1, the OsBgal1 Cter PCR product.

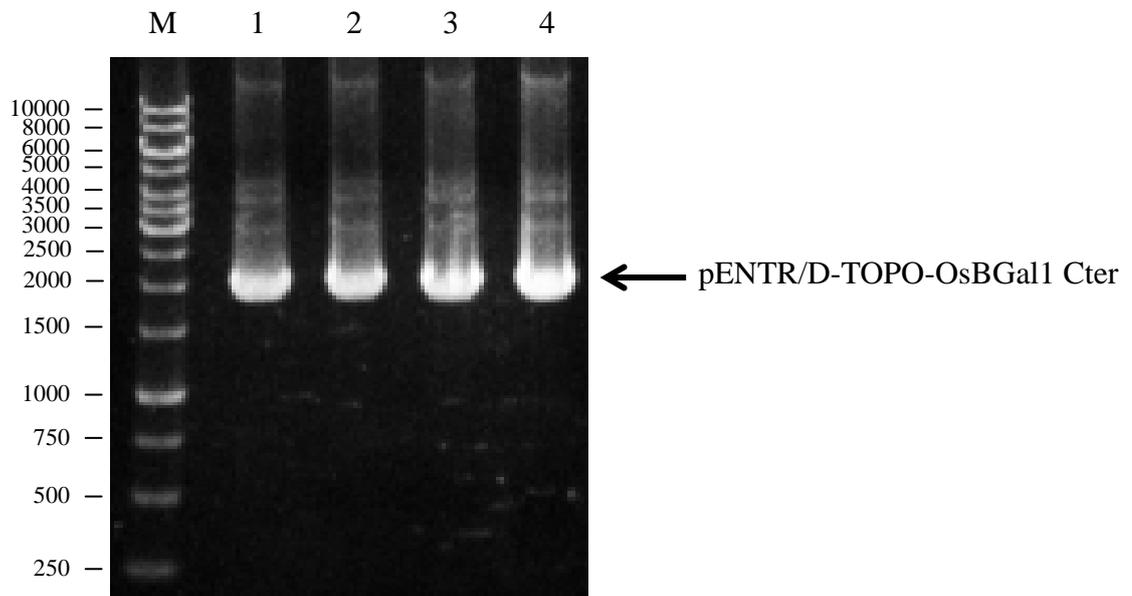


Figure 3.3 Agarose gel electrophoresis of the pENTR/D-TOPO-OsBGal1_Cter plasmid. Lane M, 1 kbp DNA marker (Invitrogen); lanes 1-4, the pENTR/D-TOPO-OsBGal1_Cter plasmid clones 1-4, respectively.

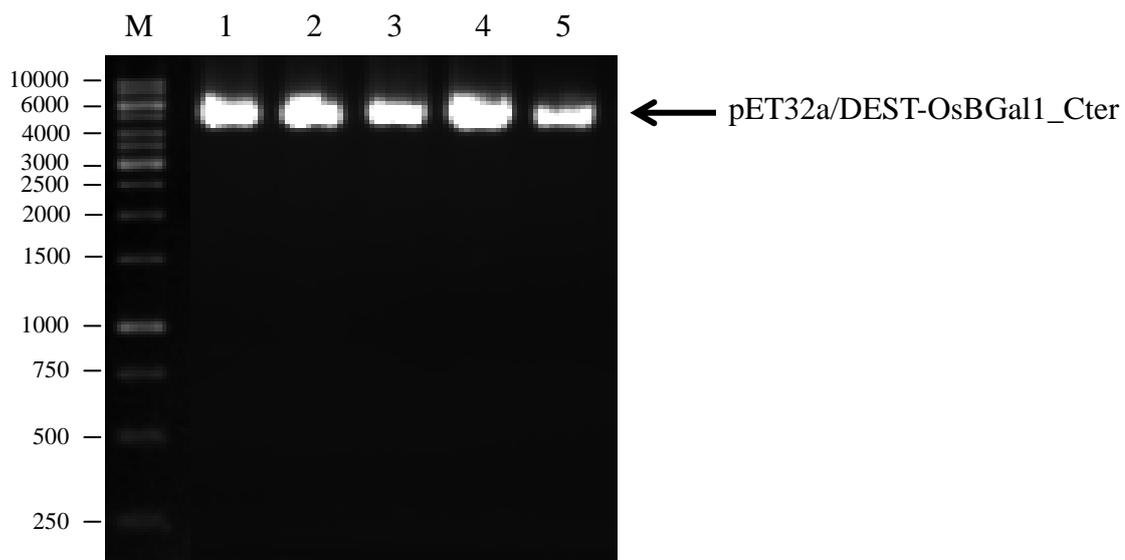


Figure 3.4 Agarose gel electrophoresis of the pET32a/DEST-OsBGal1_Cter plasmid.

Lane M, 1 kbp DNA marker (Invitrogen); lanes 1-5, the pET32a/DEST-OsBGal1_Cter plasmid clones 1-5, respectively.

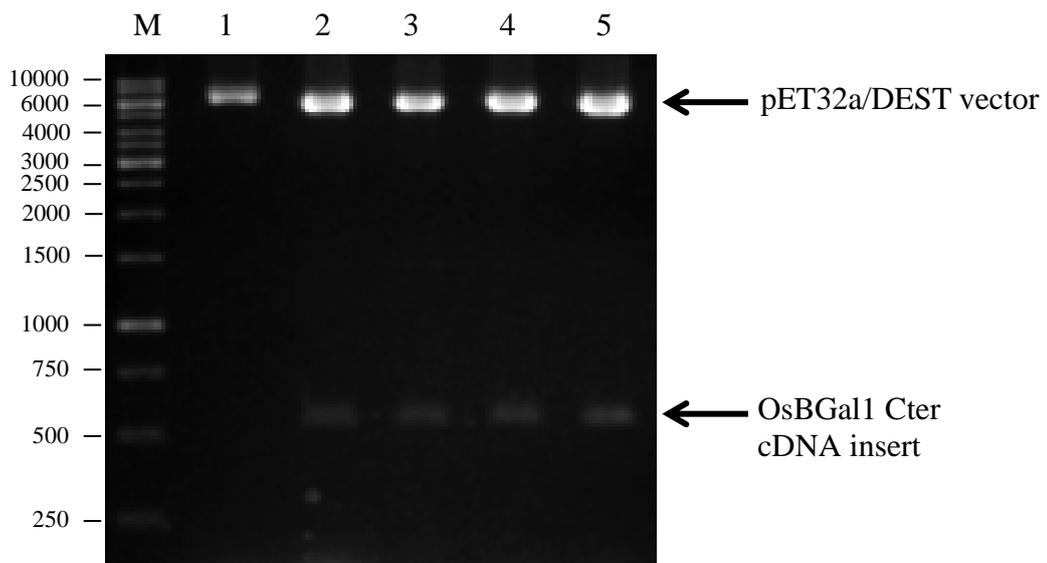


Figure 3.5 Agarose gel electrophoresis of pET32a/DEST_OsBGal1 Cter plasmids after digestion with *NcoI* and *EcoRI* restriction enzymes. Lane M, 1 kbp DNA marker (Invitrogen); lanes 1-5, *NcoI* and *EcoRI* double digests of pET32a/DEST_OsBGal1 Cter plasmid clones 1-5, respectively.

3.1.2 Recombinant expression of rice OsBGal1 Cter protein

To achieve soluble protein expression, the *E.coli* strains Origami(DE3) and Origami B(DE3) were used in this study. The pET32a/DEST-OsBGal1_Cter (Figure 3.6) expressed in Origami(DE3) and Origami B(DE3) produced the N-terminally thioredoxin-tagged fusion protein of the OsBGal1 Cter protein, which was detected at 31 kDa on SDS-PAGE (Figure 3.7). However, almost all of the expressed protein was localized to inclusion bodies. More soluble protein was obtained from the construct in Origami B(DE3) than in Origami(DE3) *E. coli*.

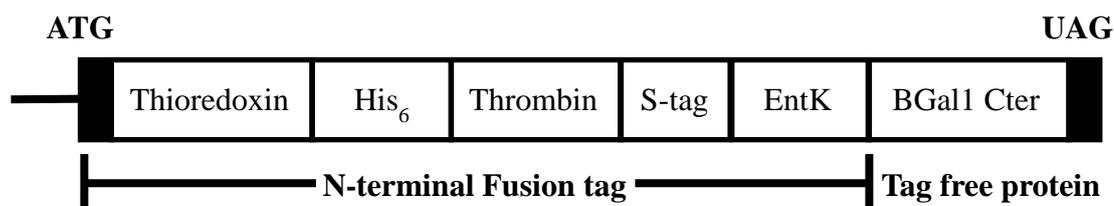


Figure 3.6 Map of the protein sequence encoded by the recombinant pET32a/DEST with the *OsBGal1Cter* cDNA inserted after the enterokinase cleavage site.

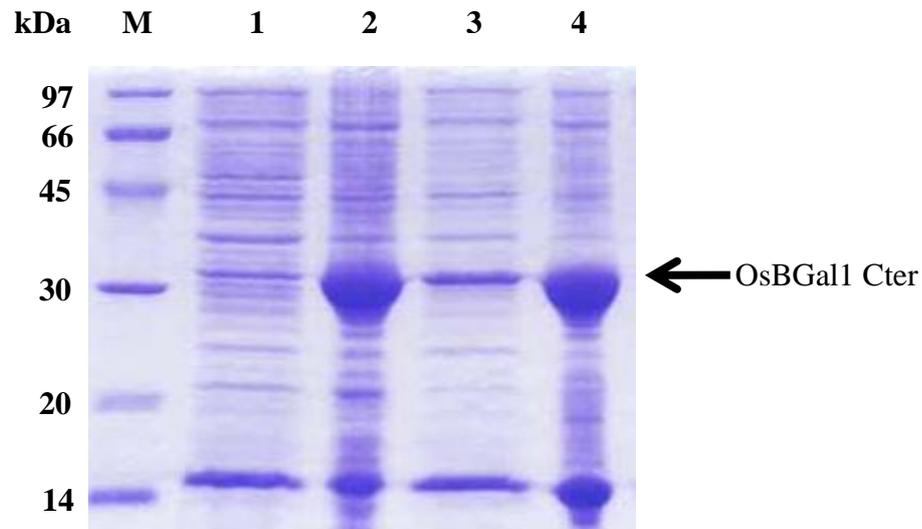


Figure 3.7 SDS-PAGE of pET32a/DEST-OsBGal1 Cter protein expressed in *E. coli* strains Origami(DE3) and Origami B(DE3). Lane M, Bio-Rad low molecular weight markers; lane 1, the soluble protein of OsBGal1 Cter expressed in Origami(DE3) after incubation with 0.4 mM IPTG, at 20°C for 18 h; lane 2, the inclusion bodies of OsBGal1 Cter expressed in Origami(DE3) after incubation with 0.4 mM IPTG, at 20°C for 18 h; lane 3, the soluble protein of OsBGal1 Cter expressed in Origami B(DE3) after induction with 0.4 mM IPTG, at 20°C for 18 h; lane 4, the inclusion bodies of OsBGal1 Cter expressed in Origami B(DE3) after incubation with 0.4 mM IPTG, at 20°C for 18 h. The arrow points to the position of OsBGal1 Cter fusion protein at 31 kDa.

3.1.3. Purification of OsBGal1 Cter protein.

The thioredoxin-His tag OsBGal1 Cter fusion protein expressed in Origami B(DE3) was purified by Co^{2+} agarose immobilized metal affinity chromatography (IMAC, Figure 3.8). The OsBGal1 Cter protein, which was detected in major band at 31 kDa and had small amounts of contaminate proteins above and below the major band on SDS-PAGE.

For further purification and to simplify protein crystallization and NMR studies, it was desirable to remove the N-terminal thioredoxin and His₆ tags from the fusion protein. To test digestion, the fusion protein was digested with thrombin protease or enterokinase. The result show that the recombinant OsBGal1 Cter fusion protein could be completely digested with enterokinase and thrombin proteases. Even though, enterokinase could cut the thioredoxin, His₆ and S-tags from the OsBGal1Cter protein, this digest appeared to give protein digested at more than one site (Figure 3.9). Therefore, the OsBGal1Cter fusion protein was digested with thrombin protease. For separation of the N-terminal fusion protein containing the thioredoxin and His tags from the OsBGal1 Cter protein, a second step of IMAC was used to adsorb this fusion tag protein. After the protein that passed through the column was concentrated and the buffer changed, the purified OsBGal1 Cter had approximately 80% purity (Figure 3.10).

The free OsBGal1 Cter was further purified by S75 gel filtration. The protein fraction passed through gel filtration as a single peak on the chromatogram (in fractions 50 to 57) as shown in Figure 3.11. The SDS-PAGE in Figure 3.11 shows that free protein was eluted concurrently with contaminating

protein. This suggested that we could not use gel filtration to separate contaminate protein from free OsBGal1 Cter.

Nonetheless, the purified protein was submitted to crystallization trials. No crystals were observed in microbatch trials of protein mixed with the precipitants in the Crystal Screen High Throughput HR2-130 (Hampton Research, CA, USA), WizardI&II (Emerrald BioSystems, Inc.) and JR screening kits. Therefore a new construct of OsBGal1 Cter was made to remove the long N-terminal linker and try to improve the purity of the protein.

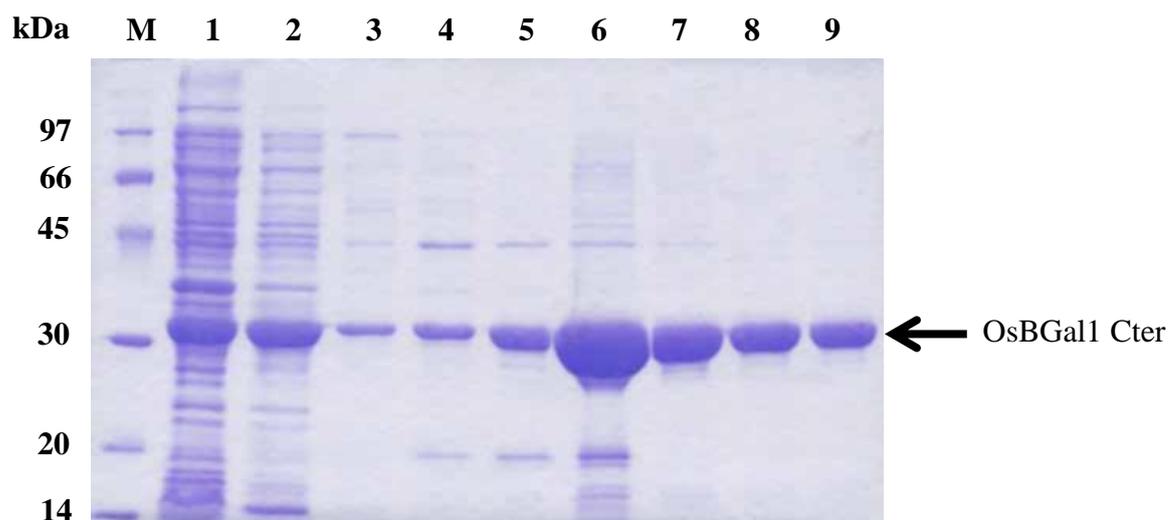


Figure 3.8 SDS-PAGE of OsBGal1 Cter purified by immobilized Co^{2+} affinity chromatography (IMAC). Lane M, Bio-Rad low molecular weight markers; lane 1, crude protein extract of induced Origami B(DE3) cells; lane 2, flow-through fraction of proteins that passed through the Co^{2+} column; lane 3, wash 0 fraction; lane 4, wash 1 fraction; lane 5, wash 2 fraction; lanes 6-9, purified OsBGal1 Cter in the IMAC elution fractions No. 1-4, respectively.

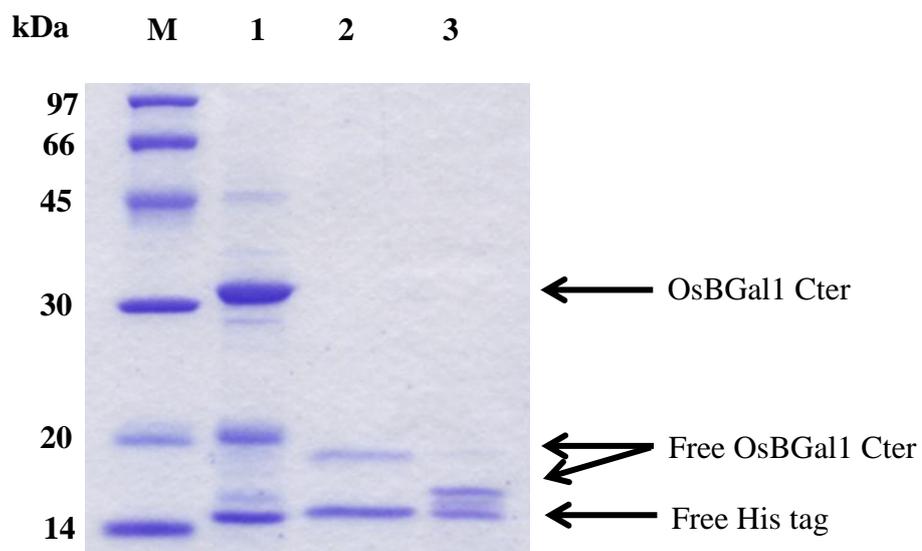


Figure 3.9 SDS-PAGE of OsBGal1 Cter after digestion with thrombin or enterokinase protease. Lane M, Bio-Rad low molecular weight markers; lane 1, protein after the first IMAC; lane 2, fusion protein digested with thrombin protease; lane 3, fusion protein digested with enterokinase.

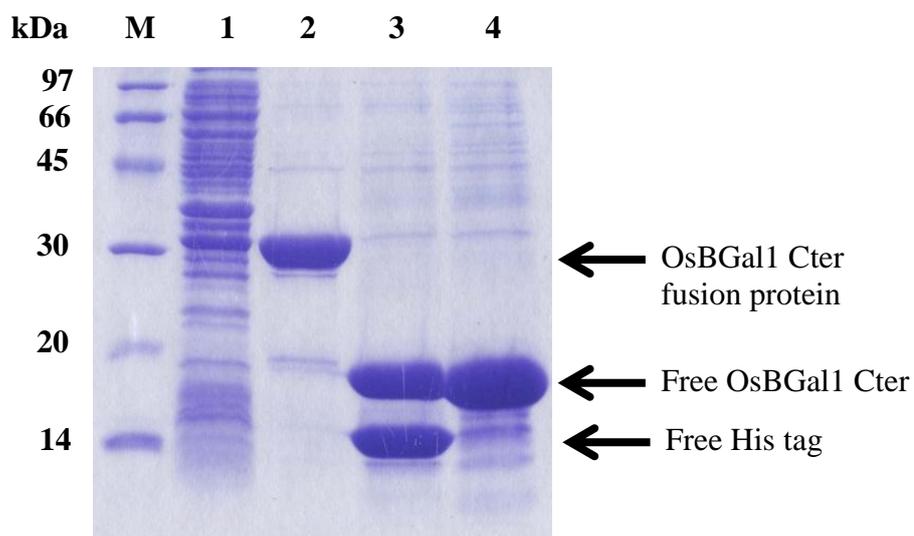


Figure 3.10 SDS-PAGE of OsBGal1 Cter from the first expression construct after purification by 1st and 2nd IMAC steps. Lane M, Bio-Rad low molecular weight markers; lane 1, crude protein extract of induced Origami B(DE3) cells; lane 2, OsBGal1 Cter protein purified by IMAC; lane 3, protein after digestion with thrombin protease; lane 4, protein after purification by 2nd IMAC.

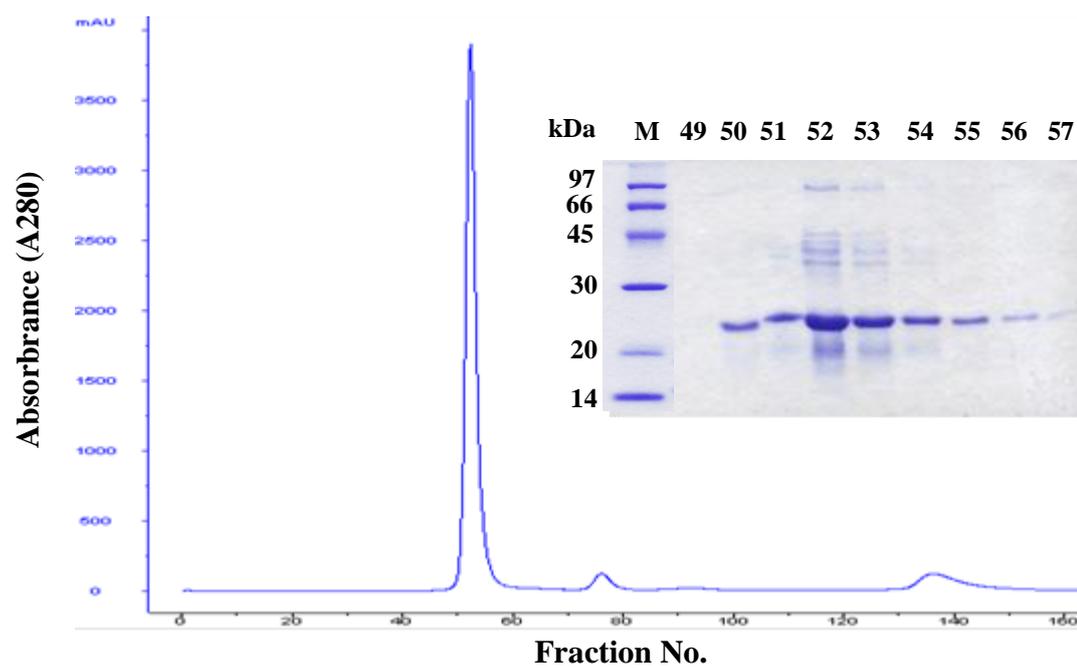


Figure 3.11 Elution profile of the free OsBGal1 1 Cter protein passed through an S75 gel filtration column and SDS-PAGE of fractions at the protein peak.

3.2 Expression of OsBGal1 from a new construct with a convenient thrombin cleavage site.

3.2.1 Recombinant plasmid production.

In order to make a new construct with the thrombin cleavage site closer to the OsBGal1 Cter protein, the cDNA was amplified by PCR with the OsBGal1 Cter_ *NdeI*_For and OsBGal1 Cter_ *BamHI*_Rev as primers and the pET32a/DEST_OsBGal1 Cter as template. The PCR product of OsBGal1 Cter (Figure 3.12) was cloned into pET15b(+) and subcloned into pET32b(+) (Figure 3.13). The new construct encoded two His tags before the OsBGal1 Cter encoding region and a thrombin cleavage site just before this region, as shown in Figure 3.14.

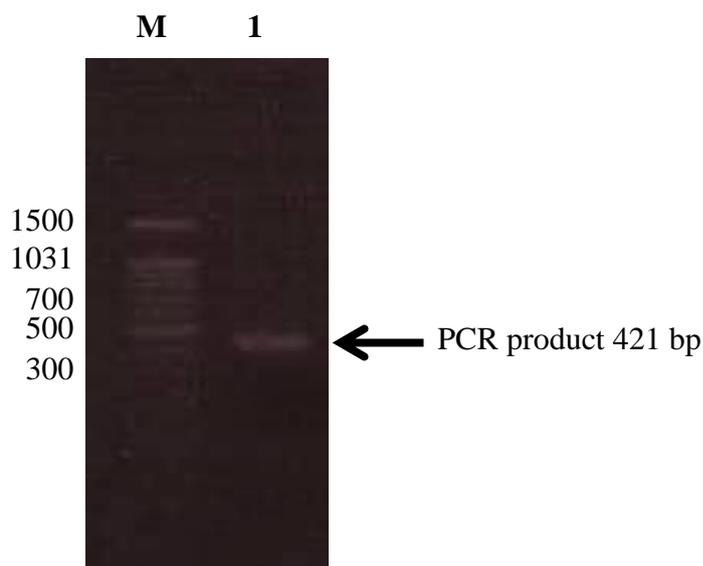


Figure 3.12 Agarose gel electrophoresis of OsBGal1 Cter cDNA fragment amplified with OsBGal1_Cter_ *NdeI*_For and OsBGal1_Cter_ *BamHI*_Rev primers. Lane M, 1 kbp DNA marker (Fermentas); lane 1, the OsBGal1 Cter PCR product.

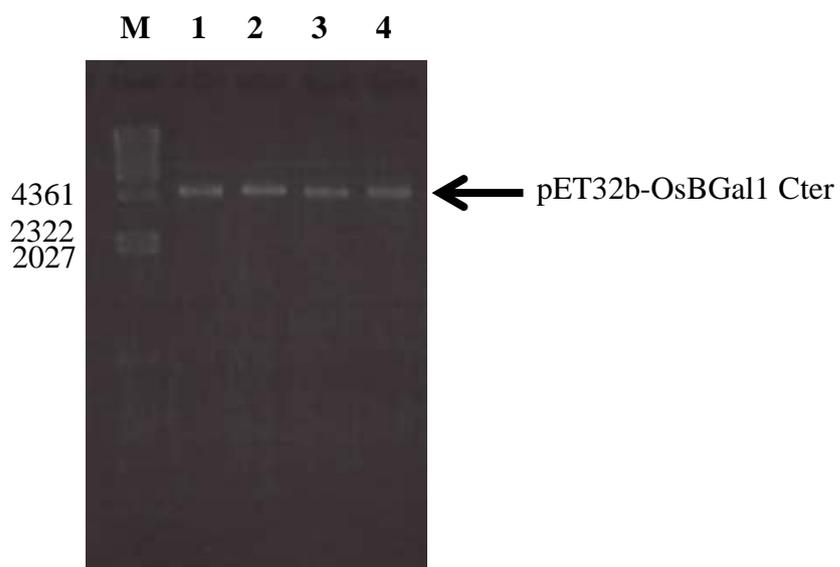


Figure 3.13 Agarose gel electrophoresis of pET32b-OsBGal1_Cter_jp_new plasmid. Lane M, 1 kbp DNA marker (Fermentas); lanes 1-4, pET32b-OsBGal1 Cter plasmid_jp_new clones 1-4, respectively.

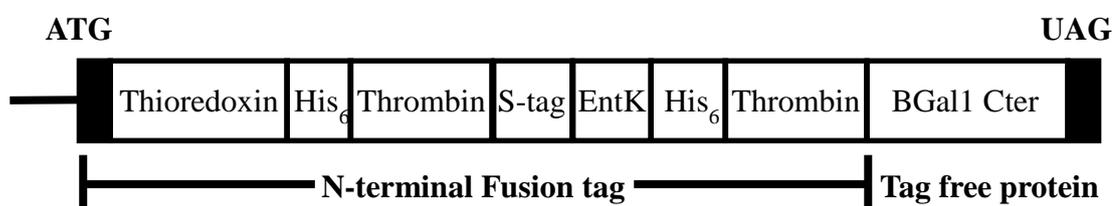


Figure 3.14 Map of the protein sequence encoded by the recombinant pET32b-OsBGal1 Cter_jp_new plasmid with the cDNA encoding OsBGal1 Cter inserted after a thrombin cleavage site.

3.2.2. Expression and purification of the new OsBGal1 Cter fusion protein

The new construct was expressed in *E.coli* strain Origami B(DE3) as before. The thioredoxin OsBGal1 Cter fusion was purified by Co^{2+} agarose affinity chromatography (Figure 3.15). The molecular weight of the recombinant OsBGal1 Cter fusion protein was about 33 kDa, as shown by SDS-PAGE and immunoblot analysis with an anti-polyhistidine antibody (Figure 3.16). A time course of thrombin protease digestion was done to determine the optimal time of digestion. The fusion protein was digested with thrombin protease at 4°C and the time varied 0-24 h. Figure 3.14 shows that the OsBGal1 Cter was completely cleaved from the fusion protein with thrombin protease within 2 h. The free OsBGal1 Cter protein was purified by a 2nd IMAC column, onto which the thioredoxin-His-tag was adsorbed. The purified OsBGal1 Cter protein, which eluted from the IMAC column in 10 mM imidazole in equilibration buffer, had an apparent mass of 13 kDa on SDS-PAGE, while thioredoxin and His₆ could be eluted with 250 mM imidazole in equilibration buffer (Figure 3.17). Figure 3.18 shows the protein after each step for purification.

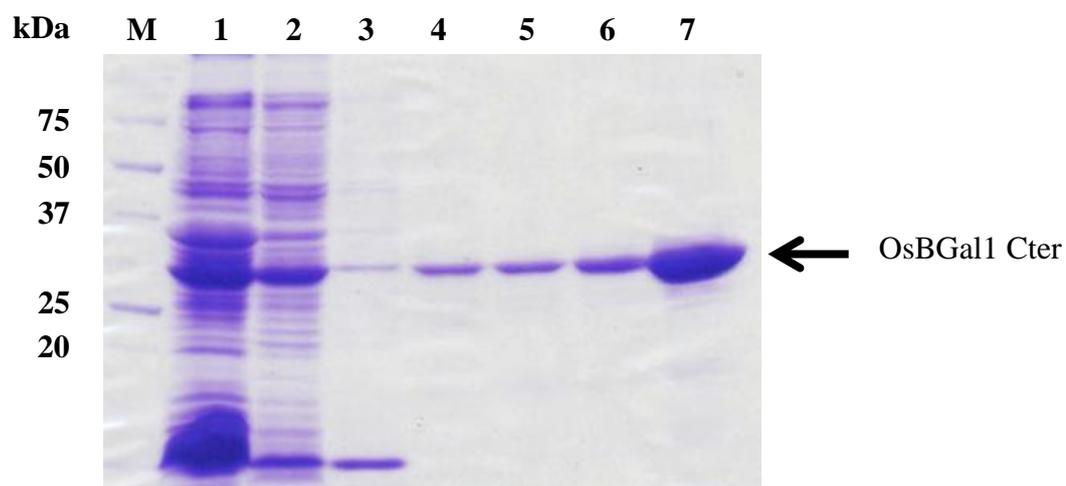


Figure 3.15 SDS-PAGE of the OsBGal1 Cter_{jp}_new fusion protein purified by IMAC. Lane M, Bio-Rad low molecular weight markers; lane 1, crude protein extract of induced Origami B(DE3) cells; lane 2, flow-through fraction of proteins that passed through the Co^{2+} column; lane 3, wash 0 fraction; lane 4, wash 1 fraction; lane 5, wash 2 fraction; lanes 6-7, purified OsBGal1 Cter in the IMAC elution fractions 1 and 2, respectively

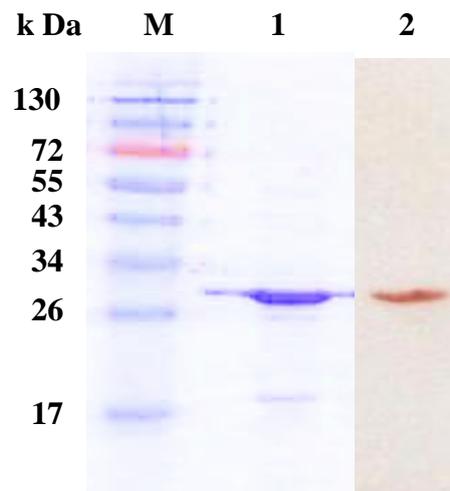


Figure 3.16 The OsBGal1 Cter protein detected by SDS-PAGE and western immunoblot analysis with a mouse monoclonal anti-poly-Histidine IgG2a isotype antibody and peroxidase conjugated goat anti-mouse IgG. Lane M; prestained protein marker, lane 1; OsBGal1 Cter protein detected by Coomassie brilliant blue staining of SDS-PAGE gel, lane 2, OsBGal1 Cter protein detected by western immunoblot analysis.

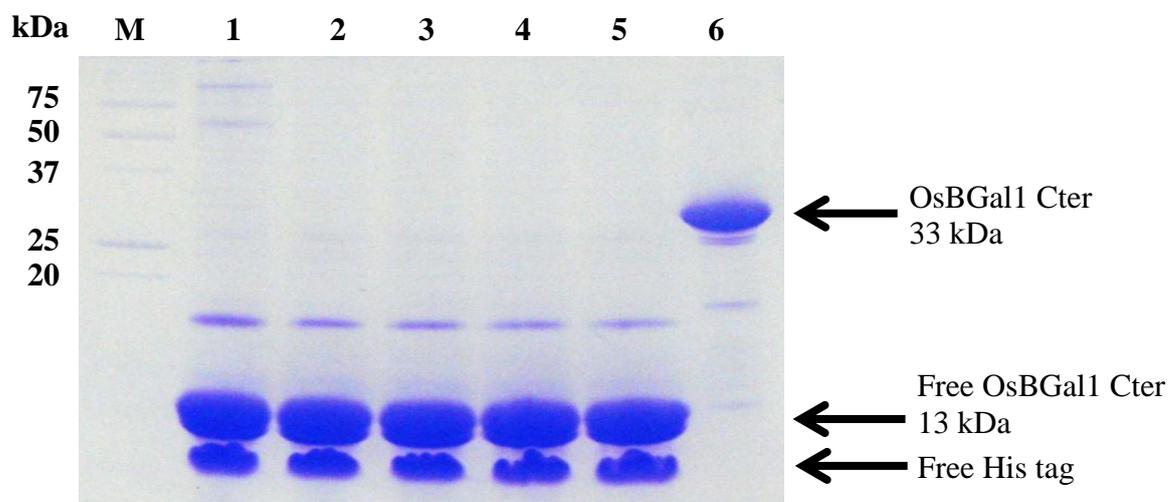


Figure 3.17 Time course of thrombin protease digestion of thioredoxin-OsBGal1 Cter fusion protein analyzed by SDS-PAGE. Lane M, LMW marker; lane 1, thrombin protease digest for 2 h; lane 2, thrombin protease digest for 4 h; lane 3, thrombin protease digest for 6 h; lane 4, thrombin protease digest for 8 h; lane 5, thrombin protease digest for 24 h; lane 6, undigested protein.

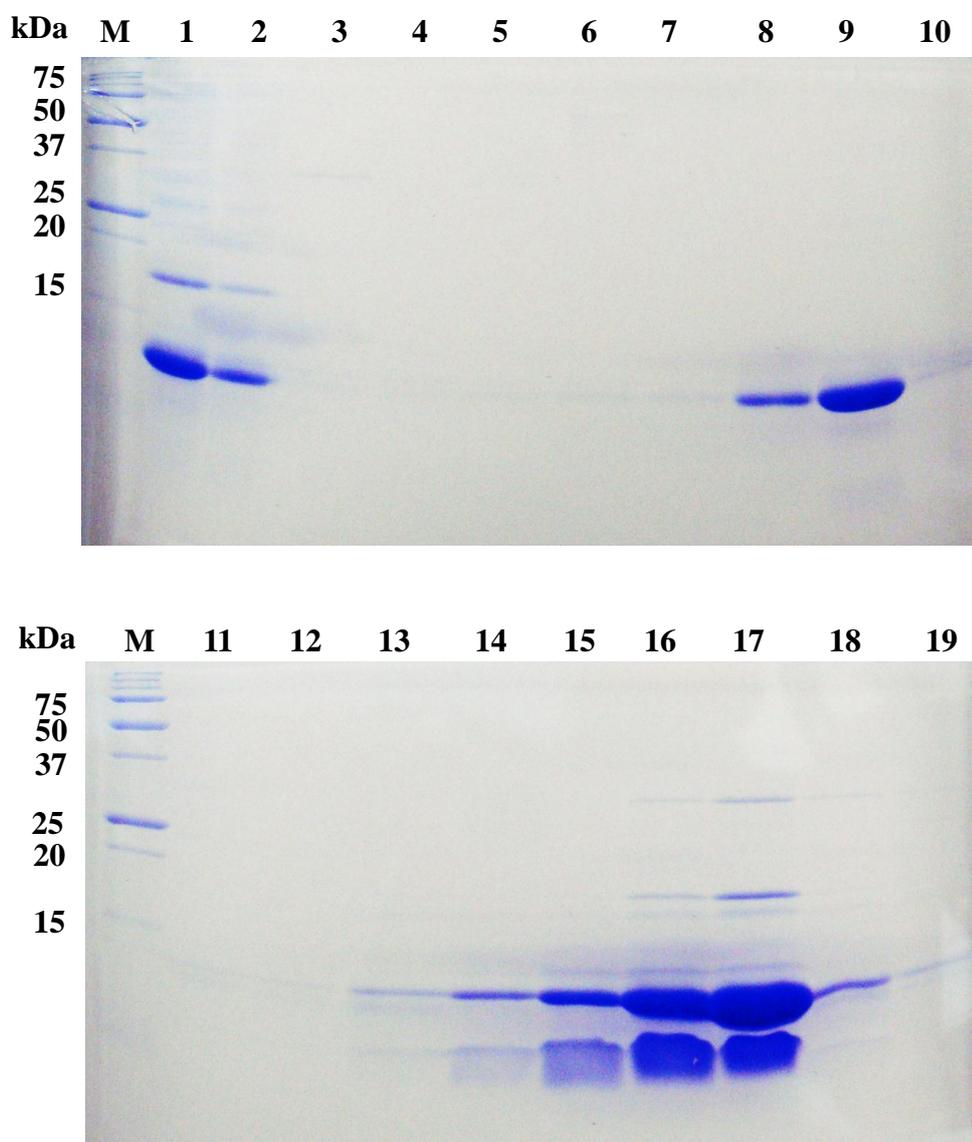


Figure 3.18 17% SDS-PAGE analysis of free OsBGal1 Cter purified by 2nd immobilized Co²⁺ affinity chromatography (IMAC) column. Lane M, Bio-Rad low molecular weight markers; lane 1, fraction of proteins that flowed through the Co²⁺ column; lanes 2-6, wash with equilibration buffer; lanes 7-10, wash with equilibration buffer containing 10 mM imidazole; lanes 11-14, wash with equilibration buffer containing 20 mM imidazole; lanes 15-19, His-tag protein eluted with 250 mM imidazole.

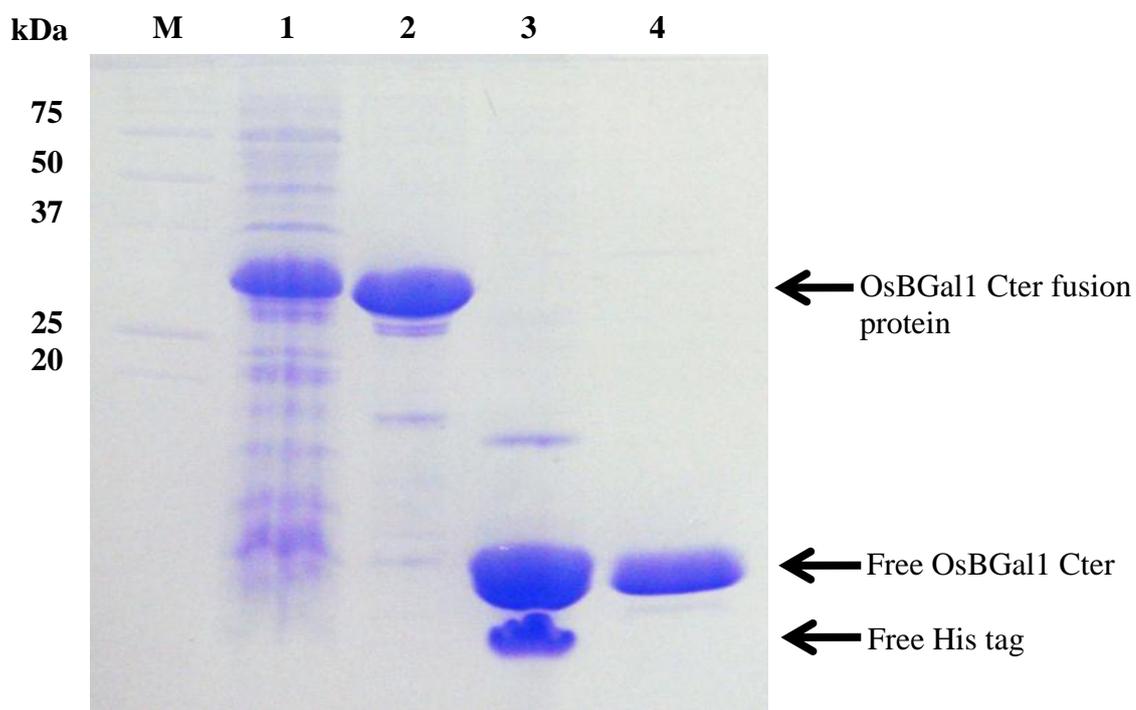
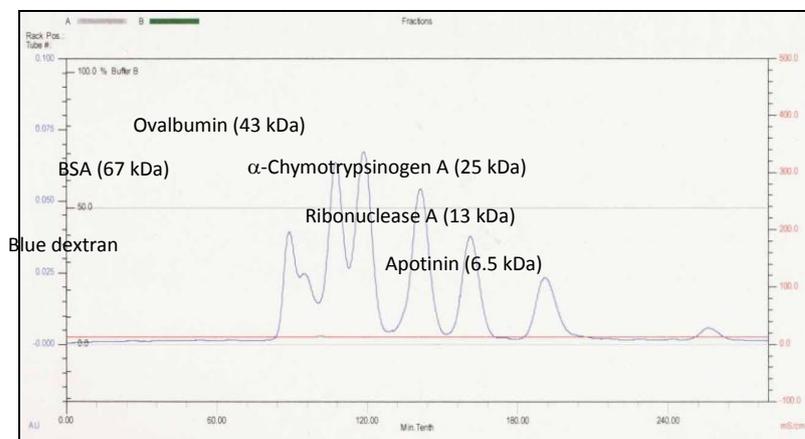


Figure 3.19 SDS PAGE analysis of OsBGal1 Cter fractions from the steps of purification for structure determination by NMR. Lane M, Bio-Rad low molecular weight markers; lane 1, soluble protein extract of Origami B(DE3) cells; lane 2, fusion protein after initial IMAC; lane 3, thrombin protease digest; lane 4, OsBGal1 Cter after 2nd IMAC.

3.3 Determination of native molecular weight of OsBGal1 Cter by gel filtration on Superdex S75 column

The native molecular weight of the free OsBGal1 Cter protein was determined by gel filtration on a Superdex S75 column chromatography. The molecular weight of the native form of the OsBGal1 Cter protein was estimated by of the K_{av} value with the standard curve of those derived from the elution volumes of proteins of known molecular mass. Since the one peak of the purified protein eluted at fraction 41 (78.9 mL), the molecular weight of the native form was estimated to be about 15 kDa (1 subunit), as shown in Figures 3.20 and 3.21. The monomeric quaternary structure of OsBGal1 Cter was appropriate for structure determination by NMR.

A



B

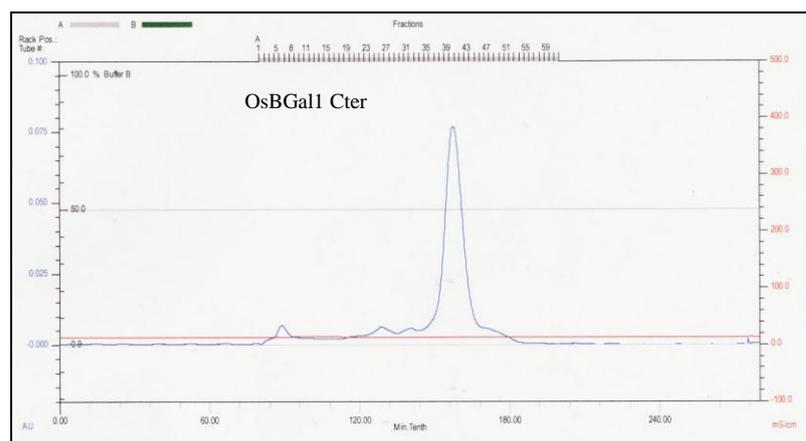


Figure 3.20 Elution profiles for estimation of molecular weight of the native form of the purified OsBGal1 Cter protein by Superdex S75 gel filtration chromatography. A, The 280 nm absorbance profile of the standard proteins elution from gel filtration on the Superdex S75 column. B, The 280 nm absorbance profile of the OsBGal1 Cter protein elution from the Superdex S75 column. The protein was eluted with 20 mM sodium phosphate buffer, pH 7.4, containing 100 mM NaCl, at flow rate of 0.5 mL/min.

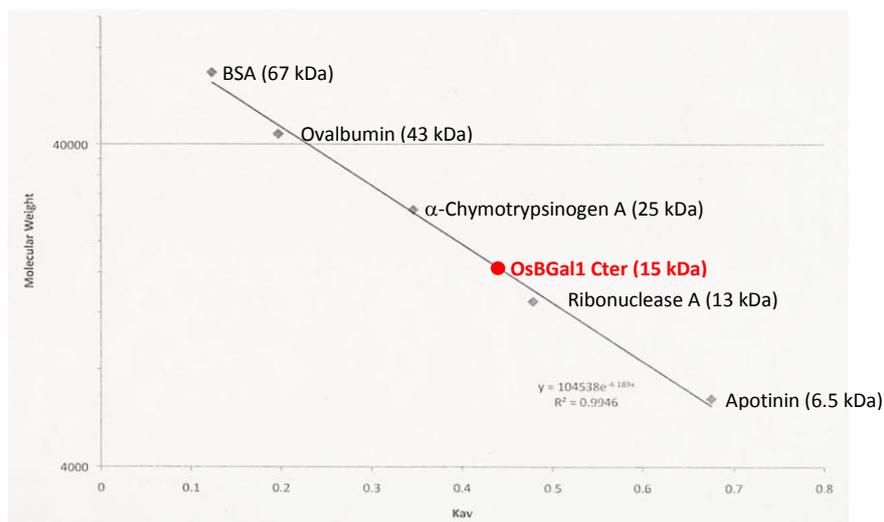


Figure 3.21 Estimation of molecular weight of the native form of the purified OsBGal1 Cter protein from the Superdex S75 gel filtration chromatography standard curve. The standard curve of the log molecular weight versus K_{av} was calibrated with the standard proteins, and the position of the OsBGal1 Cter on this curve is shown. The gel filtration standards are BSA (67 kDa), ovalbumin (43 kDa), α -chymotrypsinogen A (25 kDa), ribonuclease A (13 kDa) and aprotinin (6.5 kDa).

3.4 Determination structure of OsBGal1 Cter protein by NMR

The purified OsBGal1 Cter was labeled with ^{15}N , ^{13}C and $^{15}\text{N}/^{13}\text{C}$. Backbone assignments of OsBGal1 Cter were constructed from 3D HNCO, CBCA(CO)NH and HNCACB spectra. Side chain peaks for the OsBGal1 Cter were assigned from C(CO)NH and HCCH-TOCSY spectra.

3.4.1 Primary structure OsBGal1 Cter

The OsBGal1 Cter protein is composed 118 amino acid residues, including eight cysteine residues. Plant β -galactosidase C-terminal domains have been described as lectin-like domains, based on homology with animal lectins (Trainotti et al., 2001). Comparison of the OsBGal1 Cter sequence with those of other plant β -galactosidase C-terminal domains shows that the eight cysteine residues are conserved in rice, *Arabidopsis thaliana*, barley and strawberry (Figure 3.22). Moreover, a comparison of the sequence of the OsBGal1 Cter domain to those of lectin-like domains with known function and/or structures from animal species, such as the lectin from chum salmon (CSL3) and sea urchin (SUEL) and mouse latrophilin 1 (Lphn1), shows high conservation of eight cysteine residues in all the sequences (Figure 3.23).

3.4.2 Secondary structure of OsBGal1 Cter

The secondary structure of OsBGal1 Cter was extrapolated from chemical shift indexes (CSI) based on C α and C β chemical shifts (Metzler et al., 1993; Wishart et al., 1995). The result shows that the OsBGal1 Cter contains five β -strands (residues 731-735, 758-763, 768-784, 809-814, and 831-840) and a single, short α -helix (residues 798-803), as shown in Figure 3.23.

```

OsBa11 726 RTVSG-VCAADVSEYHPN-IKNWQI---ESYGEPE-FH-----TAKVHLKCA-PGQTIISAIKFASFGTPLGTGTFQQGECH-SINSNSVLEKKCIGLQRCVVAISPSNFGGD---PCPEVMKRVAVEAVCSTAA----- 843
OsBa13 718 RSIGV-VCAADVSEWQPS-MKNWHT---KDYEKA-----KVLQCD-NGQKITEIKFASFGTPQSGCSYQEGGCH-AHKSDFWFKNCVQGERCVSVVPEIFGGD---PCPGTMKRVAVEAVC----- 827
OsBa14 736 RTVAS-VCSFVSEHYPS-IDLESW---DRNTQNDGRD-----AAKQLSCP-KGKSISSVKFVDFGNPSGTCRSYQQGCH-HPNSISVVEKACLNMGCTVSLSDGFGED---LCPGVTKTLAIEADCS----- 851
OsBa15 739 VVPGA-VCT-----SGEAGD-----A--VTLSGG-GHAVSSVDVAFSGVGRGRGGYE-GGCE-SKAAAYEAFTAAACVGKESCTVEITGAFAGAG---C--LSGVLTVQATC----- 827
OsBa16 669 MSVTT-VCGNVDFSVPLQ-----SRGKV-----PKVRIWQ-QGGRISSEIEFASFGNPGDGRSFRIGSCH-AESSESVVKQSCIGRRCSIP VMAAKFGGDP---CPGIQKSLLVVADCR----- 775
OsBa18 743 RQTSS-ICAHVSEMHPAQIDSWIS---PQOTSQTQGP-----A--LRLEPREGQVISNIKFAFSGTSPGTCGNYNHGECSSQALAVVQEAACVGMTCNSVPPVSSNPFGDP---CSGVTKSLVVEAACS----- 956
OsBa110 711 VKRDN-ICTFMTEKNPAHVR-WSWESKDSQPKAVAGAGAGGLKPTAVLSCP-TKTKIQSVVFASFGNPLGICGNYTVGSCCH-APRTKEVVEKACIGRKTCSLVVSEVYGGDV---HCPGTTGTLAVQAKCSKRPPRSAATAQ--- 848
OsBa112 737 VVAGS-VCV-----SAEVD-----A--ITLSCGQHSTIISTIDVTSFGVARGQCGAYE-GGCE-SKAAAYKAFTACLGKESCTVQIINALTGSG---C--LSGVLTVQASC----- 828
OsBa113 781 HYAKT-VCSRISENYPPLSAWS--HLSSGRASVNA-----TPELRLOCD-DGHVISEITFASFGTSPGGCLNFSKGNCH-ASSTLDLVTEACVGNTKCAISVSNDFV-GD---PCRGVLKDLAVEAKCSPPSTTKEPRGEM-- 919
OsBa114 737 VVEGS-VCA-----SAEVD-----T--VTLSGGAHGRTISSVDVAFSGVARGRCGSYD-GGCE-SKVAYDAFAAACVGKESCTVLVTDAFANAG---C--VSGVLTVQATC----- 828
Arabi1 732 REVDS-VCADIYEWQST-LVNYQL---HASGKVNKPL-----HPKAHLCCG-PGQKITTVKFASFGTPEGTGGSYRQGSCH-AHHSYDAFNKLCVGNWCSVTVAPEMFGGD---PCPNVMKKLAVEAVCA----- 847
Arabi3 741 RSVSG-VCAEVSEYHPN-IKNWQI---ESYGGKQTFH-----RPKVHLCS-PGQAIASIKFASFGTPLGTGCSYQQGCH-AATSYAILERKCVGKARCAVTISNSNFGKD---PCPNVLKRLTVEAVCAPETSVSTWRP--- 856
Arabi7 731 VVVTG-VCARAHEHN-----KVELSC--HNRPISAVKFASFGNPLGHCGSFAVGTGQGDKAAKTVAKECVGKLNCTVNVSSDTFGSTL---DGDSPKKLAVELEC----- 826
Arabi8 735 KQTGSNLCLTVSQSHPPPVDTWTS---DSKISNRNRT-----RVLSLKCPISTQVIFSIKFAFSGTTPKGTGGSFTQGCN-SSRSLSLVQKACIGLRSNVEVSTRVFGEP---CRGVVKS LAVEASC----- 852
Arabi9 758 VTAGI-LCGQVSESHYPPLRKWPDPYINGTMSINSV-----APEVHLHCE-DGHVISSIEFASFGTSPRSGCDGFSIGKCH-ASNLSIVSEACKGRNSCFIEVSNTAFISD---PCSGTLKTLAVMSRCSPSQNMDSLFS--- 887
Arabi11 719 VNRDT-VCSYVGENYTPSVR--HWTRKQDQVAITD-----NVSLTATLCS-GTKKIAAEVAFSGNPIGVCNFTLGTCH-APVSKQVIEKHLGKAEVPIPNKSTFQDQKDCSKNVVMLAVQVRCGR----- 842
Arabi13 722 VNRDT-VCSYVGENYTPSVR--HWTRKNDQVAITD-----DVHLTANLCS-GTKKISAVEFASFGNPNGTGCFNFTLGSCH-APVSKVVEKYCLGKAEVPIPNKSTFQDQKDCSKVEKLLAVQVRCGR----- 845
Arabi14 774 VNRDT-ICSNVGEDYPPSVK--SWKREGPKIVRSK-----DMRLKAVMRCP-PEKQMVVQFASFGDPTGTGCFNFTMGKCS-ASKSKEVVEKECLGRNYSIIVVARETFG-DK--GPEIVKTLAVQVRCCKEKGKQDEKCKKED 853
Arabi15 785 IGVGS-VCANVYEKN-----VLELSC--NGKPISAIKFASFGNPGGDCGSFEKGTCEASNAAAAILTQECVGKEKCSIDVSEDKFGAA---ECGALAKRLAVEAIC----- 779
Arabi16 695 VSVTE-VCGHVSNTNPHFVI-----SPRKKGLNRKNTLYRYDRKPKVQLQCP-TGRKISKILFASFGTTPNGSCGSYSIGCH-SPNSLAVVQKACLKKSRCVSPVWSKTFGGDS---CPHTVKSLLVRAACS----- 815
Barley 707 VRRDD-ICVFISEHNPAQIK--TWDKGGQIKLIAEDHSTRG-----ILKCP-PKKTIQEVVAFSGNPEGSCANFTAGTCH-TPNAKDIVAKECLGKKSCLVPLVHTVYGADI--NCPPTTATLAVQVRCCH---PKNGEPE--- 833
Straw1 728 REVDS-VCADIYEWQPN-LMSWQM---QVSGRVNKPL-----RPKAHLCCG-PGQKISSIKFASFGTPEGVGGSFREGGCH-AHKSYNAFERSCIGQNSCSVTVSPENFGGD---PCPNVMKKLSVEAIC----- 843
Straw2 726 RQVES-LCSHVSESHPSFVDMWSS---DSKAGSKSRP-----R--LSLECPFPNQVISSIKFASFGTSPGTCGSFSGHSCR-SSRALSIVQKACVGSKCSIEVSTHTFGDP---CKGLAKSLAVEASC----- 840

```

Figure 3.22 Amino acid sequence comparison of OsBGal1 Cter., with other C-terminal domain from plant. The cysteine residues are shown in yellow. The sequence alignment was constructed by the Cobalt constraint-based multiple protein alignment tool from the National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov>).

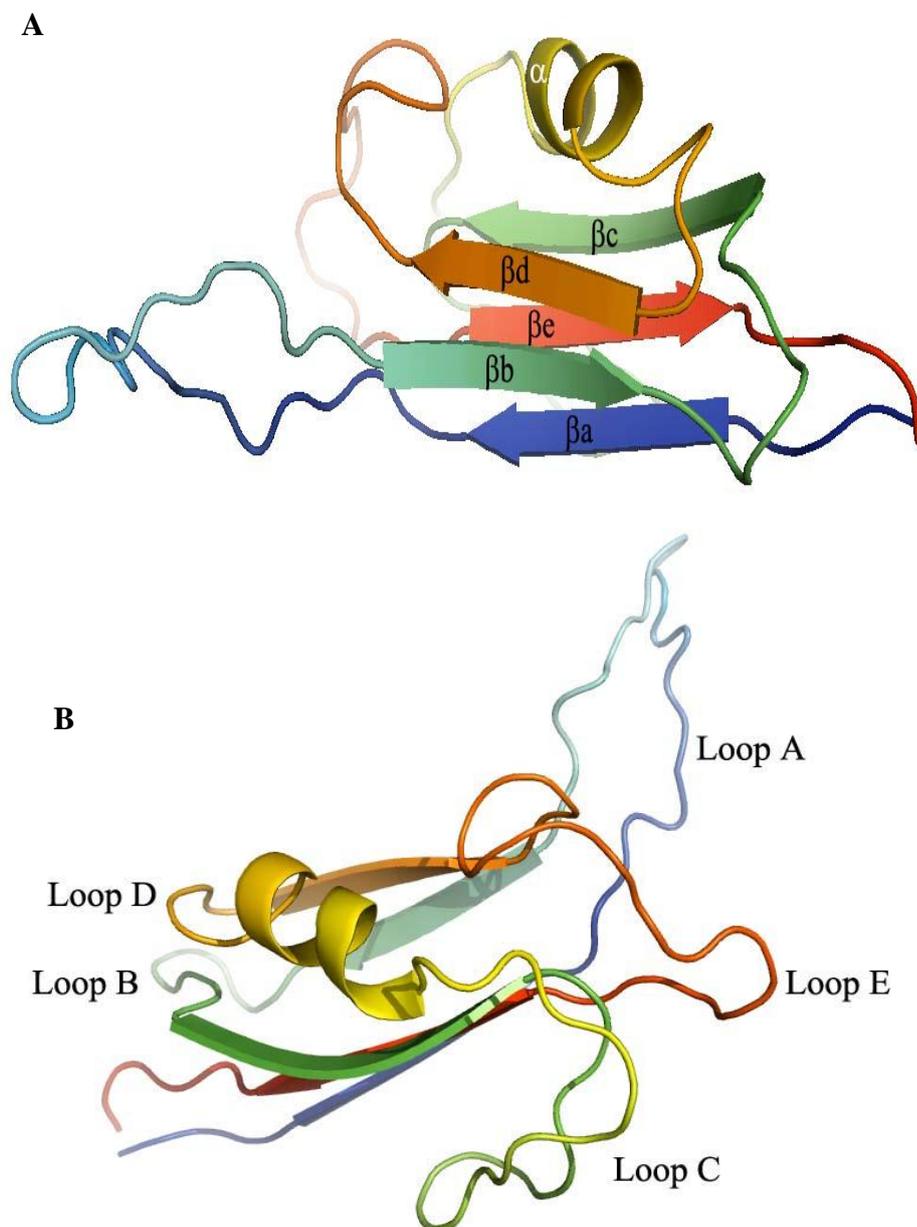


Figure 3.23 Amino acid sequence comparison of OsBGal1 Cter., chum salmon (CSL3), sea urchin (SUEL) and mouse latrophilin 1 (Lphn1). The cysteine residues are shown in yellow highlight with their linkages indicated with residue numbers from OsBGal1 Cter. Cyan highlight indicates residues involved in carbohydrate binding in the animal lectins. The sequence alignment was constructed by the Cobalt constraint-based multiple protein alignment tool from NCBI (<http://www.ncbi.nlm.nih.gov>).

3.4.3 The tertiary structure of OsBGal1 Cter

The tertiary structure of OsBGal1 Cter was determined from 1,820 distance geometry constraints. A complete analysis of the model characteristics is provided in Table 3.1. The overall fold of the OsBGal1 Cter shows a β sandwich with two antiparallel sheets (one composed of β -strands β_a , β_e , and β_c , and the other of β -strands β_b and β_d) enclosing the hydrophobic core (Figures 3.24 A). Moreover, this structure includes 5 loops (Figure 3.24 B), loop A connecting β_a and β_b (residues 736-757), loop B connecting β_b with β_c (residues 764-770), loop C connecting β_c and the α -helix (residues 779-797), Loop D connecting the α -helix and β_d (residues 806-810) and loop E connecting β_d with β_e (residues 816-832).

Four disulfide bridges are found in the OsBGal1 Cter structural core, one interconnecting β_a and loop C (Cys732–Cys785), one connecting the beginning of loop B with β_e (Cys764–Cys839), one between loop C and loop E (Cys793–Cys826), and one connecting loop D with β_d (Cys806–Cys812), as shown in Figures 3.24 and 3.25. The final ten lowest energy model structures of OsBGal1 Cter show critical differences in loop A (residues 735-763), as shown in Figure 3.26. This shows that its position was not well-defined by the data, suggesting it is very flexible.



Figures 3.24 The overall fold of the OsBGal1 Cter. **A**, the OsBGal1 Cter shows a β sandwich with two antiparallel sheets (one composed of strands βa , βe , and βc , and the other of strands βb and βd). **B**, Five loops in the OsBGal Cter structure (Loop A-Loop E). The figure was produced with the Pymol program (DeLano, 1991).

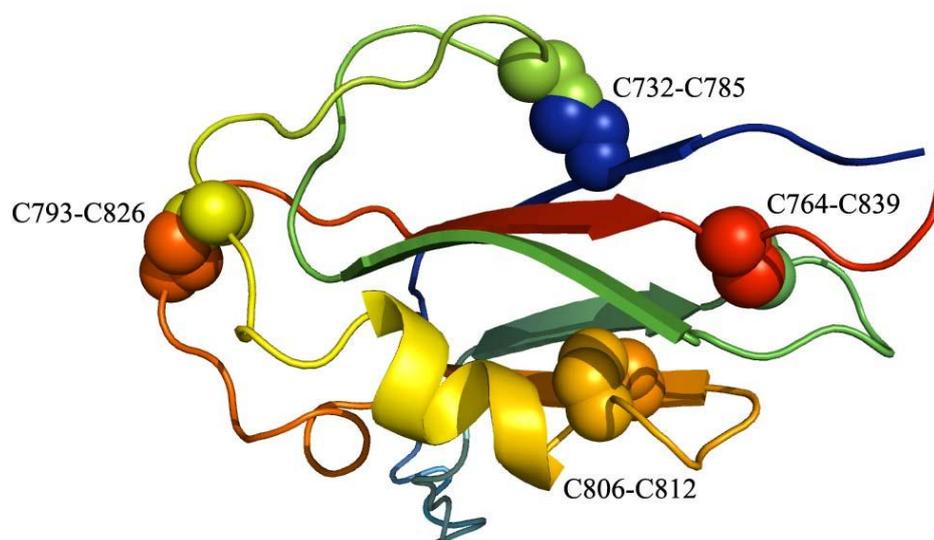


Figure 3.25 The disulfide bridges of OsBGal1 Cter. The cysteine residues forming disulfide bridges are indicated in space-filling representation, while the rest of the molecule is shown in cartoon representation. The structure is colored in a rainbow spectrum from the N-terminus (blue) to the C-terminus (red). The figure was produced with the Pymol program (DeLano, 1991).

Table 3.1 Structural statistics of the 30 structures of the C-terminal domain of OsBGal1^a

Rms deviations from experimental distance restraints (Å)	
All (1820)	0.016 ± 0.003
Interresidue sequential ($ i - j = 1$) (425)	0.019 ± 0.007
Interresidue short-range ($1 < i - j \leq 5$) (230)	0.021 ± 0.005
Interresidue long-range ($ i - j > 5$) (625)	0.011 ± 0.003
Intraresidue (528)	0.013 ± 0.008
Disulfide bridge (12)	0.011 ± 0.001
Energies (kcal mol ⁻¹)	
FNOE ^b	14.4 ± 4.2
Fcdih ^b	0.60 ± 0.22
Average rms differences (Å)	
Residues in α -helix and β -sheets	0.54 ± 0.06 (1.13 ± 0.13)

^a The number of each type of restraints used in the structure calculation is given in parenthesis. ^b FNOE and Fcdih were calculated using force constants of 50 kcal mol⁻¹ Å⁻² and 200 kcal mol⁻¹ rad⁻², respectively. ^c The average pairwise rms differences are given for selected residues. The value for backbone atoms (N, Ca, and C') is followed by that for all heavy atoms in parenthesis.

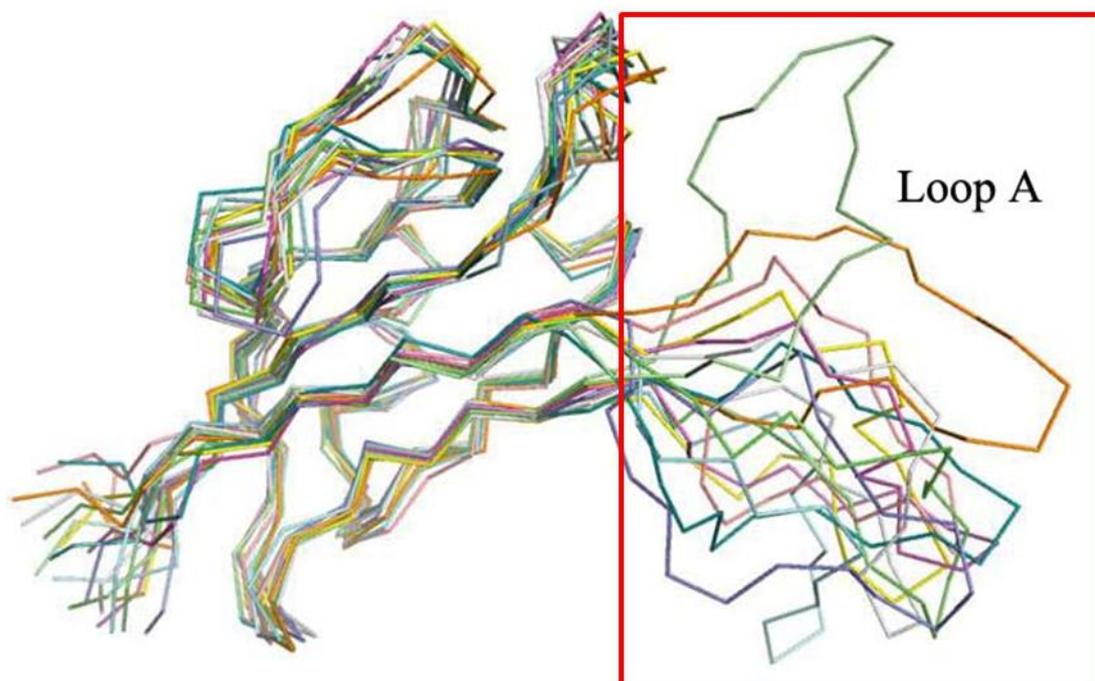


Figure 3.26 Superimposition of the final 10 lowest energy structures of the OsBGal1 Cter. Loop A of OsBGal1 Cter was not well-defined, suggesting it is flexible, as shown in the red box. The figure was produced with Pymol (DeLano, 1991).

3.4.4 Overall structure comparison of OsBGal1 Cter with related animal lectins

OsBGal1 Cter shares 13% and 17% amino acid sequence identity with the lectin domains with known structures from mouse latrophilin (2JXA, Vakonakis et al., 2008) and chum salmon (2ZX4, Shirai et al., 2009), respectively. Although these levels of overall similarity are too low to confirm sequence homology, the conservation of the disulfide bonding cysteines and a few other structurally important residues was enough to identify them as homologous (Trainotti et al., 2001). The animal lectin structures were superimposed on the rice OsBGal1 Cter structure for comparison (Figure 3.27).

The structural superimposition showed that the core structures of OsBGal1 Cter and the animal lectins are similar in shape, while differences are seen in the loops. Three variable loop regions have been described that connect the β -strands and α -helix: Loop A between strands β a and β b, Loop C between strand β c and the α -helix, and Loop E between strands β d and β e. Loop A has the largest difference in length (it is much longer in OsBGal1 Cter, as seen in Figures 3.23 and 3.24) and structural alignment for the three structures. As noted above, loop A appears to be very flexible (Figure 3.26). Loop C contains the conserved cysteine residue and is a longer loop in CSL3 and Lphn1 than in Os1BGal Cter. Loop E is similar to the corresponding loop, designated variable loop 2 (Vakonakis et al., 2008, Shirai et al., 2009), in Lphn1 and CSL3 and contains a conserved proline-containing sequence, which is GDPCP in OsBGal1 Cter and SUEL, PDPCPG in Lphn1, and GDPCV in CSL3 (Figure 3.23)

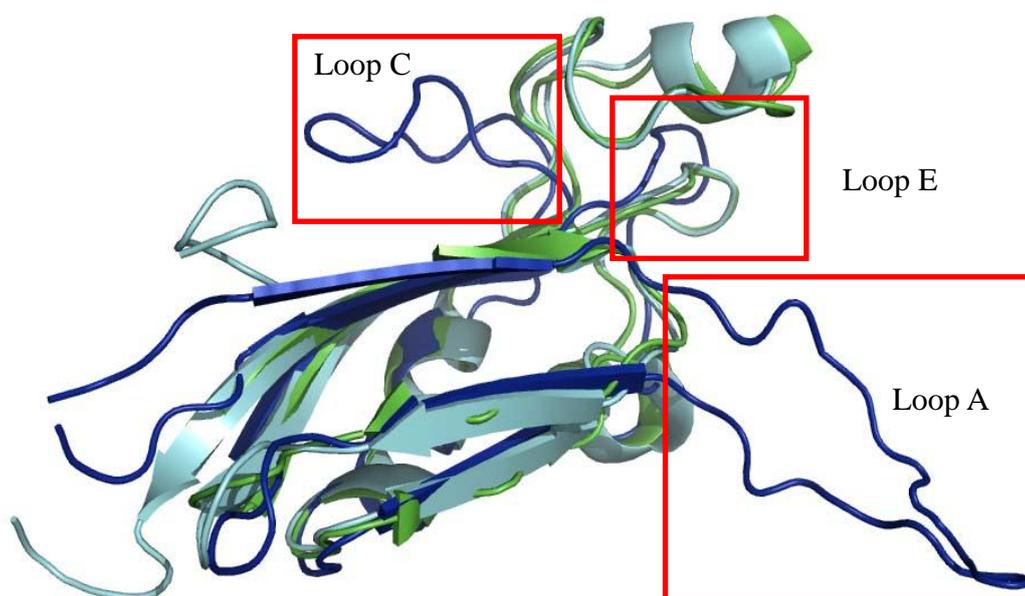


Figure 3.27 Superimposition of the structures of rice OsBGal1 Cter (blue), with other lectin domain of mouse latrophilin (2JXA, cyan) and chum salmon (2ZX4, green). The loops discussed in the text are boxed. The figure was produced with the Pymol program (DeLano, 1991)

The carbohydrate binding site of the mouse latrophilin 1 RBL domain and chum salmon with rhamnose are located on an exposed pocket formed by loop E, also called loop 2, as noted above. In the structure, two residues of loop 2, N and N^ζ of Lys120 and N of Gly117 from mouse latrophilin 1, and the corresponding N and N^ζ of Lys185 and N of Gly182 from chum salmon are directly involved in hydrogen bonding interactions with rhamnose O4, O3 and O2, respectively (Figure 3.28). The side chain hydroxyl of Tyr63 (mouse latrophilin 1)/Tyr126 (chum salmon egg) from β3 bind to rhamnose at O3 and the O^{ε1} and O^{ε2} of Glu42 (mouse latrophilin 1)/Glu106 (chum salmon egg) from end of β1 contribute two hydrogen bonds to rhamnose O3 and O4, respectively (Shirai et al., 2009 and Vakonakis et al., 2008). In comparison, mouse latrophilin 1 bind galactose via forming hydrogen bonds from N and N^ζ of Lys120 to Gal O2 and Gal O3, N of Gly117 to Gal O4, the O^{ε1} of Glu42 to galactose O2 and O3. The hydroxyl of Tyr63 bind to galactose at O3 (Figure 3.29).

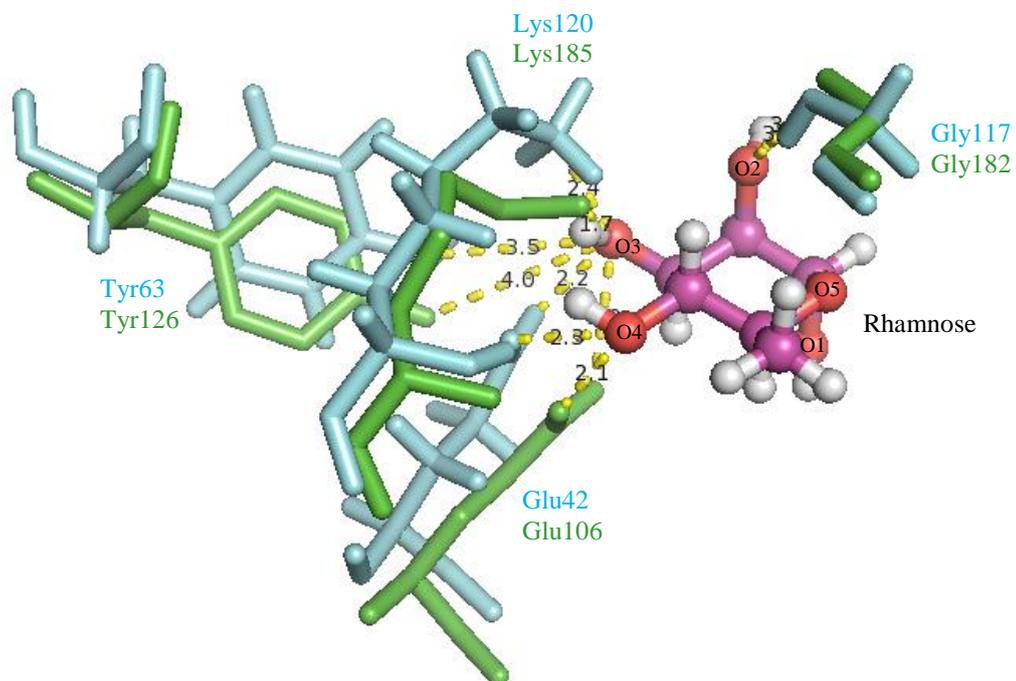


Figure 3.28 Superimposition of the structures of other lectin domain of mouse latrophilin (2JXA, cyans) and chum salmon (2ZX4, green). The figure was produced with the Pymol program (DeLano, 1991).

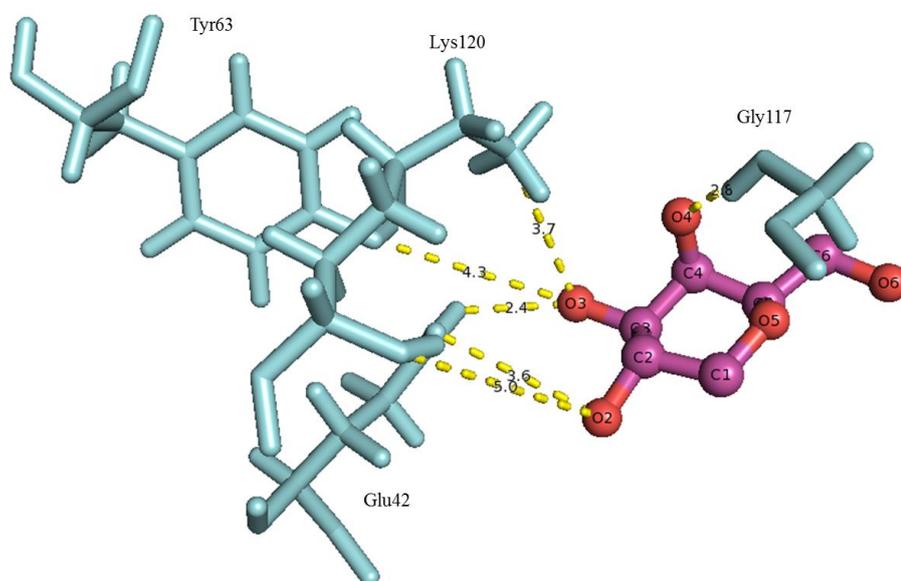


Figure 3.29 Residues use for binding to galactose from mouse latrophilin.. The figure was produced with the Pymol program (DeLano, 1991).

3.5 Investigation of binding of OsBGal1 Cter protein to sugar

3.5.1 HSQC NMR method

3.5.1.1 The ^1H - ^{15}N HSQC spectra of OsBGal1 Cter was observed by NMR and this data was used as the reference (Figure 3.30). After that the OsBGal1 Cter was mixed with galactose (Figure 3.31), rhamnose (Figure 3.32), glucose (Figure 3.33) and raffinose (Figure 3.34) to look for binding. If a peak from the HSQC spectrum of the protein alone was lost, shifted or split into two peaks that would mean it was involved in binding or moved upon binding. No obvious significant losses, shifts or splitting of peaks were observed in the HSQC spectra, suggesting that none of these sugars bound to the OsBGal1 Cter protein under the conditions tested. Although, many paper denote this domain as a galactose-binding or carbohydrate-binding domain, this data provided no evidence that it bound to sugar. Although OsBGal1 Cter had similar structure to rhamnose binding lectin domains, it cannot bind to rhamnose and galactose (Figure 3.29 and Figure 3.30). The residues used for binding to rhamnose and galactose of the mouse latrophilin 1 RBL domain are Glu42, Tyr63, Gly117 and Lys120 and the corresponding to residues Glu106, Tyr126, Gly182 and Lys185 in chum salmon lectin. Comparison of the residues that related to binding to sugar between OsBGal1 Cter, mouse latrophilin 1 and chum salmon lectin, show that residues from OsBGal1 Cter not same in mouse latrophilin 1 and chum salmon lectin (Figure 3.35).

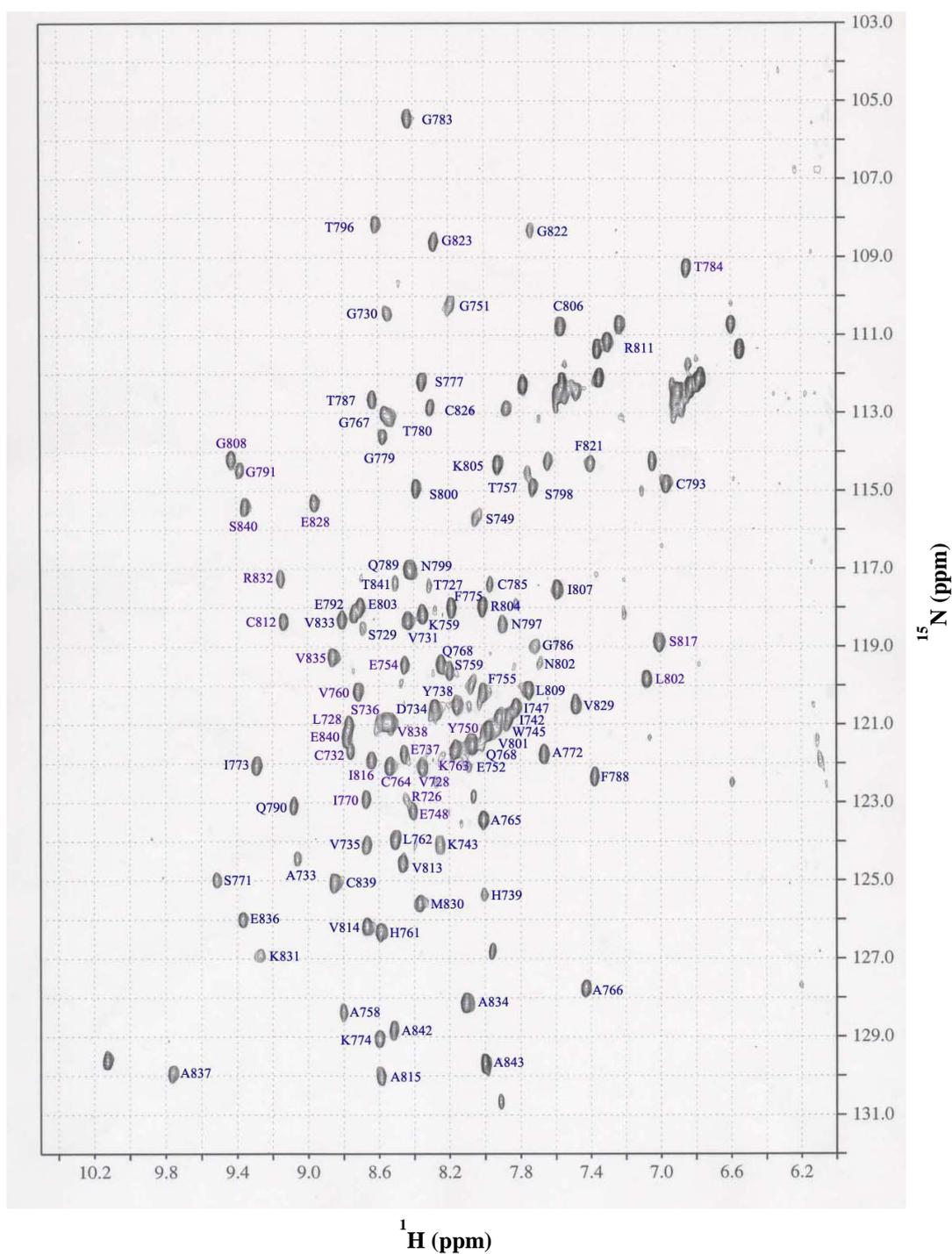


Figure 3.30 ^1H - ^{15}N HSQC NMR spectrum of OsBGal1 Cter in phosphate buffer in D_2O (20 mM sodium phosphate buffer, 100 mM NaCl pH 8.0) at 25 °C.

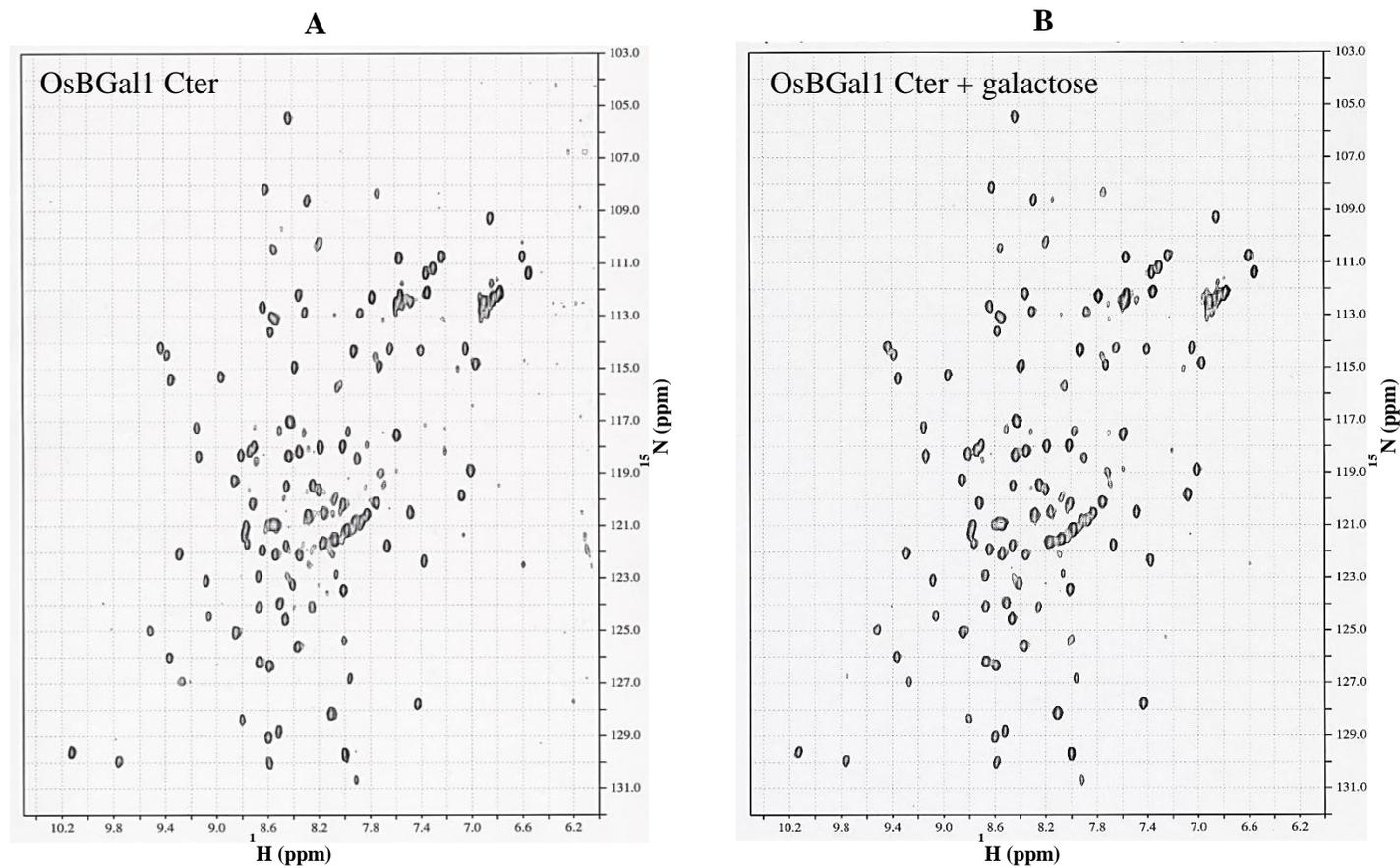


Figure 3.31 Test of the binding of OsBGal1 Cter with galactose by HSQC in NMR. A, Control OsBGal1 Cter. B, OsBGal1 Cter mixed with 10 molar equivalents of galactose compared to protein in phosphate buffer in D_2O (20 mM sodium phosphate buffer, 100 mM NaCl pH 8.0).

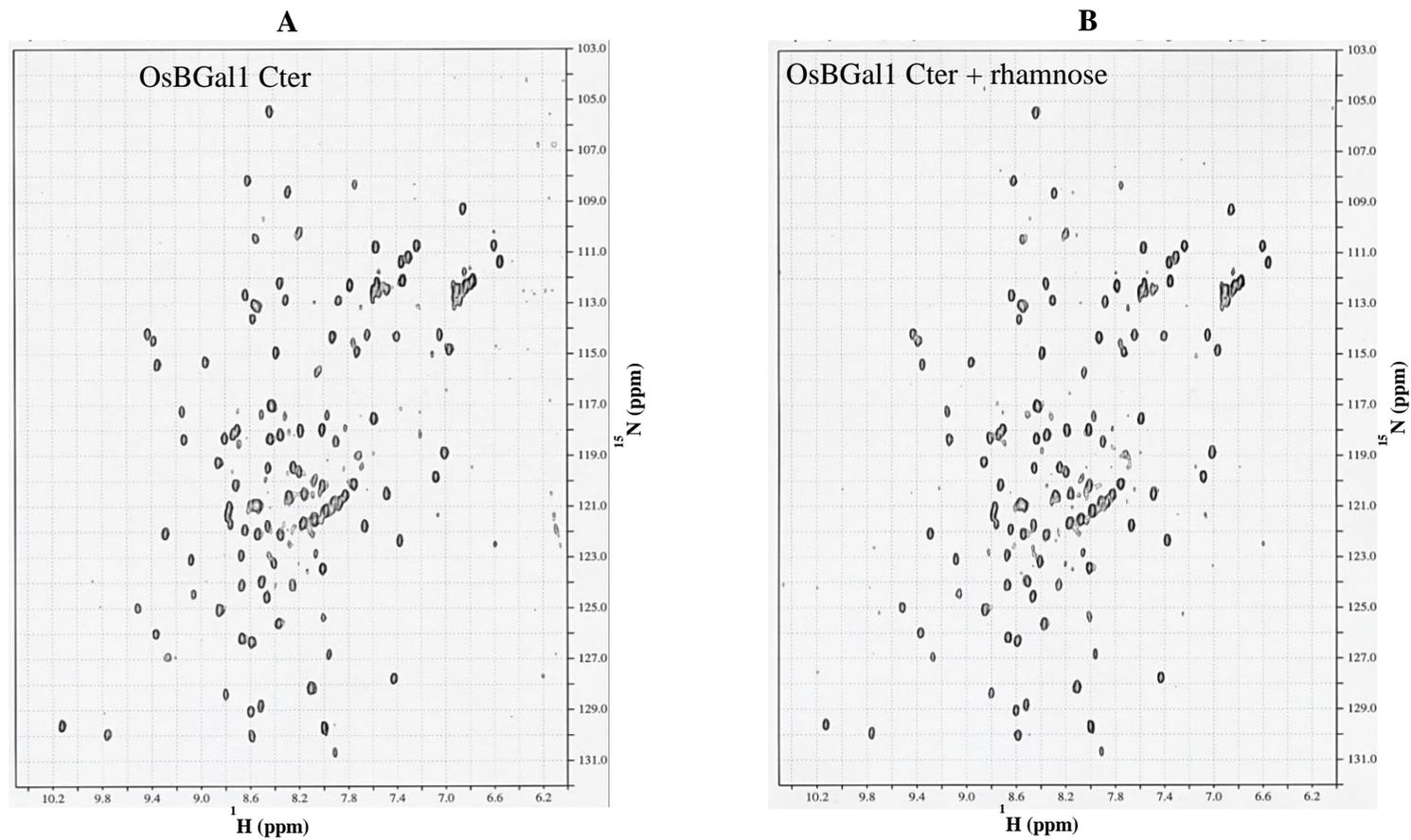


Figure 3.32 Test of the binding of OsBGal1 Cter with rhamnose by HSQC in NMR. A, Control OsBGal1 Cter. B, OsBGal1 Cter mixed with 10 molar equivalents of rhamnose compared to protein in phosphate buffer in D₂O (20 mM sodium phosphate buffer, 100 mM NaCl pH 8.0).

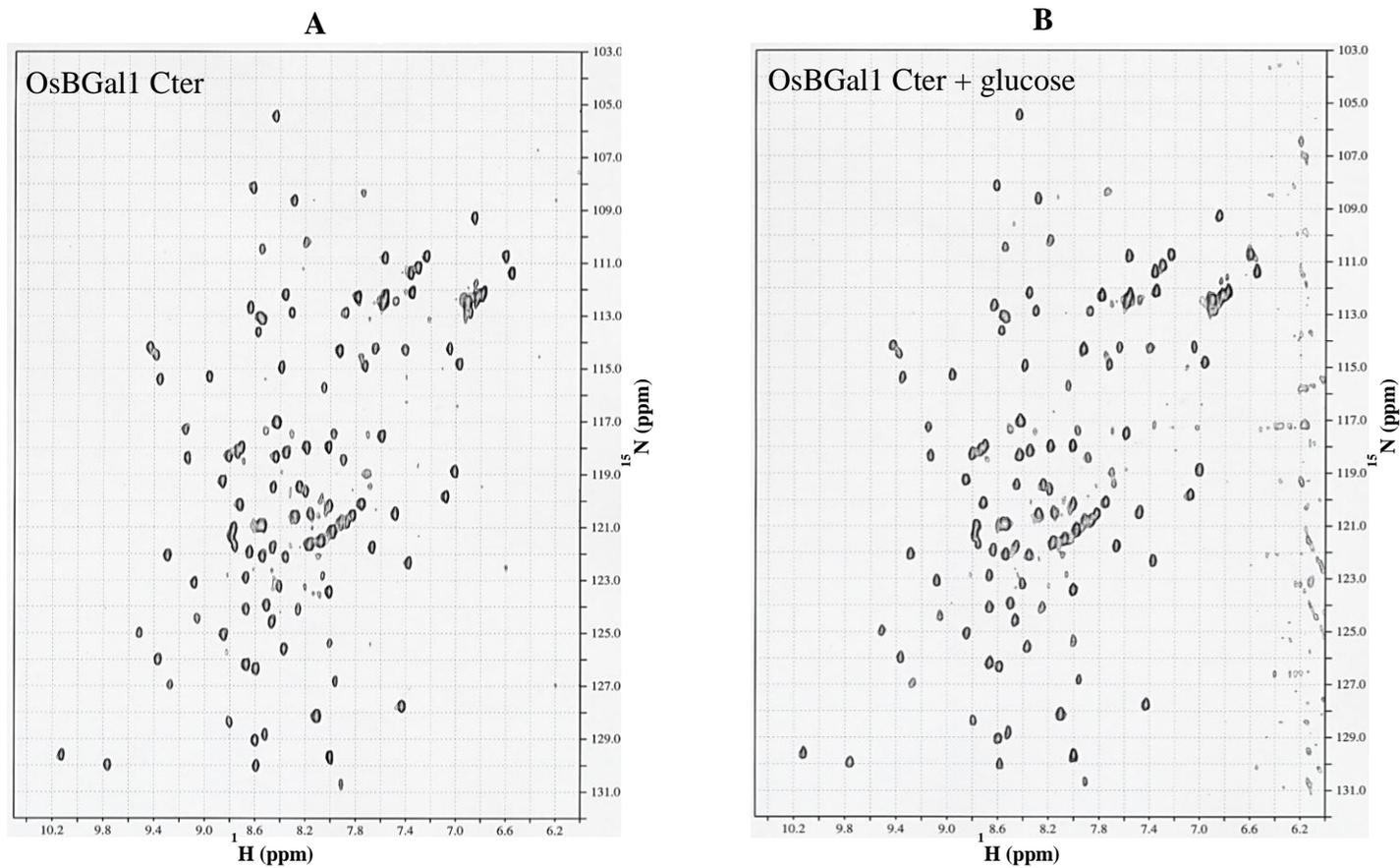


Figure 3.33 Test of the binding of OsBGal1 Cter with glucose by HSQC in NMR. A, Control OsBGal1 Cter. B, OsBGal1 Cter mixed with 10 molar equivalents of glucose compared to protein in phosphate buffer in D_2O (20 mM sodium phosphate buffer, 100 mM NaCl pH 8.0).

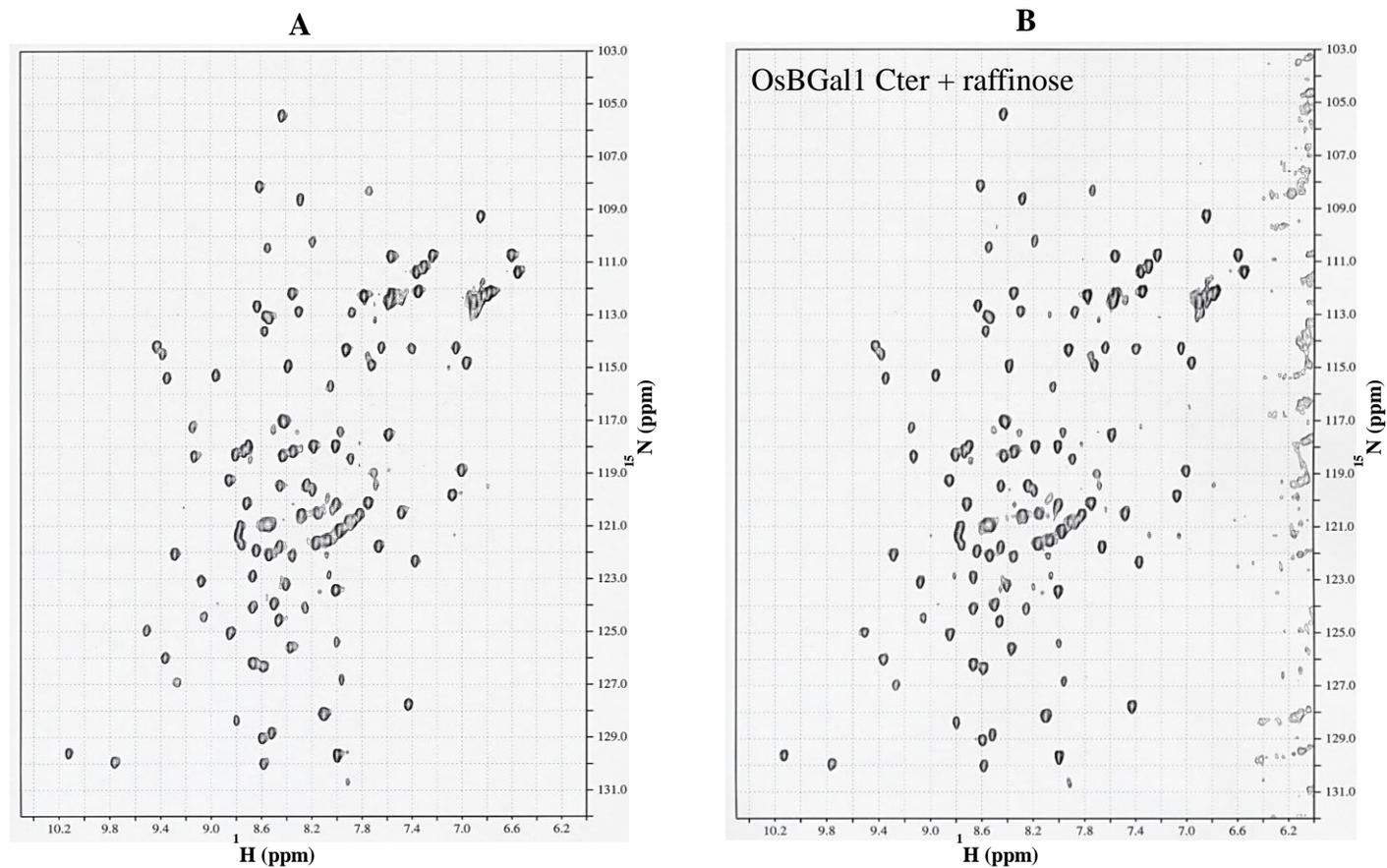


Figure 3.34 Test of the binding of OsBGal1 Cter with raffinose by HSQC in NMR. A, Control OsBGal1 Cter. B, OsBGal1 Cter mixed with 10 equivalents of raffinose compared to protein in phosphate buffer in D_2O (20 mM sodium phosphate buffer, 100 mM NaCl pH 8.0).

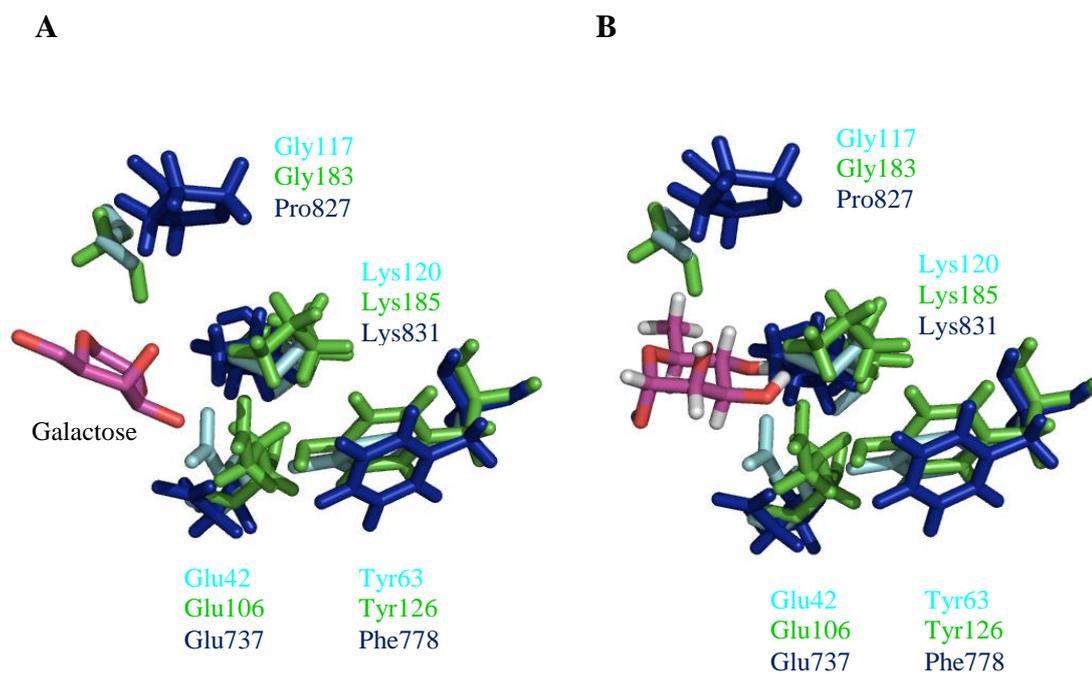


Figure 3.35 Residues comparison for binding between other lectin domain of mouse latrophilin (2JXA, cyans) and chum salmon (2ZX4, green) and OsBGal1 Cter (blue). The figure was produced with the Pymol program (DeLano, 1991).

3.5.1.2 Stability of OsBGal 1Cter in HSQC experiments

The stability of OsBGal 1Cter was observed by ^1H - ^{15}N HSQC spectra in NMR. In terms of stability, at 25°C, single labeled (^{15}N) protein was generally degraded within 2 weeks (Figure 3.36) and double labeled ($^{13}\text{C}/^{15}\text{N}$) was very stable at 25°C for around 1 month (Figure 3.37). Unlabeled protein seemed to generally be degraded faster, but it could not be studied by this technique.

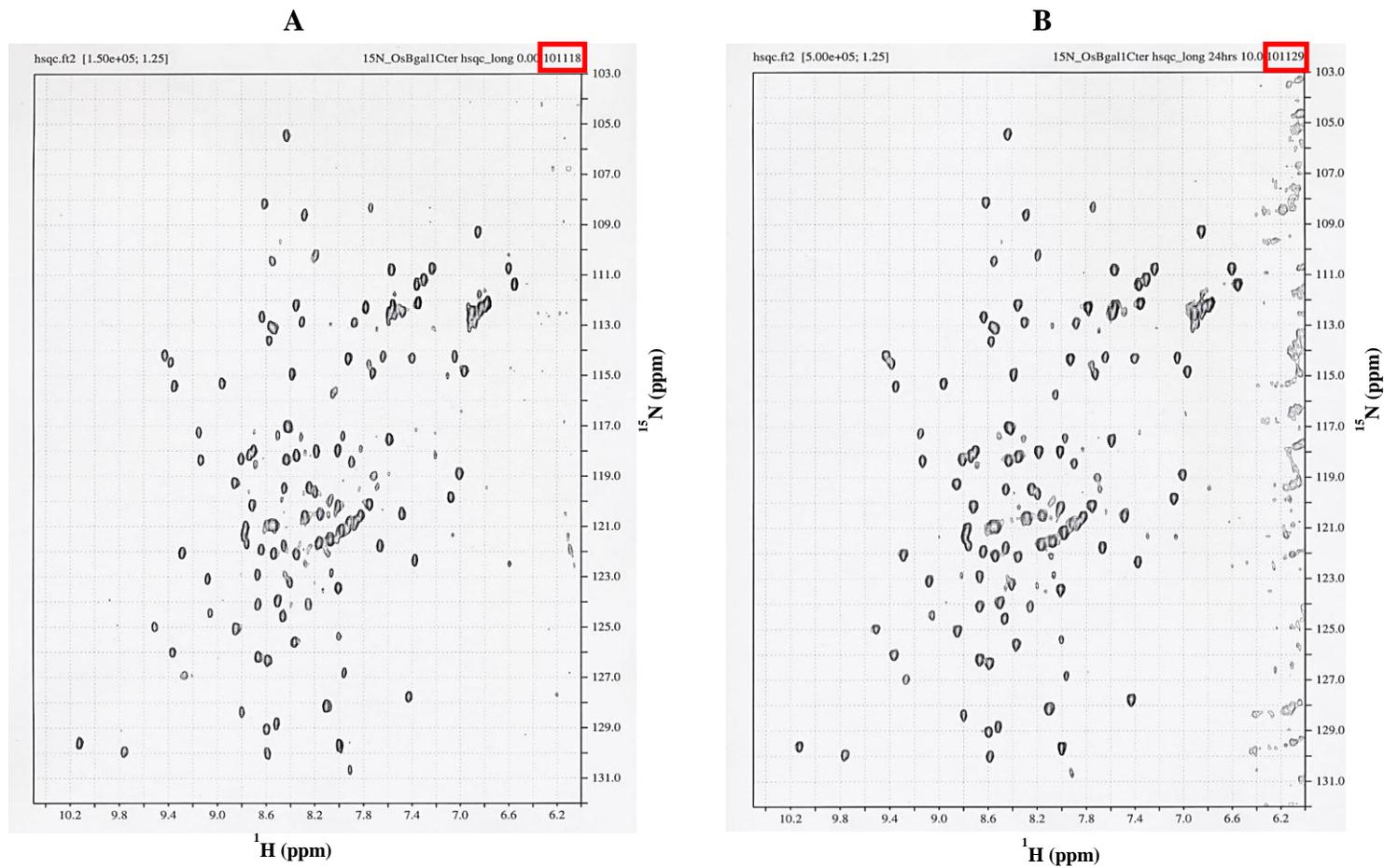


Figure 3.36 Stability label ^{15}N of OsBGal1 Cter in NMR. A, ^1H - ^{15}N HSQC spectrum start date of analysis. B, Spectrum at the end of the analysis (11 days).

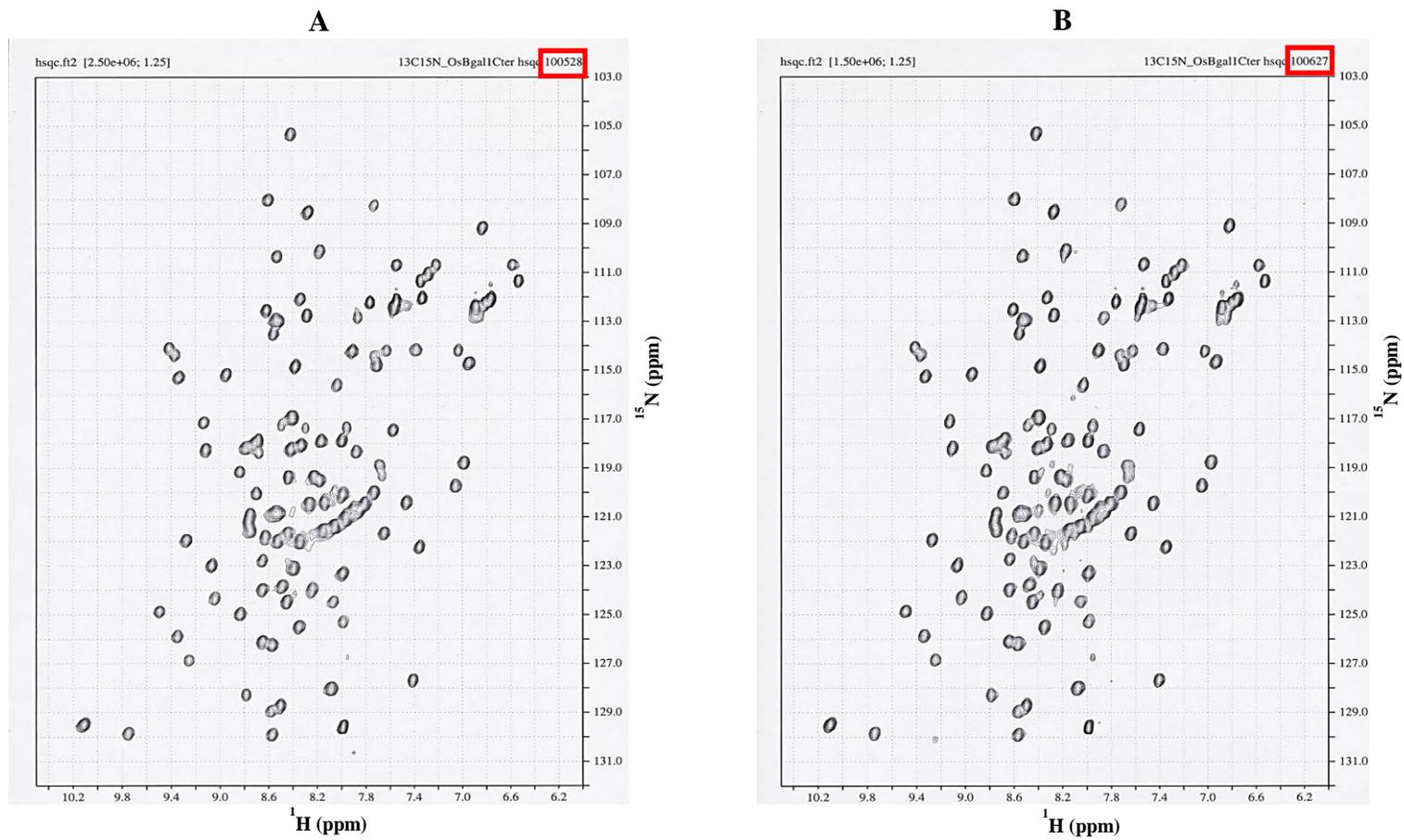


Figure 3.37 Stability label $^{13}\text{C}^{15}\text{N}$ of OsBGal1 Cter in NMR. A, ^1H - ^{15}N HSQC spectrum start date of analysis. B, Spectrum at the end of the analysis (32 days).

3.5.2 Carbohydrate microarray

The carbohydrate microarray was applied to investigate whether the OsBGal1 Cter protein binds to different sugars. The sugar spotted on the membrane included oligosaccharides and polysaccharides. The appearance of immunoreactive spots in the membrane incubated with OsBGal1 Cter, but not in the control membrane incubated with thioredoxin-tag showed that OsBGal1 Cter appeared to bind to the sugars listed in Table 3.2, as indicated in Figure 3.38.

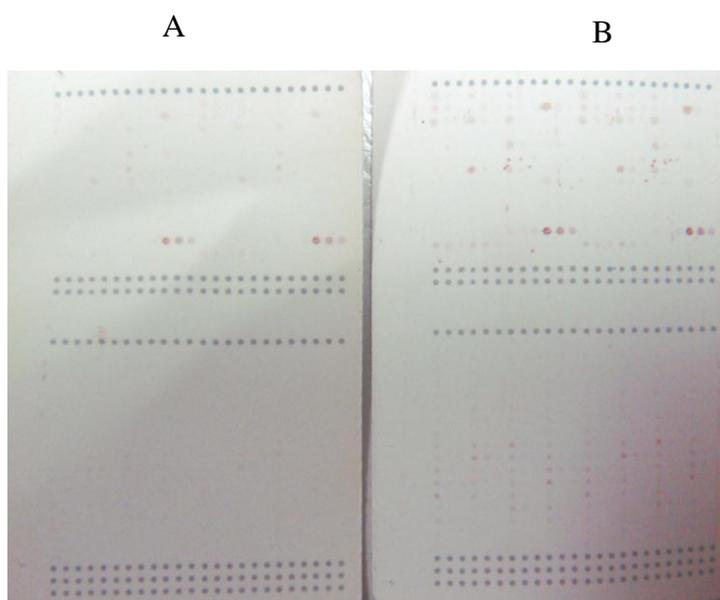


Figure 3.38 OsBGal1 Cter binding to spots on the carbohydrate array membrane. Immunoblots with anti-OsBGal Cter antiserum were developed after incubation of the membranes with A, control thioredoxin fusion tag, and B, OsBGal1 Cter.

Table 3.2 Names of sugars to which OsBGal1 Cter appeared to bind in the array.

polysaccharides	oligosaccharides
Galactomannan (crob)	α -(1→5)-L-Arabinotriose
Glucomannan (konjac)	α -(1→5)-L-Arabinohexaose
Tamarind seed xyloglucan	β -(1→4)-D-Galactobiose
MLG Lichenan, β -glucan (1→3),(1→4)- β -D-gucan	6 ¹ - α -D-Galactosyl-(1-4)- β -D-mannobiose/ manotriose
Hydroxypropyl Celulose	(1→4)- β -D-Mannotriose
Lime pectin DE: 81% (E81)	isoprimeverose (α -D-xylopyranosyl (1→6)glucose)
Sugar beet arabinan	(1→4)- β -D-Xylobiose
	<i>Hexa</i> acetyl-chitopentaose
	6 ² - β -D-Galactosyl-(1→4)- β -D-galactotriose
	β -(1→4)-D-Galactobiose, feruloylated

3.5.3 Saturation transfer difference (STD) method

The result from carbohydrate array indicated that the OsBGal1 Cter may bind to certain sugars, such as α -(1 \rightarrow 5)-L-arabinotriose and β -(1 \rightarrow 4)-D-galactobiose. To confirm the result from the carbohydrate microarray and find which sugar that bind to this protein, the STD NMR method was used to access atom-specific binding information. The sample contained OsBGal1 Cter (30 μ M) in the presence of a 50 molar excess of 1,5- α -L-arabinotriose. The protein-saturating irradiation was varied at 7.25 ppm (3629 Hz), 7.04 ppm (3524 Hz), 6.67 ppm (3436 Hz), 6.27 ppm (3139 Hz), 3.05 ppm (1528 Hz), 2.60 ppm (1300 Hz), 1.47 ppm (1011 Hz), (735 Hz) and 0.74 ppm (372 Hz), respectively. Although differences were detected in the irradiated region, no consistent difference in sugar peaks was detected, which suggested that OsBGal1 Cter protein could not bind to 1,5- α -L-arabinotriose under the conditions tested, even though irradiation energy was varied (Figure 3.39).

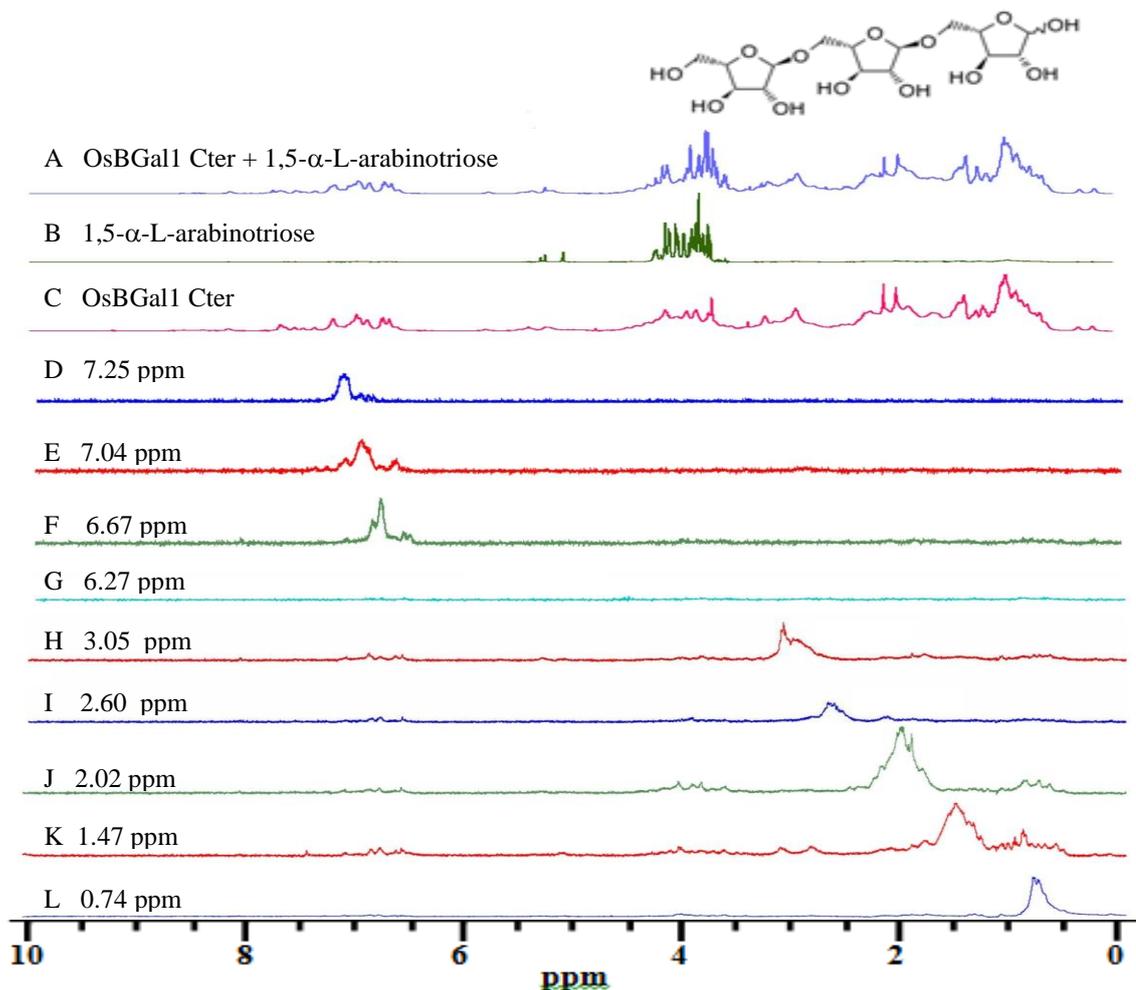


Figure 3.39 STD NMR of OsBGal1 Cter with 1,5- α -L-arabinotriose. A, reference ^1H -spectrum mixture of OsBGal1 Cter with 1,5- α -L-arabinotriose. B, reference ^1H -spectrum of 1,5- α -L-arabinotriose. C, reference ^1H -spectrum of OsBGal1 Cter. D-L, STD difference spectra of the spectrum of the solution of OsBGal1 Cter (30 μM) with 1,5- α -L-arabinotriose (1.5 mM) with the spectra when the solution was irradiated at 7.25 ppm (3629 Hz), 7.04 ppm (3524 Hz), 6.67 ppm (3436 Hz), 6.27 ppm (3139 Hz), 3.05 ppm (1528 Hz), 2.60 ppm (1300 Hz), 1.47 ppm (1011 Hz), (735 Hz) and 0.74 ppm (372 Hz), respectively.

3.6 Expression of OsBGal1 from cDNA optimized for *P. pastoris*

A construct in which the OsBGal1 cDNA was inserted in pET32a(+) was first used to express an NH₂-terminal thioredoxin-His6-tagged fusion protein in *Escherichia coli* strain Origami B(DE3) cells, as this system has been successfully used for expression of other plant glycoside hydrolases (Chantarangsee et al., 2007). Under the conditions tested, the OsBGal1 protein was relatively low in the soluble fraction upon cell extraction. As noted in the Introduction, expression of OsBGal1 in *Pichia pastoris* was also attempted, but expression was low and purification was not achieved (Chantarangsee and Ketudat Cairns, unpublished).

Expression of OsBGal1 from an optimized gene in yeast was attempted by inserting the DNA into the pPICZ α BNH₈ expression vector, to produce a protein fused to the α -factor prepropeptide for secretion (Figure 3.40). This construct contains an eight-histidine tag at the NH₂-terminus of optimized OsBGal1 (AHHHHHHHHAA) (Figure 3.41).

```

nat_OsBGal1  GTGACGTACGAC AAGAAAGCGGTGCTCGTCGACGGCCAGAGGAGGATTCTCTTCTCCGGA 60
opt_OsBGal1  GTTACTTACGATAAGAAAGCTGTGTTGGTTGACGGTCAAAGAAGAATTTGTTTTCCGGA 60
          V T Y D K K A V L V D G Q R R I L F S G

nat_OsBGal1  TCCATACATTACCCGAGGAGCACACCCGAAATGTGGGACGGGCTAATTGAGAAGGCTAAA 120
opt_OsBGal1  TCAATCCATTATCCTAGATCTACACCAGAAATGTGGGATGGTTTCGATTGAGAAGGCTAAA 120
          S I H Y P R S T P E M W D G L I E K A K

nat_OsBGal1  GATGGAGGCTTGGATGTGATCCAGACCTATGTCCTTTGGAATGGCCATGAACCAACTCCT 180
opt_OsBGal1  GATGGTGGATTGGACGTTATCCAAACTTACGTTTTCGGAATGGACATGAACCAACAACCT 180
          D G G L D V I Q T Y V F W N G H E P T P

nat_OsBGal1  GGAATTACAATTTTGAAGGAGGTACGATCTGGTCAGGTTTCATCAAGACTGTCCAGAAG 240
opt_OsBGal1  GGTAAC TACAATTTTGAAGGAGATACGATTTGGTTAGATTTCATTAAGACTGTTCAAAA 240
          G N Y N F E G R Y D L V R F I K T V Q K

nat_OsBGal1  GCTGGCATGTTTTGTTTCATCTCCGCATCGGTCCCACATTTGTGGAGCTGGAATTTTGGG 300
opt_OsBGal1  GCTGGAATGTTTTGTTCACTTGAGAATTGGACCAACATCTGTGGTGAATGGAACTTTTGGT 300
          A G M F V H L R I G P Y I C G E W N F G

nat_OsBGal1  GGATTTCCAGTTTGGTTGAAGTATGTTACCAGGCATCAGCTTCAGGACGGACAATGAACCT 360
opt_OsBGal1  GGATTTCCCTGTTTGGTTGAAGTATGTTACCAGGTATTTCTTTAGAACTGATAATCAACCT 360
          G F P V W L K Y V P G I S F R T D N E P

nat_OsBGal1  TTCAAGAATGCAATGCAGGGGTTACAGAGAAAATGTTGGGCATGATGAAGAGTGAAAAC 420
opt_OsBGal1  TTCAAAAACGCTATGCAAGGTTTACAGAAAAGATCGTTGGAATGATGAAATCAGAGAAAT 420
          F K N A M Q G F T E K I V G M M K S E N

nat_OsBGal1  CTCTTTGCTTCA CAAGGCGGTCTATTATCCTCTCTCAGATTGACAACGAGTATGGCCCA 480
opt_OsBGal1  TTGTTTCCCTTCTCAAGGTGGACCTATTATCTTGTCTCAGATTGAAAACGAGTACGGTCCA 480
          L F A S Q G G P I I L S Q I E N E Y G P

nat_OsBGal1  GAAGGTAAAGAGTTTGGGGCTGCCGGCAAGGCATATATCAACTGGGGCGCAAAGATGGCC 540
opt_OsBGal1  GAAGGAAAGGAGTTTGGTGGCTGCTGGAAAAGCTTATATCAATTGGGGCTGCTAAGATGGCT 540
          E G K E F G A A G K A Y I N W A A K M A

nat_OsBGal1  GTCGGATTGGACACCGGTGTGCCGTGGGTGATGTCGAAAGGAGATGACGCACCTGACCCA 600
opt_OsBGal1  GTTGGTTTGGATACTGGAGTTCATGGGTTATGTCGTAACAAGATGACGCTCCAGATCCT 600
          V G L D T G V P W V M C K E D D A P D P

nat_OsBGal1  GTGATCAATGCATGCAATGGTTTTCTATTGTGACACATTTTCTCCTAACAAGCCTTACAAG 660
opt_OsBGal1  GTTATTAATGCTTGTAAACGGTTTTTACTGTGACA TTTCTCACCAACAAGCCTTATAAA 660
          V I N A C N G F Y C D T F S P N K P Y K

nat_OsBGal1  CCTACGATGTGGACTGAAGCTTGGACTGGATGGTTTACTGAAATTCGGAAGGAACCATCCCT 720
opt_OsBGal1  CCAACTATGTGGACA GAAGCTTGGCTGGATGGTTTACTGAGTTCCGGTGAACAATCAGA 720
          P T M W T E A W S G W F T E F G G T I R

nat_OsBGal1  CAACGACCAGTTGAAGATCTCGCATTTGGTGTGTGCTCGCTTCGTA CAGAAGGGTGGTCTCT 780
opt_OsBGal1  CAAAGACCAGTTGAAGATTTGGCTTTTGGTGTGTGCTAGATTTCGTT CAGAAGGGTGGATCT 780
          Q R P V E D L A F G V A R F V Q K G G S

nat_OsBGal1  TTTATCAACTACTACATGTATCATGGAAGAACGAAATTTTGGTCCGACGGCTGGAGGTCCC 840
opt_OsBGal1  TTTATTAACTACTACATGTACCATGGTGAACTAACCTTCGGTAGAACAGCTGGTGGACCT 840
          F I N Y Y M Y H G G T N F G R T A G G P

```

Figure 3.40 Alignment of the nucleotide sequences of the native (upper) and optimized (lower) OsBGal1 cDNA. Both sequences encode the same 817 amino acid residues.

```

nat_OsBGal1 TTTATCACCACGAGCTATGATTATGATGCTCCACTTGATGAATATGGTCTTGCAAGCGAA 900
opt_OsBGal1 TTTATCACTACATCCTACGATTATGACGCTCCCTTGATGAATATGGTCTTGCTAGAGAG 900
          F I T T S Y D Y D A P L D E Y G L A R E

nat_OsBGal1 CCAAAGTTTGGGCACCTTAAAGAACTCCATAGAGCTGTTAAGTTATGTGAGCAGCCTTTG 960
opt_OsBGal1 CCAAAGTTTCGGACATTTGAAAGAATTGCACAGAGCTGTTAAATTTGTGAGCAACCTTTG 960
          P K F G H L K E L H R A V K L C E Q P L

nat_OsBGal1 GTTCTGCCGATCCAACCTGTGACTACCCTTGGAGTATGCAAGAGGCCCATGTGTTCCGA 1020
opt_OsBGal1 GTTCTGCTGACCCAACCTGTACTACATTTGGATCCATGCAGGAAGCTCACGTTTTCAGA 1020
          V S A D P T V T T L G S M Q E A H V F R

nat_OsBGal1 TCTTCCTCTGGCTGTGCAGCTTTCCTTGCAAACTACAATTCTAACTCGTATGCCAAAGTT 1080
opt_OsBGal1 TCTTCCTCAGGTTGTGCTGCTTTCTTGGCTAACACTACAATTCTAACTCCTACGCTAAGCTT 1080
          S S S G C A A F L A N Y N S N S Y A K V

nat_OsBGal1 ATCTTCAACAATGAAATTTACAGCCTTCCACCTTGGTCAATCAGCATCCTTCCTGATTGC 1140
opt_OsBGal1 ATTTTCAACAACGAGAACTACTCTTTGCCACCTTGGTCAATCTCTATCTTGCCAGATTGT 1140
          I F N N E N Y S L P P W S I S I L P D C

nat_OsBGal1 AAAAATGTTGTTTTTAACACTGCACAGTTGGTGTTCAGACAAATCAAATGCAAATGTGG 1200
opt_OsBGal1 AAGAAACGTTGTTTTTAACACTGCACAGTTGGTGTTCAAACTAACCAAATGCAGATGTGG 1200
          K N V V F N T A T V G V Q T N Q M Q M W

nat_OsBGal1 GCAGACGGGGCTTCTTCAATGATGTGGGACAAGTATGATGAGGAGTTGATTCATTGGCA 1260
opt_OsBGal1 GCTGACGGAGCTTCTTCCATGATGTGGGAAAAGTATGATGAAAGAGTTGACTCATTGGCT 1260
          A D G A S S M M W E K Y D E E V D S L A

nat_OsBGal1 GCTGCTCCATTGCTCACGTCAACTGGTCTACTTGACCAAGCTTAATCTCACAGAGACACC 1320
opt_OsBGal1 GCTGCTCCATTGTTGACTTCTACAGGTTTGTGGAGCAATTGAACGTTACTAGAGATAGA 1320
          A A P L L T S T G L L E Q L N V T R D T

nat_OsBGal1 AGTGATTACCTCTGGTACATTACAGGGTGGAGGTAGACCCATCTGAGAAGTTTCTACAA 1380
opt_OsBGal1 TCTGACTACTTGTGGTATATTACAGAGTTGAAGTGTGATCCTTCCGAGAAATCTTCCAA 1380
          S D Y L W Y I T R V E V D P S E K F L Q

nat_OsBGal1 GGTGGCAGCCTCTGTCACTCACTTGTGAGTCTGCTGGCCATGCCGTGCATGTCTTCAATC 1440
opt_OsBGal1 GGTGGAACCTCCATGTCTTTGACAGTTCAGTCCGCTGGTTCATGCTTGCACGTTTTTATT 1440
          G G T P L S L T V Q S A G H A L H V F I

nat_OsBGal1 AATGGGCAACTCCAAAGTTCTGCCATATGGAACCAGCGAAGATCGGAAAATCTCATATAGT 1500
opt_OsBGal1 AATGGTCAATTCAGGGATCAGCTTACGGTACTAGAGAAGACAGAAAGATCTCTATCTCT 1500
          N G Q L Q G S A Y G T R E D R K I S Y S

nat_OsBGal1 GCCAATGCTAACCTTCCTGCTGGTACAAACAAAGTTGCACCTGTTGACTGTTGCTTGTGGA 1560
opt_OsBGal1 GGTAAATGCTAACCTTGAGAGCTGGACAAACAAAGTTGCTTTGTTGCTGTTGCTTGTGGA 1560
          G N A N L R A G T N K V A L L S V A C G

nat_OsBGal1 CTGCCGAATGTCGCACTGCATTATGACACGTGGAACACTGGTGTGTTGGTCTCTGTTGTG 1620
opt_OsBGal1 TTGCCAAATGTTGCTGTTTATTACGAACTTGGAAACACAGGTGTTGTTGGACCTGTTGTT 1620
          L P N V G V H Y E T W N T G V V G P V V

nat_OsBGal1 ATTCACGGCTTGGACGAAAGGTTCCAGAGATCTGACTTGGCAGACTTGGTCCATATCAATTC 1680
opt_OsBGal1 ATTCACGGATTGGATGAGGGTTCAGAGACTTACTTGGCAACATGGTTCATATCAATTT 1680
          I H G L D E G S R D L T W Q T W S Y Q F

```

Figure 3.40 (Continued) Alignment of the nucleotide sequences of the native (upper) and optimized (lower) OsBGal1 cDNA. Both sequences encode the same 817 amino acid residues.

```

nat_OsBGal1 CAGGTTGGCCTGAAAGGTGAACAGATGAATCTAACTCCTTAGAAGGCTCAGGCTCAGTT 1740
opt_OsBGal1 CAGGTTGGTTTGAAGGGAGAACAGATGAATTTGAACTCCTTGAAGGTTCTGGATCCGTT 1740
          Q V G L K G E Q M N L N S L E G S G S V

nat_OsBGal1 GAATGGATGCAAGCATCATTGGTAGCACAAACCAACAACCGTTGGCATGGTATAGGGCA 1800
opt_OsBGal1 GACTGGATGCAAGGTTTATTGGTAGCTCAGAAATCAACAGCCATTGGCTTGGTACACAGCT 1800
          E W M Q G S L V A Q N Q Q P L A W Y R A

nat_OsBGal1 TACTTTGATACTCCAGTGGTGACGAGCCACTGGCTCTGGATATGGGCAGCATGGGTAAG 1860
opt_OsBGal1 TATTTTGATACTCCTTCAGGTGACGAAACCATGGCTTGGATATGGGATCTATGGGTAAA 1860
          Y F D T P S G D E P L A L D M G S M G K

nat_OsBGal1 GGTCAAATATGGATAAATGGGCAAAGCATTGGACGGTACTGGACAGCATATGCCGAAGGT 1920
opt_OsBGal1 GGCAAATTTGGATCAACGGACAGTCCATTGGTAGTACTGGACTGCTTATGCTGAAGGA 1920
          G Q I W I N G Q S I G R Y W T A Y A E G

nat_OsBGal1 GACTGCAAAGGTTGCCATTACACTGGGTCATACAGGCCACCAGTCTCAGGCAGGTTGT 1980
opt_OsBGal1 GACTGTAAAGGTTGTCATTACACAGGTTCAATAGAGCTCCTAAATGTCAGGCTGGTTGT 1980
          D C K G C H Y T G S Y R A P K C Q A G C

nat_OsBGal1 GGTGAGCCTACACAGCGCTGGTATCATGTGCCAAGATCTGGTTGCAACCAACTAGAAT 2040
opt_OsBGal1 GGCAACCAACTCAGAGATGGTACCAAGTTCCTAGATCTGGTTGCAACCAACAGAAAT 2040
          G Q P T Q R W Y H V P R S W L Q P T R N

nat_OsBGal1 CTGCTACTCGTTTTTGAGGAACCTGGCGGTGATTCCTCAAAGATTGCCCTTGCGAAGCGG 2100
opt_OsBGal1 TTCTTGGTTGTTTTCAAGAGTTGGTGGAGATTCACTAAGATTGCTTTGGCTAAAACA 2100
          L L V V F E E L G G D S S K I A L A K R

nat_OsBGal1 ACAGTCTCAGGTCTCTGTGCTGATGTATCTGAATATCATCCAATATCAAGAAGTGGCAG 2160
opt_OsBGal1 ACTGTTCTGGAGTTGTGCTGACGTTTCCGAATACATCCTAATATTAAGAAGTGGCAA 2160
          T V S G V C A D V S E Y H P N I K N W Q

nat_OsBGal1 ATCGAGAGCTATGGGGAACCAGAGTTCCACACGGCAAGGTCATTTAAAATGTGCACCT 2220
opt_OsBGal1 ATCGAGTCTTATGTTGAACCAGAGTTCCATACGCTAAGGTTCACTTGAATGTGCTCCT 2220
          I E S Y G E P E F H T A K V H L K C A P

nat_OsBGal1 GGCAGACCATTTCGCAATCAAATTTGCTAGCTTTGGGACACCTCTTGGAACTTCGGAA 2280
opt_OsBGal1 GGTCAAACATTTTCCGCTATCAAATTTGCTTCAATTCGGAACACCATGGGTACTTCTGGA 2280
          G Q T I S A I K F A S F G T P L G T C G

nat_OsBGal1 ACATTCAGCAAGGGGAGTGCATTCAAATTAAC TCAAACCTCTGTTCTTGAAGAAGAAATGC 2340
opt_OsBGal1 ACATTTCAACAGGGTGAATGTCACTCTAATTAATCAAACCTCTGTTTGGAGAAGAAATGT 2340
          T F Q Q G E C H S I N S N S V L E K K C

nat_OsBGal1 ATTGGACTACAAAGATGTGTCTGCGCAATCTCTCCAGCAACTTTGGTGGAGATCCCTGC 2400
opt_OsBGal1 ATTGGTTTGCAAAGATGTGTTGTTGCTATCTCCCTTCAAACCTTGGTGGAGATCCATGT 2400
          I G L Q R C V V A I S P S N F G G D P C

nat_OsBGal1 CCCGAGGTCATGAAAAGGGTGGCGTTGAGGCGGTATGCTCTACCGCTGCATAG 2454
opt_OsBGal1 CCTGAACTTATGAACAGAGTTGCTCTTGAGGCTGTTGTTCTACTGCTGCTTAA 2454
          P E V M K R V A V E A V C S T A A *

```

Figure 3.40 (Continued) Alignment of the nucleotide sequences of the native (upper) and optimized (lower) OsBGal1 cDNA. Both sequences encode the same 817 amino acid residues.

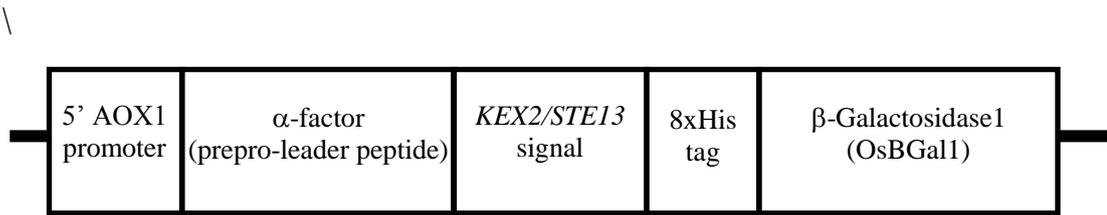


Figure 3.41 Expression cassette of the of the pPICZαBNH₈_OsBGal1 plasmid construct.

The plasmid provides alcohol oxidase 1 (AOX1) promoter-controlled expression of OsBGal1 in *P. pastoris*. OsBGal1 was produced as an N-terminal prepro-α-factor-8xHis-tagged fusion protein, containing the KEX2/STE13 cleavage signal (Glu Lys Arg * Glu Ala * Glu Ala *). The *kex2* gene product cleaves between Arg and Glu and the *ste13* gene product cleaves after the Glu-Ala repeats, with the cleavage site indicated by asterisks in the sequence above.

3.7 Protein purification of optimized OsBGal1

The codon-optimized OsBGal1 cDNA was expressed in the protease-deficient *P. pastoris* strain SMD1168H and protein expression was induced at a temperature of 20°C. High levels of OsBGal1 activity, measured with *p*NPGal, were detected from the codon-optimized OsBGal1 cDNA fusion protein-expressing yeast media. The activity in the media slowly increased until 4 days of induction, after which it decreased (Figure 3.42). This protein was purified from the culture media by IMAC, which bound OsBGal1 containing the His₈-tag. The OsBGal1 protein was eluted from the column with imidazole. The protein band seen in the SDS-PAGE gel in Figure 3.43 is somewhat broad with some higher molecular weight smear, which could be due to hyperglycosylation in the yeast. It has been reported that OsBGal1 has two putative N-glycosylation sites at Asn366 and Asn435 (Chantarangsee et al., 2007).

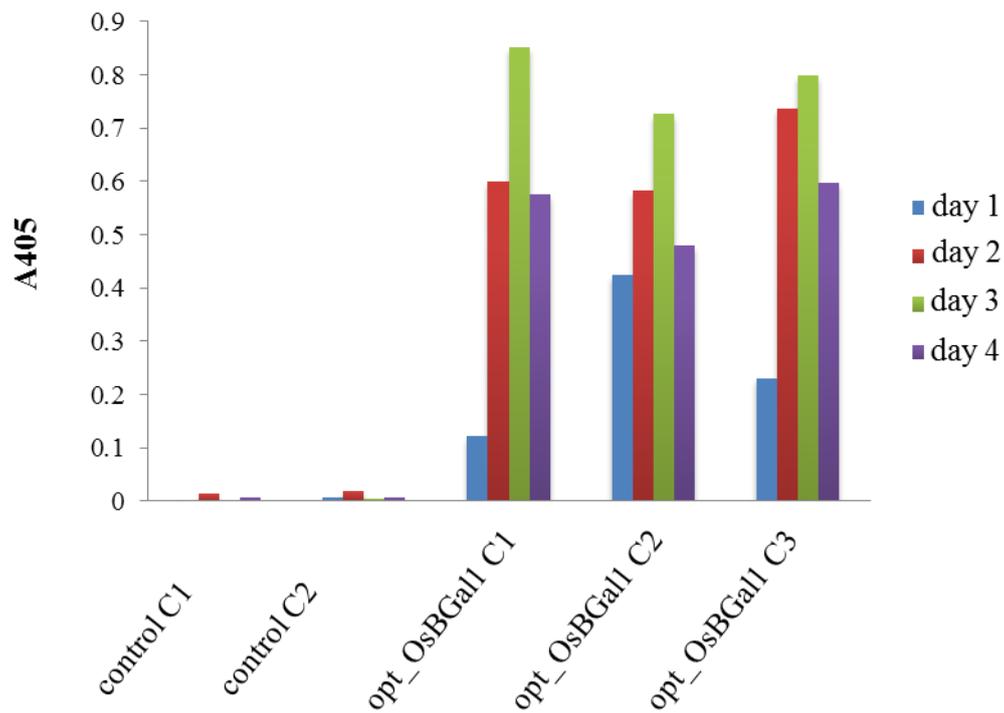


Figure 3.42 Time course of β -galactosidase activity in media from *P. pastoris* clones transformed with empty plasmid control (control C1 and C2), optimized expression vector (opt_OsBGal1 C1, C2 and C3). The β -galactosidase activity in the *P. pastoris* strain SMD1168H media upon induction of expression with 0.5% methanol at 20°C, was measured as the 405 nm absorbance of *p*NP released from *p*NPGal in the standard assay.

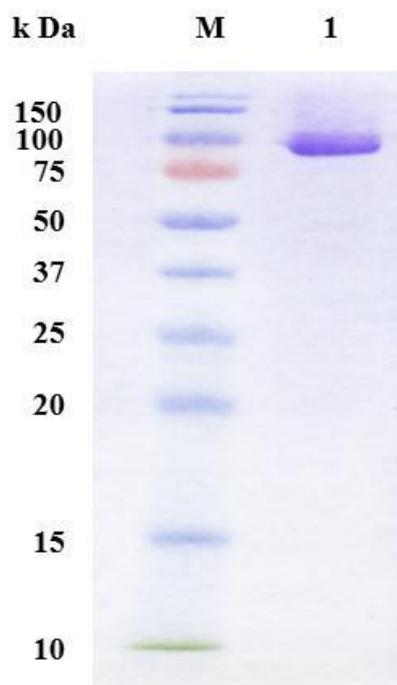


Figure 3.43 SDS-PAGE of optimized OsBGal1 after purification via IMAC. The Coomassie brilliant-blue-stained protein (lane 1) is shown in comparison to a prestained protein marker (lane M).

3.8 Effect of pH and temperature of optimized OsBGal1 enzyme activity

The pH profile of the glycosylated of OsBGal1 for *p*NPGal hydrolysis was determined in universal buffer over the pH range of 3.0-8.0 at 55°C (Figure 3.44). The universal buffer series was used to eliminate the difference of chemical components at each point. The optimum pH for OsBGal1 was found to be 4.5, while the activity of this enzyme was decreased more than 50% at pH 3.5 and 6 (Figure 3.41). The optimum temperature was found at 55°C (Figure 3.45)

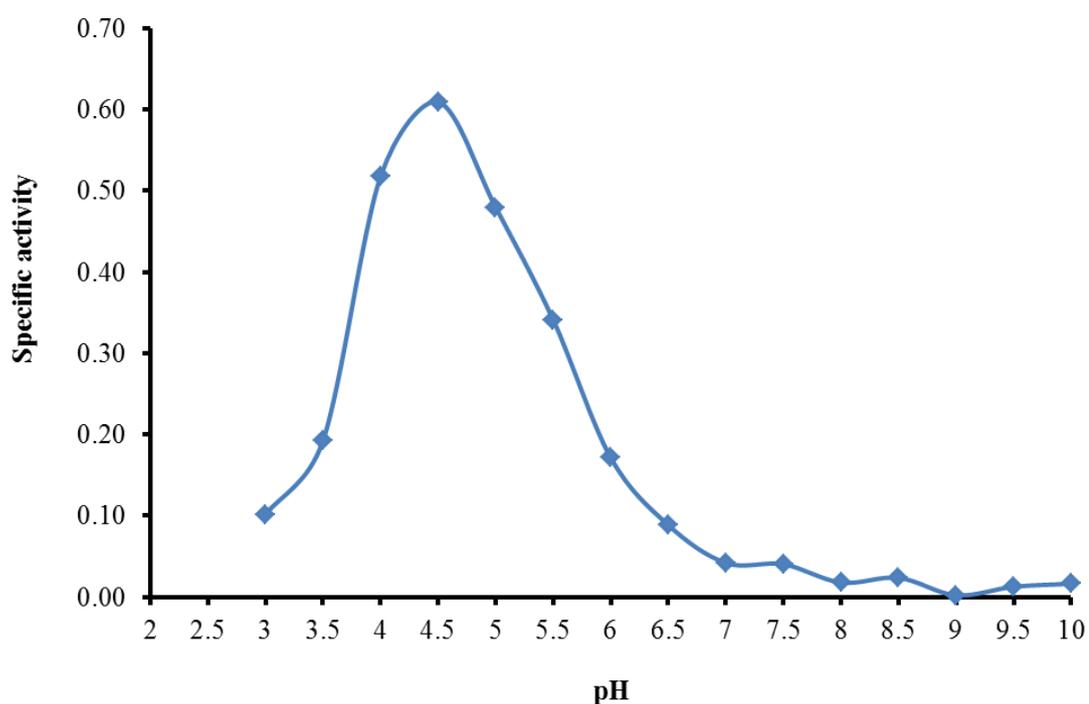


Figure 3.44 Activity versus pH profile of OsBGal1 over a pH range of 3.0-8.0.

OsBGal1 was assayed for hydrolysis of 10 mM *p*NPGal at 55 °C for 30 min

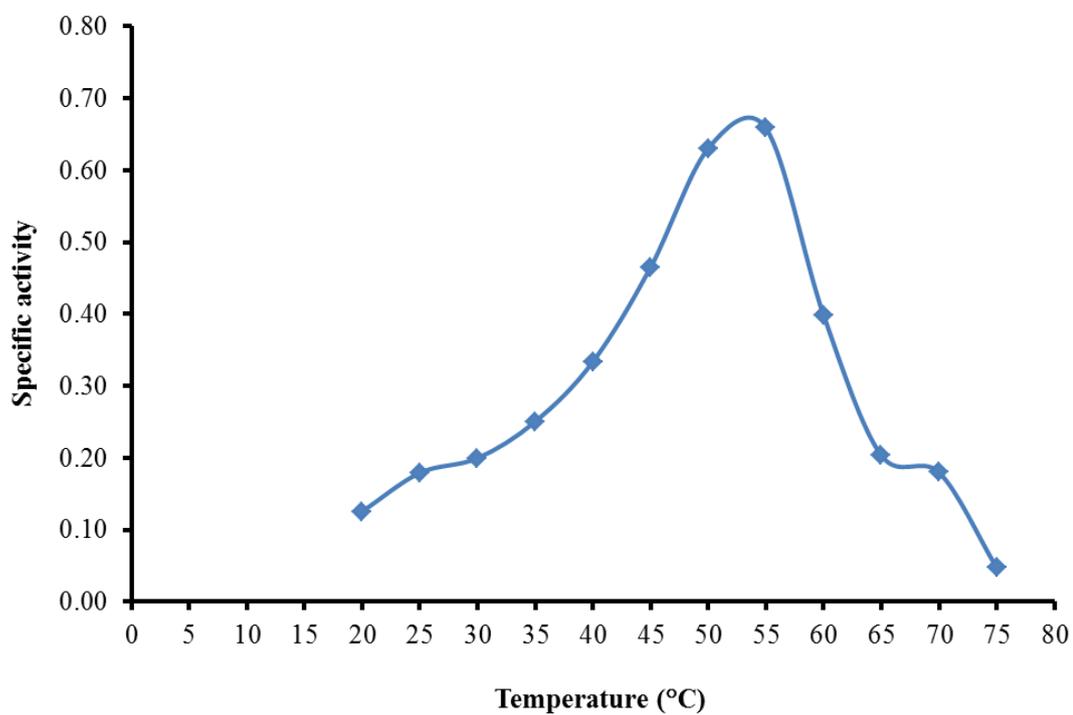


Figure 3.45 Activity profile of OsBGal1 over the temperature range 20-75 °C. OsBGal1 was assayed with 5 mM *p*NPGal in 50 mM sodium acetate, pH 4.5, for 30 min

CHAPTER IV

DISCUSSION

4.1 Protein expression and purification of OsBGal1 β -galactosidase C-terminal domain from rice

When OsBGal1 Cter was produced in *E. coli*, it was expressed at 20°C to slow protein synthesis and thereby lessen the protein aggregation into inclusion bodies. This protein could be expressed in soluble form in the *E. coli* strains Origami(DE3) and Origami B(DE3) as an N-terminal thioredoxin and hexahistidine tag fusion protein. The soluble protein in extracts of Origami B(DE3) was more than in those of Origami (DE3), suggesting that more OsBGal1 Cter protein could fold correctly in Origami B(DE3). Origami(DE3) is a K-12 derivative, while Origami B(DE3) is a BL21 Tuner strain derivative. They have mutations in the glutathione reductase (*gor*) and thioredoxin reductase (*trxB*) genes, which enhance disulfide bond formation in the cytoplasm (Besette et al., 1999). They were reported to yield 10-fold more active protein than in another host, even though overall expression levels were similar (Prinz et al., 1997). Origami(DE3) host cells are congruent with ampicillin resistant plasmids, and with pET vectors, since the λ DE3 prophage contains a T7 polymerase gene with a *lacUV5* promoter, Moreover Origami B(DE3) has a *lacZY* mutant of BL21. The *lacZY* mutation of BL21 allows precise control of expression levels by adjusting the concentration of IPTG. The *trxB* and *gor* mutations are selectable on kanamycin and tetracycline, respectively.

The OsBGal1 protein was purified by three steps. A first step of IMAC was used to eliminate most of the contaminating proteins and some nonspecifically bound proteins were removed with 10 and 20 mM imidazole washes. The fusion protein binds tightly to Co^{2+} resin and came out at 250 mM imidazole wash, suggesting that the His₆ tag is fully exposed to the solvent and the resin. A major protein band at 31 kDa was seen on SDS-PAGE. Approximately 70% pure protein was obtained from this first step of purification by IMAC (Figure 3.8). Since enterokinase gave poor cleavage results, the N-terminal thioredoxin and His₆ tags were cleaved from the fusion protein with thrombin protease and the protein was passed through a second Co^{2+} IMAC column. The OsBGal1 Cter protein from which the N-terminal thioredoxin and His₆ tags had been removed was expected to come out in the unbound fraction. On SDS-PAGE, the free OsBGal1 Cter had molecular weight approximately 17 kDa, while the N-terminal thioredoxin and His₆ tag protein and a small amount of contaminant protein were observed on SDS-PAGE of the 250 mM imidazole elution fractions. The purity of protein from the second Co^{2+} IMAC column was estimated to be approximately 80%. In the next step, free OsBGal1 Cter was purified by Superdex S75 gel filtration chromatography, but this could not eliminate all of the contaminating proteins and degradation of the protein was observed in this step. Although PMSF was added in the initial lysis buffer, soybean trypsin inhibitor was not added, because it proved difficult to remove during the purification in this case. What is more, no inhibitors of proteases other than serine proteases were added. So, evidently protease contaminants which were not removed in the IMAC steps led to OsBGal1 Cter degradation. Moreover, Superdex S75 gel filtration column could not

eliminate contaminant proteins. However, due to the excess tag left on the N-terminus of this version on the protein, further purification was not pursued in this thesis.

4.2 Protein crystallization

The quality of the protein, including the purity and age of the protein preparation are crucial for protein crystallization. For OsBGal1 Cter, no crystals were obtained from this initial protein preparation, perhaps due to the relatively low purity. Many screening kits were used for this experiment and the freshly prepared protein could not produce crystals. This version of the OsBGal1 Cter protein may be difficult to crystallize, due to the purity or the N-terminal extension that still contained the S-tag and enterokinase site from the expression vector, or perhaps not enough screening conditions were tried. Since I had the opportunity to work on the NMR structure, which does not require crystallization, further efforts were concentrated on this front.

4.3 NMR structures of OsBGal1 Cter

4.3.1 Recombinant protein expression and purification for NMR

A new construct from production of the rice β -galactosidase 1 C-terminal domain was generated to allow removal of the long N-terminal linker, and this clone name OsBGal1_Cter_jp_new. This construct also encoded a protein that had two His₆ tags. The recombinant protein that had two His₆ tags might bind to the Co²⁺ IMAC column very tightly. This might help to eliminate contaminant proteins.

As with the protein from the previous construct, the OsBGal1 Cter was purified by IMAC, followed by thrombin cleavage and a second IMAC step.

To protect from proteolysis, every step of purification was done at 4°C. The thioredoxin fusion tag was removed in the second round of IMAC. The OsBGal1 Cter was released from the resin in a 10 mM imidazole wash and contaminating proteins, including a little bit of His₆ tag came off in the later 20 mM imidazole wash. This indicates that the nontagged protein can bind to the resin nonspecifically, or possibly the carbohydrate backbone of the sepharose matrix of the resin, since the protein is homologous to carbohydrate binding proteins. After this step, no protein impurities were obvious on SDS-PAGE. A benzamidine column was used to remove the thrombin protease from the protein, before analysis by NMR.

Finally, >95% pure protein with a minimum yield of 2-4 mg protein/L cell culture was used to determine the structure by NMR. If the protein amount is <2 mg, it is difficult to determine the structure by NMR. During the process of structure determination, we expressed unlabeled protein and labeled protein in the same culture volume, but the yield of protein was not the same. For example, unlabeled protein was expressed more than labeled protein, perhaps due to the use of isotopes of nitrogen and carbon that are rare in nature, which might affect their use in biosynthesis and cell growth. In order to get 4 mg pure protein, more than 1 L of bacterial culture would be needed in the production of labeled protein, but this would require using more stable isotope-labeled media ingredients, which are expensive.

4.3.2 Determination of the native molecular weight of OsBGal1 Cter

Before structure determination, the native molecular weight of OsBGal1 Cter was determined by S75 gel filtration to make sure its native molecular weight was <30 kDa, to allow routine NMR structure determination. In this experiment, the S75 gel filtration column was very long, so it could separate protein peaks very well. If the protein had been a multimer with a native molecular weight more than 30 kDa, its structure could not be determined conveniently by NMR, because there would be many overlapping peaks. The 15 kDa native mass indicated that OsBGal1 C-ter is a monomeric protein, which is different from the RBL domain from chum salmon, which is a dimer (Shirai et al., 2009).

4.3.3 Comparison of the sequence of OsBGal1 Cter with related sequences

The sequence alignment between rice β -galactosidase 1 C-terminal domain from rice to other β -galactosidase isoenzyme indicates highly conserved positions of the eight cysteine residues (Figure 3.21). Moreover, we found the eight cysteine residues in the C-terminal domains of β -galactosidases from many plants, such as strawberry (Trainotti et al., 2001), barley (Triantafillidou et al., 2001) and *Arabidopsis thaliana* (Iglesias et al., 2006). Figure 3.22 compares the sequence between OsBGal1 Cter and those of distantly related proteins with known structures or functions from animal species, including sea urchin and chum salmon lectins and mouse latrophilin 1. It shows the conservation of the positions of the eight cysteine residues in all of these sequences, even though the sequence identity between the plant and animal sequences is only 13-17%.

Lectins have many roles, such as cell signaling, immune response and control of cellular growth (Sharon, 2007). Carbohydrate recognition domains or lectins are classified with over 20 lectin family base on amino acid sequence and functions (Vakonakis et al., 2008). Many roles involve recognition of specific carbohydrate by this domain, such as binding to galactose and rhamnose.

4.3.4. Protein structure determination by NMR.

We have described the first structure of a plant β -galactosidase C-terminal domain. The OsBGal1 Cter was determined by NMR. Nowadays, the NMR technique still has a problem with large proteins. For example if a protein has a molecular weight >30 kDa, it is difficult and more expensive to determine the structure. Moreover, the protein must be soluble at the high concentration required (0.1-2 mM) and should be stable at 25°C for the length of the experiments (up to one week). Luckily, the OsBGal1 Cter is a small protein (13 kDa). In terms of stability, at 25°C unlabeled protein was degraded in 1 week (data not shown), single labeled (^{13}C or ^{15}N) protein was generally degraded after approximately 2 weeks and double labeled ($^{13}\text{C}/^{15}\text{N}$) protein was very stable at 25°C for around 1 month. So, the properties of the OsBGal1 Cter were adequate to investigate structure in NMR, although some proteins had to be prepared repeatedly due to degradation.

The structure of OsBGal1 Cter was compared with known carbohydrate binding domain structures from animal proteins that show the same general fold and distant sequence homology. The core beta sandwich fold was similar to the known structures of chum salmon egg lectin and the sugar-binding domains of mouse latrophilin 1. The main structural differences were in three loops (A, C and E).

In particular, loop A is very flexible and approximately five times longer than the corresponding loops in mouse latrophilin 1 and chum salmon egg lectin.

4.3.5 Relationship of structure to function

The carbohydrate recognition of chum salmon lectin (egg) and mouse latrophilin 1 involves interaction of the sugars with amino acid residues in loop 2. This loop's interaction with rhamnose and galactose are similar to carbohydrate binding with other proteins, such as C-type lectin, which suggests that loop 2 is responsible for binding (Zelensky and Gresdy, 2005). Although, the dissociation constant (K_d) of rhamnose binding is approximately 100-200 μM by sea urchin (Hosono et al., 1999), which indicates tighter binding than mouse latrophilin 1 ($K_d = 1.8 \text{ mM}$) (Vakonakis et al., 2008). In fact, rhamnose is found rarely in animals (Tymiak et al., 1993). This monosaccharide has no evident reason to bind to latrophilins, because no biosynthetic pathway for it is found in mice. Although loop D of the OsBGal1 Cter is related to loop 2 in RBL from animals and contains the lysine residue involved in sugar binding in those proteins, OsBGal1 Cter does not have the residues from other loops that help with sugar binding and could not bind to galactose and rhamnose. This loop is not very flexible, but we could not find any monosaccharide ligand bound to it in the HSQC binding experiments.

The structures of RBL in complexes with D-galactose and L-rhamnose show that Glu42 and Lys120 of latrophilin-1 and Glu7 and Lys86 of chum salmon lectin have important interactions with the sugars (Vakonakis et al., 2008 and Shirai et al., 2009). When Glu42/Glu7 and Lys120/Lys86 in loop2 change to another amino acid, the K_d values decrease (Vakonakis et al., 2008). For example, when E42 was

substituted with aspartate (D) or glutamine (Q), show that the K_d increased from 1.8 mM to 144 mM and 142 mM, respectively. Moreover, when K120 was changed to arginine (R) or alanine (A), binding was not detected.

For OsBGal1 Cter, the expected monosaccharides did not seem to bind in the OsBGal1 Cter in the HSQC experiment, so binding of oligosaccharides and polysaccharides was screened by carbohydrate microarray. We found candidate oligosaccharides and polysaccharides and went on to test binding by STD NMR. However, the STD NMR did not show any signal for binding for α -(1,5)-L-arabinotriose, despite trying to saturate the protein at several frequencies, in order to ensure transfer of energy to any bound ligand. There may have been a problem with nonspecific binding in the carbohydrate microarray, since the signal was not that strong. Further STD-NMR experiments with β -(1,4)-galactopyranobiose will be done in the future to verify whether this can bind to the C-term domain or not

4.4 Expression of full length OsBGal1 β -galactosidase in *Pichia*

4.4.1 Sequence analysis of native and optimized OsBGal1 cDNA

The native OsBGal1 cDNA sequence was analyzed for *Escherichia coli* and *Pichia pastoris* codon usage by the online facility at Genscript Corp (<http://www.genscript.com>). The Codon Adaptation Index (CAI), a simple measure of synonymous codon bias was used to measure codon bias patterns. Comparing the codon preference of the expression hosts (*E. coli* and *P. pastoris*) with the codon usage in the *OsBGal1* gene by the CAI can be used to predict whether it will affect the gene expression level. A CAI of 0.8 is considered to be ideal, so if the CAI value of the gene of interest is close to 0.8, the level of gene expression should be high.

The native OsBGal1 cDNA had a CAI value of 0.64 for *E. coli* and 0.67 for *P. pastoris*, which means it was likely to be poorly expressed in these hosts. The native OsBGal1 had several codons which are rarely used in *E. coli*, such as CTA (Leu), ATA (Met), TCT/C (Ser), ACA (Thr) and AGA/G (Arg), as shown in Table 5.1, while the codons rarely used in *P. pastoris* include CTC (Leu), GTA (Val), CCG (Pro), ACG (Thr), GCG (Ala), GGG (Gly) and AGC (Ser), as shown in Table 5.2 (<http://www.kazusa.or.jp/codon/>). To improve the expression level, the OsBGal1 gene was optimized for *P. pastoris* by GenScript Corp (Piscataway, NJ, USA). The sequence differed from the native cDNA at 234 nucleotide positions and had a G+C content of 41.5%, as shown in Figure 4.1.

Table 4.1 The codon usage of *Escherichai coli*. Genetic codons are indicated as three capital characters, the frequencies of triplet codon per thousand codons are indicated by the numbers next to the codon and the codon counts out of 5,122 codons are indicated as the number in brackets (<http://www.kazusa.or.jp/codon/>).

TTT 19.7 (101)	Phe	TCT 5.7 (29)	Ser	TAT 16.8 (86)	Tyr	TGT 5.9 (30)	Cys
TTC 15.0 (77)		TCC 5.5 (28)		TAC 14.6 (75)		TGC 8.0 (41)	
TTA 15.2 (78)	Leu	TCA 7.8 (40)	Pro	TAA 1.8 (9)	Stop	TGA 1.0 (5)	Stop
TTG 11.9 (61)		TCG 8.0 (41)		TAG 0.0 (0)		TGG 10.7 (55)	
CTT 11.9 (61)	Leu	CCT 8.4 (43)	Pro	CAT 15.8 (81)	His	CGT 21.1 (108)	Arg
CTC 10.5 (54)		CCC 6.4 (33)		CAC 13.1 (67)		CGC 26.0 (133)	
CTA 5.3 (27)	Leu	CCA 6.6 (34)	Pro	CAA 12.1 (62)	Gln	CGA 4.3 (22)	Arg
CTG 46.9 (240)		CCG 26.7 (137)		CAG 27.7 (142)		CGG 4.1 (21)	
ATT 30.5 (156)	Ile	ACT 8.0 (41)	Thr	AAT 21.9 (112)	Asn	AGT 7.2 (37)	Ser
ATC 18.2 (93)		ACC 22.8 (117)		AAC 24.4 (125)		AGC 16.6 (85)	
ATA 3.7 (19)	Met	ACA 6.4 (33)	Thr	AAA 33.2 (170)	Lys	AGA 1.4 (7)	Arg
ATG 24.8 (127)		ACG 11.5 (59)		AAG 12.1 (62)		AGG 1.6 (8)	
GTT 16.8 (86)	Val	GCT 10.7 (55)	Ala	GAT 37.9 (194)	Asp	GGT 21.3 (109)	Gly
GTC 11.7 (60)		GCC 31.6 (162)		GAC 20.5 (105)		GGC 33.4 (171)	
GTA 11.5 (59)	Val	GCA 21.1 (108)	Ala	GAA 43.7 (224)	Glu	GGA 9.2 (47)	Gly
GTG 26.4 (135)		GCG 38.5 (197)		GAG 18.4 (94)		GGG 8.6 (44)	

Table 4.2 The codon usage of *Pichai pastoris*. Genetic codons are indicated as three capital letters, the frequencies of triplet codon per thousand codons are indicated by the numbers next to the codon and the codon counts out of 81,301 codons (from 137 proteins) are indicated as the number in brackets (<http://www.kazusa.or.jp/codon/>).

TTT 24.1(1963)	Phe	TCT 24.4(1983)	TAT 16.0(1300)	TGT 7.7(626)
TTC 20.6(1675)		TCC 16.5(1344)	TAC 18.1(1473)	TGC 4.4(356)
TTA 15.6(1265)	Leu	TCA 15.2(1234)	TAA 0.8(69)	TGA 0.3(27)
TTG 31.5(2562)		TCG 7.4(598)	TAG 0.5(40)	TGG 10.3(834)
CTT 15.9(1289)	Leu	CCT 15.8(1282)	CAT 11.8(960)	CGT 6.9(564)
CTC 7.6(620)		CCC 6.8(553)	CAC 9.1(737)	CGC 2.2(175)
CTA 10.7(837)		CCA 18.9(1540)	CAA 25.4(2069)	CGA 4.2(340)
CTG 14.9(1215)		CCG 3.9(320)	CAG 16.3(1323)	AGG 1.9(158)
ATT 31.1(2532)	Ile	ACT 22.4(1820)	AAT 25.1(2038)	AGT 12.5(1020)
ATC 19.4(1580)		ACC 14.5(1175)	AAC 26.7(2168)	AGC 7.6(621)
ATA 11.1(906)	Met	ACA 13.8(1118)	AAA 29.9(2433)	AGA 20.1(1634)
ATG 18.7(1517)		ACG 6.0(491)	AAG 33.8(2748)	AGG 6.6(539)
GTT 26.9(2188)	Val	GCT 28.9(2351)	GAT 35.7(2899)	GGT 25.5(2075)
GTC 14.9(1210)		GCC 16.6(1348)	GAC 25.9(2103)	GGC 8.1(655)
GTA 9.9(804)		GCA 15.1(1228)	GAA 37.4(3043)	GGA 19.1(1550)
GTG 12.3(998)		GCG 3.9(314)	GAG 29.0(2360)	GGG 5.8(468)

4.5 Development of an expression system for active OsBGal1

Bacteria and yeast are common host cells used for production of interesting proteins. Even though bacterial expression is convenient, many eukaryotic proteins are produced in low protein yields (Batas et al., 1999; Patra et al., 2000; Oagnaesyant et al., 2005). OsBGal1 could be expressed in bacteria, but its expression was not very high (Chantarangsee et al., 2007). Therefore, a yeast expression system was introduced that was reported to be able to produce plant enzymes in active forms (Ketudat Cairns et al., 2000). The OsBGal1 cDNA amplified from rice seedling shoots produced low protein yields in *E. coli* (Chantarangsee et al., 2007) and *P. pastoris* (Chantarangsee and Ketudat Cairns, unpublished data).

The OsBGal1 β -galactosidase was successfully expressed from a codon-optimized cDNA in a protease-deficient strain of *P. pastoris* at a low temperature of 20°C, following the methodology used by Luang et al (2010) for barley ExoI exoglucanase. In the yeast expression system, the OsBGal1 was secreted out of the cells and the protein recovered from the media. We tried to precipitate protein with ammonium sulfate to allow further IMAC purification. After redissolving the protein, no β -galactosidase activity was detectable, suggesting that the conformation of protein changed. In fact, previous attempts to purify OsBGal1 from *E. coli* in procedures that used high salt resulted in loss of protein activity (Chantarangsee, unpublished).

These results indicate that this enzyme is sensitive to high salt. We solved this problem by adjusting the pH of the media to 7.5 with dipotassium hydrogen phosphate before purification by IMAC. The OsBGal1 could be purified by a single IMAC step. This enzyme is very sensitive to salt and cold temperature. For example,

if the enzyme was precipitated with ammonium sulfate or frozen, no activity was detectable. The OsBGal1 expressed in *E. coli* was similar and was kept at 4°C (Chantarangsee et al., 2007). So, we had to complete purification in a short time and every step must be done at 4°C.

4.6 Catalytic properties of OsBGal1

The pH optimum and temperature of the OsBGal1 were similar to those previously reported for OsBGal1 produced in *E. coli*. The calculated molecular mass of the OsBGal1 polypeptide was 90 kDa (<http://www.expasy.org>). Based on the sizes of N-linked carbohydrates at the N-glycosylation sites, the presence of these carbohydrates might affect the enzyme properties, such as pH optimum and temperature optimum. However, when the OsBGal1 protein was deglycosylated by EndoH at pH 5.5 for 1 day, the protein degraded. Protease inhibitor (PMSF) was added in the solution before purification by IMAC, but the enzyme could not tolerate being kept at pH 5.5 for 1 day, suggesting proteases that were not inhibited by PMSF were present and the protein did not exhibit stability in the presence of these proteases. The pH optimum determination showed that the OsBGal1 produced in *P. pastoris* had a pH optimum at 4.5, and lost approximately 50% activity at pH 5.5. This pH optimum is higher than those of 3.2 for β -galactosidase from mango and 3.6 for β -galactosidase from mung bean seedling (Li et al., 2001 and Ali et al., 1995). Nonetheless, the result suggests that this enzyme may function in an acidic environment. The optimum temperature of optimized OsBGal1 Cter protein was found to be 55°C. In comparison, the optimum temperatures of β -galactosidases from peach (*Prunus persica*) (Lee et al., 2003) and apricots (*Prunus armeniaca kaisa*)

(Ali et al., 1995) were 50°C and 40°C, respectively. These values are similar to those of β -galactosidases from other fruits.

CHAPTER V

CONCLUSION

The rice β -galactosidase 1 C-terminal domain was initially expressed as a fusion protein with N-terminal thioredoxin and hexahistidine (His₆) tags, which were designed to be removed with enterokinase, in *E. coli* Origami B(DE3) and purified for structural studies. The 31 kDa thioredoxin fusion protein was purified by IMAC. Because enterokinase digest was incomplete and gave extra protein bands, the thioredoxin fusion tag was removed by thrombin digestion followed by second round of IMAC, which left an S-tag and enterokinase site on the N-terminus of the OsBGal1 Cter protein. With this initial expression construct, the purity of OsBGal1 Cter was approximately 80%, and further purification by S75 gel filtration could remove only a fraction of the contaminating protein. When this protein was used for crystallization, no protein crystals were observed in any condition.

A new construct, pET32a/OsBGal1 Cter_jp_new was designed to place a thrombin site up against the OsBGal1 Cter sequence in order to remove nearly all of the fusion tag by thrombin digest. After expression from this construct, a 33 kDa thioredoxin-OsBGal1-Cter fusion protein was purified by IMAC. The OsBGal1 Cter was completely cleaved from the fusion protein with thrombin protease within 2 h at 4°C. The thioredoxin fusion tag was removed by thrombin cleavage followed by second round of IMAC, and removal of the thrombin protease from the sample by benzamidine column chromatography. Approximately 0.15-1 mmol of 13 kDa

OsBGal1 Cter protein was obtained per liter of cell culture, with >95% purity. The OsBGal1 Cter was a monomer ie. protein and had an apparent native molecular weight of approximately 15 kDa, as determined by S75 gel filtration chromatography.

The OsBGal1 Cter comprises 118 amino acid residues, including 8 cysteine residues. A comparison of the OsBGal1 Cter sequence to those of rhamnose-binding lectins (RBL) and related domains from animal species, such as sea urchin egg lectin, mouse latrophilin and chum salmon (egg) lectin, shows that the positions of the eight cysteine residues are highly conserved in all sequences. Despite this, the percent identities between the OsBGal1 Cter sequence and the animal sequences are very low, ranging from 13-17%.

To determine its structure by NMR, the recombinant OsBGal1 Cter was expressed with and without labeling with ^{15}N , ^{13}C or $^{15}\text{N}/^{13}\text{C}$. The backbone assignments of OsBGal1 Cter were constructed from 3D HNC(O), CBCA(CO)NH and HNCACB NMR spectra. Side chain peaks for the OsBGal1 Cter were assigned from C(CO)NH and HCCH-TOCSY spectra. The OsBGal1 Cter structure is composed of five β -strands, β_a to β_e (residues 731-735, 758-763, 771-777, 811-814, and 833-840, respectively) and a single, short α -helix (residues 798-805). The overall fold of the OsBGal1 Cter can be described as a β sandwich with two antiparallel sheets (one composed of β -strands β_a , β_e , and β_c , and the other of β_b and β_d). Moreover, this structure includes 5 loops connecting strands β_a and β_b (residues 736-757, loop A), strands β_b with β_c (residues 764-770, loop B), strand β_c and the α -helix (residues 778-797, loop C), the α -helix and strand β_d (residues 806-810, loop D) and strand β_d with strand β_e (residues 816-832, loop E).

Four disulfide bridges are found in the OsBGal1 Cter structural core, interconnecting strand β a and loop C (Cys732–Cys785), loop B with strand β e (Cys764–Cys839), loop C and loop E (Cys793–Cys826), and loop D with strand β d (Cys806–Cys812). Loop A (residues 735-763) is very flexible, based on the fact that it was poorly defined by the NMR data, with few long-range restraints in this region. Because OsBGal1Cter is homologous to galactose/rhamnose-binding lectins and is an accessory domain from a β -galactosidase that could act in carbohydrate binding, its binding to galactose and rhamnose was tested. Titration with L-rhamnose, D-galactose, D-glucose and raffinose showed no shifts in the ^1H - ^{15}N HSQC NMR spectra, suggesting that OsBGal Cter did not bind these sugars. The C-terminal domain of β -galactosidase 1 from rice and carbohydrate binding sites of the animal RBL were compared. The comparison showed that the structure of OsBGal1 Cter is quite similar to those of the RBL, but the residues involved in sugar binding are not conserved.

A carbohydrate array was used to assess other possible binding partners for OsBGal1 Cter. The OsBGal1Cter appeared to bind to some oligosaccharides and polysaccharides on the carbohydrate array, including 1,5- α -L-arabinotriose. However, STD NMR experiments showed that the OsBGal1 Cter could not bind to 1,5- α -L-arabinotriose under the conditions tested.

To improve the expression of the full-length OsBGal1 β -galactosidase for structural and functional studies, a codon-optimized gene was used to express the protein in *Pichia pastoris*. The OsBGal1 protein expressed from the optimized cDNA was secreted into the *P. pastoris* media. The enzyme was purified by IMAC and characterized. The pH profile of the glycosylated OsBGal1 for pNPGal hydrolysis was determined in universal buffer over the pH range of 3.0-8.0 at 55°C. The optimum pH

for OsBGal1 was found to be 4.5, while the activity of this enzyme was decreased more than 50% at pH 2 and 6. The optimum temperature was at 55°C.

REFERENCES

REFERENCES

- Abe, A., Tonozuka, T., Sakano, Y., and Kamitori, S. (2004). Complex structures of *Thermoactinomyces vulgaris* R-47 alpha-amylase 1 with malto-oligosaccharides demonstrate the role of domain N acting as a starch-binding domain. **J. Mol. Biol.** 335: 811-822.
- Ali, Z.M., Armugam, S., and Lazan, H. (1995). β -Galactosidase and its significance in ripening mango fruit. **Phyto-chemistry.** 38: 1109-1114.
- Benie, A.J., Moser, R., Bäuml, E., Blaas, D., and Peters, T. (2003). Virus-ligand interactions: identification and characterization of ligand binding by NMR Spectroscopy. **J. Am. Chem. Soc.** 125: 14-15
- Bessette, P.H., Aslund, F., Beckwith, J., and Georgiou, G. (1999). Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. **Proc. Natl. Acad. Sci. USA.** 96: 13703-13708.
- Blanchard, J.E., Gal, L., He, S., Foisy, J., Warren, R.A., and Withers, S.J. (2001). The identification of catalytic nucleophiles of two β -galactosidases from glycoside hydrolase family 35, **Carbohydrate Res.** 333: 7-17.
- Boel, E., Brady, L., Brzozowski, A.M., Derewenda, Z., Dodson, G.G., Jensen, V.J., Peterson, S.B., Swift, H., Thim, L., and Woldike, H.F. (1990). Calcium binding in α -amylase: an X-ray diffraction study at 1.2 Å resolution of two enzymes from *Aspergillus*. **Biochemistry.** 9: 6244-6249.

- Bodenhausen, G. and Ruben, D.J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. **Chemical Physics Letters**. 69(1): 185-189.
- Bolam, D.N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M.P., Rixon, J. E., Boraston, A., Hazlewood, G.P., and Gilbert, H.J. (1998). *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. **Biochem. J.** 331: 75-781.
- 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.
- Boraston, A.B., Bolam, D.N., Gilbert, H.J., and Davies, G.J. (2004). Carbohydrate binding modules: fine-tuning polysaccharide recognition. **Biochem. J.** 382: 769-781.
- Boraston, A.B., Ficko-Blean, E., and Healey, M. (2007). Carbohydrate recognition by a large sialidase toxin from *Clostridium perfringens*. **Biochemistry**. 46: 11352-11360.
- Boraston, A. B., Nurizzo, D., Notenboom, V., Ducros, V., Rose, D.R., Kilburn, D.G., and Davies, G.J. (2002). Differential oligosaccharide recognition by evolutionarily related beta-1,4 and beta-1,3 glucan-binding modules. **J. Mol. Biol.** 319: 1143-1156.
- Brett, C.T. and Waldron, K.W. (1996). **Physiology and Biochemistry of Plant Cell Wall**, 2nd ed. London: Chapman and Hall.

- Callahan, J.W. (1999). Molecular basis of GM1 gangliosidosis and Morquio disease, type B. Structure-function studies of lysosomal β -galactosidase and the non-lysosomal β -galactosidase-like protein. **Biochem. Biophys. Acta.** 1455: 85-103.
- Canales, A., Matesanz, R., Gardner, N., Andreu, J. M., Paterson, I., Diaz, F., and Jimenez-Barbero, J. (2008). The bound conformation of microtubule-stabilizing agents: NMR insights into the bioactive 3D structure of discodermolide and dictyostain. **Chemistry.** 14: 7569-7575
- Cantarel, B.L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. **Nucleic Acids Res.** 37: D233-D238.
- Carey, A.T., Holt, K., Picard, S., Wilde, R., Tucker, G.A., Bird, C.R., Schuch, W., and Seymour, G.B. (1995). Tomato exo-(1-4)- β -D-galactanase. Isolation, changes during ripening in normal and mutant tomato fruit, and characterization of a related cDNA clone. **Plant Physiol.** 108: 1099-1107.
- Chang, T.W. (1983). Binding of cells to matrixes of distinct antibodies coated on solid surface. **J. Immunol. Methods.** 65(1-2): 217-23.
- Chantarangsee, M., Tanthanuch, W., Fujimura, T., Fry, S.C., and Ketudat Cairns, J.R. (2007). Molecular characterization of β -galactosidases from germinating rice (*Oryza sativa*). **Plant Sci.** 173: 118-134.
- Claasen, B., Axmann, M., Meinecke, R., and Meyer, B. (2005). Direct observation of ligand binding to membrane proteins in living cells by a saturation transfer double difference (STDD) NMR spectroscopy method shows a significantly

- higher affinity of integrin $\alpha(\text{IIb})\beta_3$ in native platelets than in liposomes. **J. Am. Chem. Soc.** 127(3): 916-919.
- Cohn, M. and Monod, J. (1951). Purification and properties of the β -galactosidase (lactase) of *Escherichia coli*. **Biochim Biophys Acta.** 7: 153.
- Coutinho, P.M. and Henrissat, B. (1999). Carbohydrate-active enzymes: an integrated database approach. In "Recent Advances in Carbohydrate Bioengineering", H.J. Gilbert, G. Davies, B. Henrissat and B. Svensson eds., pp. 3-12, The Royal Society of Chemistry, Cambridge, UK.
- Durand, P., Lehn, P., Callebaut, I., Fabrega, S., Henrissat, B., and Moron, J.P. (1997). Active-site motifs of lysosomal acid hydrolases: invariant features of clan GH-A glycosyl hydrolases deduced from hydrophobic cluster analysis. **Glycobiology.** 7: 277-284.
- Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. **Structure.** 3: 853-859.
- Fujimoto, Z., Kuno, A., Kaneko, S., Kobayashi, H., Kusakabe, I., and Mizuno, H. (2002). Crystal structures of the sugar complexes of *Streptomyces olivaceoviridis* E-86 xylanase: sugar binding structure of the family 13 carbohydrate binding module. **J. Mol. Biol.** 316: 65-78.
- Gasparini, F., Franchi, N., Spolaore, B., and Ballarin, L. (2008). Novel rhamnose-binding lectins from the colonial ascidian *Botryllus schlosseri*. **Dev. Comp. Immunol.** 16: 944-953.
- Giannakouros, T., Karagiorgos, A., and Simos, G. (1991). Expression of β -galactosidase multiple forms during barley (*Hordeum vulgare*) seed

- germination. Separation and characterization of enzyme isoforms, **Physiol. Plant.** 82: 413-418.
- Gilkes, N.R., Warren, R.A., Miller, R.C. Jr., and Kilburn, D.G. (1988). Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. **J. Biol. Chem.** 263: 10401-10407.
- Gurdev, S. K. and Grary, H. Y. (1991). **Rice Biotechnology.** The International Rice Research Institute, UK. pp. 1-2.
- Husain, Q. (2010). β -Galactosidases and their potential applications: a review. **Biotechnology.** 30(1): 41-62.
- Hashimoto, H. (2006). Recent structural studies of carbohydrate-binding modules. **Cell Mol Life Sci.** 63: 2954-2967.
- Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. **Biochem. J.** 280: 309-316.
- 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.
- Herfurth, L., Ernst, B., Wagner, B., Ricklin, D., Strasser, D. S., Magnani, J. L., Benie, A. J., and Peters, T. (2005). Comparative epitope mapping with saturation transfer difference NMR of sialyl Lewis(a) compounds and derivatives bound to a monoclonal antibody. **J. Med. Chem.** 48: 6879-6886.

- Hidaka, M., Fushinobu, S., Ohtsu, N., Motoshima, H., Matsuzawa, H., Shoun, H., and Wakagi, T. (2002). Trimeric crystal structure of the glycoside hydrolase family 42 beta-galactosidase from *Thermus thermophilus* A4 and the structure of its complex with galactose. **J. Mol. Biol.** 322: 79-91.
- Hosono, M., Kawauchi, H., Nitta, K., Takatanagi, Y., Shiokawa, H., Mineki, R., and Muramoto, K. (1993). Three rhamnose-binding lectins from *Osmerus eperlanus mordax* (Olive rainbow smelt) roe. **Biol. Pharm. Bull.** 16, 239-243.
- Ito, Y., Tomita, T., Roy, N., Nakano, A., Sugawara-Tomita, N., Watanabe, S., Okai, N., Abe, N., and Kamio, Y. (2003). Cloning, expression, and cell surface localization of *Paenibacillus* sp. strain W-61 xylanase 5, a multidomain xylanase. **Appl. Environ. Microbiol.** 69: 6969-6978.
- Jacobson, R.H., Zhang, X.J., DuBose, R.F., and Matthews, B.W. (1994). Three-dimensional structure of β -galactosidase from *E. coli*. **Nature.** 369: 761-766.
- Jenkins, J., Lo Leggio, 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.
- Juers, D.H., Jacobson, R.H., Wigley, D., Zhang, X.J., Huber, R.E., Tronrud, D. E., and Matthews, B.W. (2000). High resolution refinement of β -galactosidase in a new crystal form reveals multiple metal-binding sites and provides a structural basis for alpha-complementation. **Protein Sci.** 9: 1685-1699.

- Juers, D.H., Huber, R.E., and Matthews, B.W. (1999). Structural comparisons of TIM barrel proteins suggest functional and evolutionary relationships between β -galactosidase and other glycohydrolases. **Protein Sci.** 8: 122-136.
- Juge, N., Gal-Coeffet, M.L., Furniss, C., Gunning, A., Kramhoft, B., Morris, V.J., Williamson, G., and Svensson, B. (2002). The starch binding domain of glucoamylase from *Aspergillus niger*: overview of its structure, function, and role in raw-starch hydrolysis. **Biologia Bratislava** 57: 239-245.
- Kaneko, S. and Kobayashi, H. (2003). Purification and characterization of extracellular β -galactosidase secreted by suspension cultured rice (*Oryza sativa* L.) cells. **Biosci. Biotechnol. Biochem.** 67: 627-630.
- Kawazu, T., Nakanishi, Y., Uozumi, N., Sasaki, T., Yamagata, H., Tsukagoshi, N., and Udaka, S. (1987). Cloning and nucleotide sequence of the gene coding for enzymatically active fragments of the *Bacillus polymyxa* beta-amylase. **J. Bacteriol.** 169: 1564-1570.
- Konno, H. and Katoh, K. (1992). An extracellular β -galactosidase secreted from cell suspension cultures of carrot. Its purification and involvement in cell wall polysaccharide hydrolysis. **Physiol. Plant.** 85: 507-514.
- Konno, H. and Tsumuki, H. (1993). Purification of β -galactosidase secreted from rice shoots and its involvement in hydrolysis of the natural substrate in cell walls. **Physiol. Plant.** 89: 40-47.
- Koshland, D.E.J. (1953). Stereochemistry and the mechanism of enzymatic reactions. **Biol. Rev.** 28: 416-436.
- Kotake, T., Dina, S., Konishi, T., Kaneko, S., Igarashi, K., Samejima, M., Watanabe, Y., Kimura, K., and Tsuraya, Y. (2005). Molecular cloning of a

- β -galactosidase from radish that specifically hydrolyzes β -(1 \rightarrow 3) and β -(1 \rightarrow 6)-galactosyl residues of arabinogalactan protein. **Plant Physiol.** 182: 1563-1576.
- Lee, D.H., Kang, S.G., Suh, S.G., and Byun, J.K. (2003). Purification and Characterization of a β -Galactosidase from Peach (*Prunus persica*) **Molecules and Cells.** 15: 68-74.
- Li, S.C., Han, J.W., Chen, K.C., and Chen, C.S. (2001). Purification and characterization of isoforms of β -galactosidases in mung bean seedlings, **Phytochemistry.** 57: 349-359.
- Ly, H.D. and Withers, S.G. (1999). Mutagenesis of glycosidases. **Annu. Rev. Biochem.** 68: 487-522.
- Mari, S., Serrano-Gómez, D., Cañada, F.J., Corbí, A.L., and Jiménez-Barbero, J. (2005). 1D-STD NMR experiments on living cells. The DC-SIGN/oligomannose interaction, **Angew. Chem. Int. Ed.** 44(2): 296-298.
- McCarter, J. and Withers, S.G. (1994). Mechanisms of enzymatic glycoside hydrolysis. **Curr. Opin. Struct. Biol.** 4: 885-892.
- McLean, B.W., Bray, M.R., Boraston, A.B., Gilkes, N.R., Haynes, C.A., and Kilburn, D.G. (2000). Analysis of binding of the family 2a carbohydrate-binding module from *Cellulomonas fimi* xylanase 10A to cellulose: specificity and identification of functionally important amino acid residues. **Protein Eng.** 13: 801-809.
- Mikami, B., Sato, M., Shibata, T., Hirose, M., Aibara, S., Katsube, Y., and Marita, Y. (1992). Three dimensional structure of soybean α -amylase determined at 3.0 Å resolution-Preliminary chain tracing of the complex with α -cyclodextrin. **J. Biochem.** 112: 541-546.

- Naganuma, T., Ogawa, T., Hirabayashi, J., Kasai, K., Kamiya, H., and Muramoto, K. (2006). Isolation, characterization and molecular evolution of a novel pearl shell lectin from a marine bivalve, *Pteria penguin*. **Mol. Divers.** 10: 607-618.
- Notenboom, V., Boraston, A.B., Chiu, P., Freelove, A.C.J., Kilburn, D.G., and Rose, D.R. (2001). Recognition of cello-oligosaccharides by a family 17 carbohydrate-binding module: an X-ray crystallographic, thermodynamic and mutagenic study. **J. Mol. Biol.** 314: 797-806.
- Oberli, M.A., Tamborrini, M., Tsai, Y.H., Werz, D.B., Horlacher, T., Adibekian, A., Gauss, D., Moller, H.M., Pluschke, G., and Seeberger, P.H. (2010). Molecular analysis of carbohydrate-antibody interactions: case study using a *Bacillus anthracis* tetrasaccharide. **J. Am. Chem. Soc.** 132: 10239-10241
- O'Donoghue, E.M., Somerfield, S.D., Sinclair, B.K., and King, G.A. (1998). Characterization of the harvest-induced expression of β -galactosidase in *Asparagus officinalis*. **Plant Physiol.** 10: 721-729.
- 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: 38.
- Ozeki, Y., Matsui, T., Suzuki, M., and Titani, K. (1991). Amino acid sequence and molecular characterization of a D-galactoside-specific lectin purified from sea urchin (*Anthocidaris crassispina*) eggs. **Biochemistry.** 30: 2391-2394.
- Palmiano, P.E. and Juliano, O.B. (1973). Change in the activity of some hydrolases, peroxidase, and catalase in the rice seed during germination. **Plant Physiol.** 52: 274-277.

- Park, S., Gildersleeve, C.J., Blixt, O., and Shin, I. (2013). Carbohydrate microarrays. **Chem. Soc. Rev.** 42: 4310-4326
- Rojas A.L., Nagem, R.A.P., Neustroev, K.N., Arand, M., Adamska, M., Eneyskaya, E.V., Kulminskaya, A.A., Garratt, R.C., Golubev, A.M., and Polikarpov, I. (2004). Crystal structures of β -galactosidase from *Penicillium* sp. and its complex with galactose. **J. Mol. Biol.** 343: 1281-1292.
- Ross, G.S., Wegrzyn, T., Macrae, E.A., and Redgwell, R.J. (1994). Apple β -galactosidase: Activity against cell wall polysaccharides and characterization of related cDNA clone. **Plant Physiol.** 106: 521-528.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). **Molecular Cloning: a Laboratory Manual, 2nd ed.** Cold Spring Harbor Laboratory Press, New York.
- Sekimato, M., Ogura, K., Tsumuraya, Y., Hashimoto, Y., and Yamamoto, S. (1989). A β -galactosidase from radish (*Raphanus sativus* L.) seed. **Plant Physiol.** 90: 567-574.
- Seymour, G.B., Tatlor, J.E., and Tucker, G.A. (1993). **Biochemistry of Fruit Ripening.** Chapman & Hall, London. pp. 418-422.
- Sinnott, M.L. (1990). Catalytic mechanism of enzymatic glycosyl transfer. **Chem. Rev.** 90: 1171-1202.
- Smith, D.L. and Gross, K.C. (2000). A family of at least seven β -galactosidase genes is expressed during tomato fruit ripening. **Plant Physiol.** 123: 1173-1183.
- Stirk, H.J., Woolfson, D.N., Hutchinson, E.G., and Thornton, J. M. (1992). Depicting topology and handedness in jellyroll structures. **FEBS Letters.** 308(1): 1-3.

- Tanphanuch, W., Chantarangsee, M., Maneesan, J., and Ketudat Cairns, J. (2008). Genomic and expression analysis of glycosyl hydrolase family 35 genes from rice (*Oryza sativa* L.). **BMC Plant Biology**. 1-17.
- Tateno, H., Ogawa, T., Muramoto, K., Kamiya, H., Hirai, T. and Saneyoshi, M. (2001). A novel rhamnose-binding lectin family from eggs of steelhead trout (*Oncorhynchus mykiss*) with different structures and tissue distribution. **Biosci. Biotechnol. Biochem.** 65: 1328-1338.
- Tateno, H., Saneyoshi, A., Ogawa, T., Muramoto, K., Kamiya, H., and Saneyoshi, M. (1998). Isolation and characterization of rhamnose-binding lectins from eggs of steelhead trout (*Oncorhynchus mykiss*) homologous to low density lipoprotein receptor superfamily. **J. Biol. Chem.** 273: 19190-19197.
- Tateno, H., Yamaguchi, T., Ogawa, T., Muramoto, K., Watanabe, T., Kamiya, H., and Saneyoshi, M. (2002). Immunohistochemical localization of rhamnose-binding lectins in the steelhead trout (*Oncorhynchus mykiss*). **Dev. Comp. Immunol.** 26: 543-550.
- Terada, T., Watanabe, Y., Tateno, H., Naganuma, T., Ogawa, T., Muramoto, K., and Kamiya, H. (2007). Structural characterization of a rhamnose-binding glycoprotein (lectin) from Spanish mackerel (*Scomberomorus niphonius*) eggs. **Biochim. Biophys. Acta.** 1770: 617-629.
- Tomme, P., Van Tilbeurgh, H., Pettersson, G., van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T., and Claeysens, M. (1988). Studies of the cellulolytic system of *Trichoderma reesei* QM 9414: analysis of domain function in two cellobiohydrolases by limited proteolysis. **Eur. J. Biochem.** 170: 575-581.

- Tomme, P., Warren, R.A., and Gilkes, N.R. (1995). Cellulose hydrolysis by bacteria and fungi. **Adv. Microb. Physiol.** 37:1-81.
- Tormo, J., Lamed, R., Chirino, A.J., Morag, E., Bayer, E.A., Shoham, Y., and Steitz, T.A. (1996). Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. **EMBO J.** 15: 5739-5751.
- Trainotti, L., Spinello, R., Piovan, A., Spolaore, S., and Casadoro, G. (2001). β -galactosidases with a lectin-like domain are expressed in strawberry. **J. Exp. Botany.** 361: 1635-1645.
- Triantafillidou, D. and Georgatsos, J.G. (2001). Barley β -galactosidase: structure, function, heterogeneity, and gene origin. **J. Prot. Chem.** 20: 551-562.
- Vernon, C.A. (1967) The mechanisms of hydrolysis of glycosides and their relevance to enzyme-catalyzed reactions. **Proc. R. Soc. Ser B.** 167: 389-401.
- Viegas, A., Manso, J., Nobrega, F.L., and Cacrita, E.J. (2011). Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. **J. Chem. Edu.** 88: 990-994.
- Wallenfels, K. and Weil, R. (1972) **The Enzymes** (P.D. Boyer, ed.), 3rd edition, vol. 7: 617-663. Academic Press, NY.
- Wang, D., Liu, S., Trummer, J.B., Deng, and Wang, A. (2002). Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells. **Nature Biotech.** 20: 275-281.

- Watanabe, Y., Shiina, N., Shinozaki, F., Yokoyama, H., Kominami, J., and Nakamura-Tsuruta, S. (2008). Isolation and characterization of rhamnose-binding lectin, which binds to microsporidian *Glugea plecoglossus* from ayu (*Plecoglossus altivelis*) eggs. **Dev. Comp. Immunol.** 32: 487-499.
- Wüthrich, K. (2001). The way to NMR structures of proteins. **Nature Struct. Biol.** 8: 923-925.

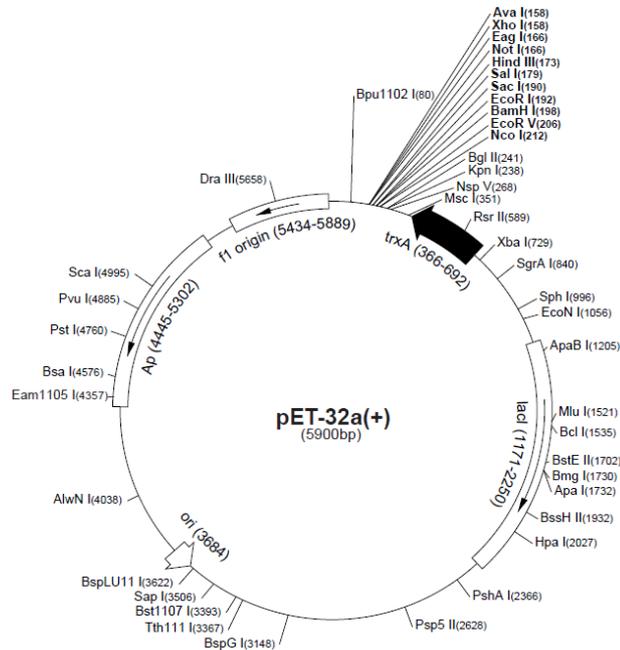
APPENDICES

APPENDIX A

PLASMID MAPS

1. pET32a(+) and pET32b(+) vector

The pET32a(+) and pET32b(+) vectors were designed for cloning and high level expression of protein sequences fused with 109 amino acids of thioredoxin protein (Trx-tag). Cloning sites are available for producing fusion proteins also containing cleavable His•Tag and S•Tag sequences for detection and purification. The sequence is numbered by pBR322 and the T7 expression region sequence is given and its position shown on the map. The cloning and expression region of the coding strand is transcribed by T7 RNA polymerase. pET32b(+) is the same as pET32a(+), except pET32b(+) is a 5899bp plasmid with a 1 bp shift in the reading from from the *Bam*HI site; subtract 1bp from each site beyond *Bam*H I at 198.

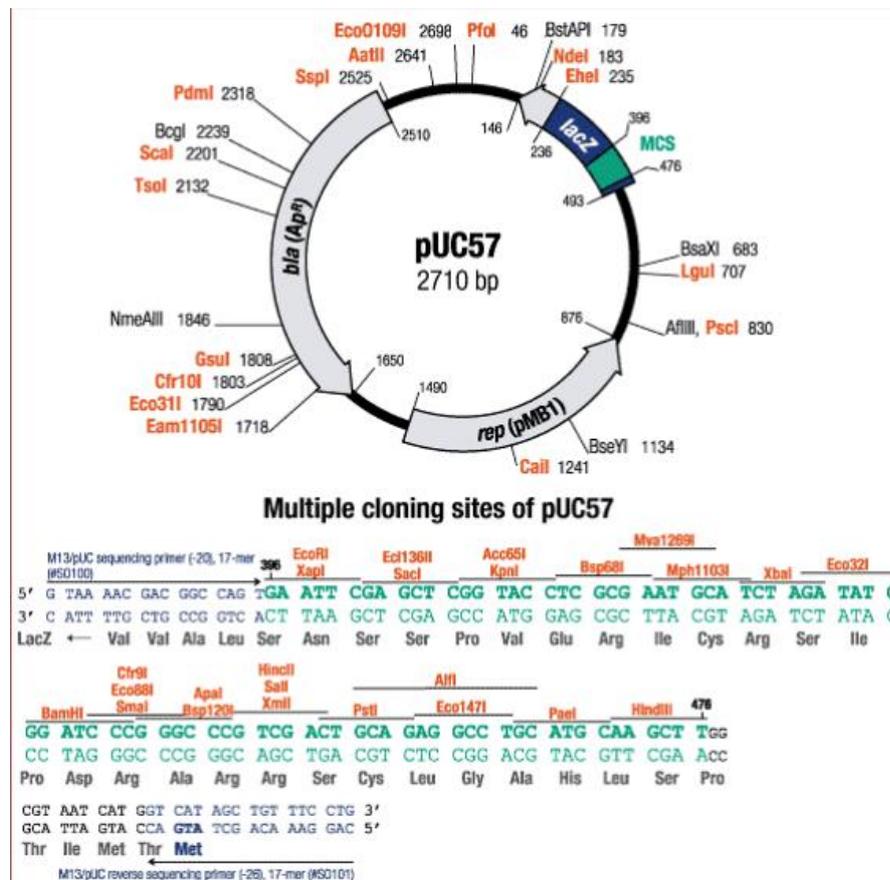


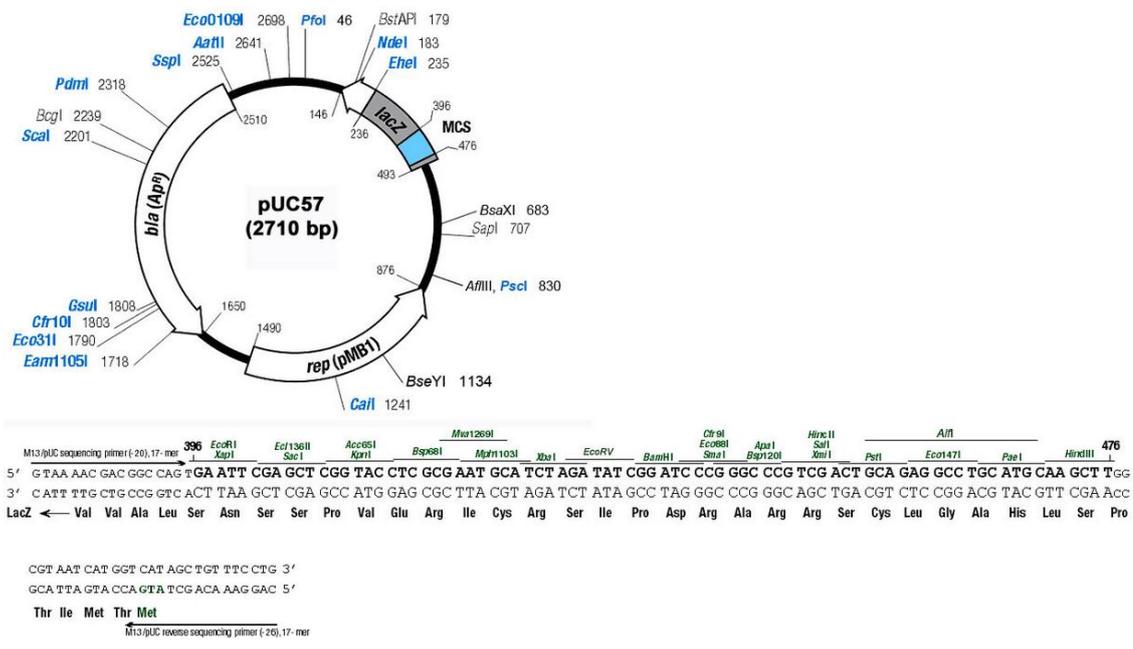
T7 promoter → **lac operator** *Xba I* **rbs**
 TAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGA
Trx•Tag *Msc I* **His•Tag**
 TATACATATGAGC...315bp...CTGGCCGGTTCTGGTTCTGGCCATATGCACCATCATCATCATCTTCTTCTGGTCTGGTGCACGCGTTCT
 MetSer 105aa...LeuAlaGlySerGlySerGlyHisMetHisHisHisHisHisHisSerSerGlyLeuValProArgGlySer
S•Tag *Nsp V* **S•Tag primer #69945-3** *Bgl II* *Kpn I* **thrombin I**
 GGTATGAAAGAAACCGCTGCTGCTAAATTCGAACGCCAGCACATGGACAGCCAGATCTGGGTACCGACGACGACGACAAG
 GlyMetLysGluThrAlaAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAspAspAspLys
pET-32a(+) *Nco I* *EcoR V* *BamH I* *EcoR I* *Sac I* *Sal I* *Hind III* *Eag I* *Not I* *Ava I* *Xho I* **His•Tag** **enterokinase**
 GCCATGGCTGATATCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGGCCGCACTCGAGCACCACCACCACCACCCTGAGATCCGGCTGCTAA
 AlaMetAlaAspIleGlySerGluPheGluLeuArgArgGlnAlaCysGlyArgThrArgAlaProProProProProLeuArgSerGlyCysEnd
 GCCATGGCGATATCTGTGGATCCGAATTCGAGCTCCGTCGACAAGCTTGGCCGCACTCGAGCACCACCACCACCACCCTGAGATCCGGCTGCTAA **pET-32b(+)**
 AlaMetAlaIleSerAspProAsnSerSerSerValAspLysLeuAlaAlaAlaGluGluHisHisHisHisHisHisEnd
 GCCATGGGATATCTGTGGATCCGAATTCGAGCTCCGTCGACAAGCTTGGCCGCACTCGAGCACCACCACCACCACCCTGAGATCCGGCTGCTAA **pET-32c(+)**
 AlaMetGlyTyrLeuTrpIleArgIleArgAlaProSerThrSerLeuArgProHisSerSerThrThrThrThrThrThrThrGluIleArgLeuLeuThr
Bpu1102 I **T7 terminator**
 CAAAGCCCGAAAGGAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACCTAGCATAACCCCTTGGGGCTCTAAACGGTCTTGAGGGTTTTTGG
 LysProGluArgLysLeuSerTrpLeuLeuProProLeuSerAsnAsnEnd
 ← **T7 terminator primer #69337-3**

pET-32a-c(+) cloning/expression region

2. pUC57

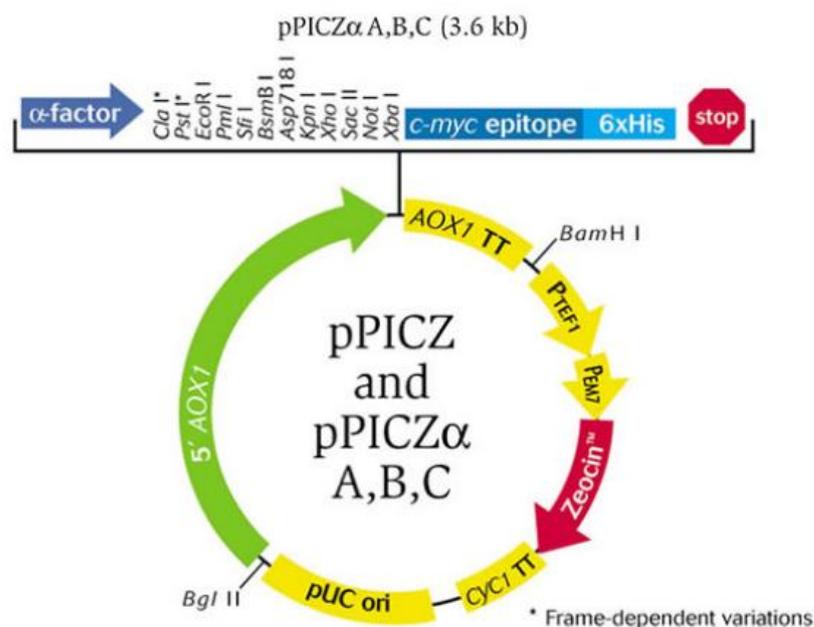
pUC57 is a plasmid cloning vector commonly used in *E. coli*. The vector length is 2,710 bp and is isolated from *E. coli* strain DH5 α by standard procedures. The map and multiple cloning site sequence are shown below.





3. pPICZ α B vector

pPIC α B vector is 3.6 kb used to express and secrete recombinant proteins in *Pichia pastoris*. Recombinant proteins are expressed as fusions to an N-terminal peptide encoding the *Saccharomyces cerevisiae* α -factor secretion signal. This vector allows high-level, methanol inducible expression of the gene of interest in *Pichia pastoris* strain, such as X-33, SMD1168H, and KM71H. pPICZ α vectors contain AOX1 promoter for tightly regulated, methanol-induced expression of the gene of interest, reading frames are provided to facilitate in-frame cloning with the C-terminal peptide and α -factor secretion signal for directing secreted expression of the recombinant protein. The also contain the zeocin resistance gene for selection in both *E. coli* and *P. pastoris* and a sequence encoding a C-terminal peptide containing the c-myc epitope and a polyhistidine (6xHis) tag for detection and purification of a recombinant fusion protein.



APPENDIX B

NMR CHEMICAL SHIFTS FOR OsBGal1-Cter

Table chemical shifts of OsBGal1 Cter

Amino acid residue	NH	H	C=O	C α	C β	H α	H β	Other
R726			176.02	55.97	31.32	4.44	1.74, 1.84	C γ 27.16 H γ 1.57, 1.63 C δ 43.42 H δ 3.16, -
T727	117.4	8.29	174.06	61.83	70.20	4.44	4.14	C γ 2 21.78 H γ 2 1.13
V728	122.03	8.34	175.69	61.77	33.88	4.55	2.05	C γ 1 21.36 H γ 2 0.92 C γ 2 20.33 H γ 2 0.90
S729	118.42	8.68	173.82	57.70	64.83	4.67	3.91, -	
G730	110.34	8.53	173.29	45.53		4.01, 4.66		
V731	118.24	8.42	174.15	60.93	34.66	4.67	2.17	C γ 1 21.59 H γ 2 0.94
C732			173.64	53.74	49.4	5.76	3.17, 3.73	

Table chemical shifts of OsBGal1 Cter (Continued)

Amino acid residue	NH	H	C=O	Ca	Cβ	Hα	Hβ	Other
Q79C	123.01	9.07	176.11	58.41	29.18	3.94	1.79, 1.89	Cy 33.77 Hy 1.90, 2.09
G79I	114.42	9.37	174.1	42.84		3.99, 4.42		Cy 26.41
F792	118.04	8.72	175.72	58.75	29.89	3.89	1.95, -	Hy 2.24, 2.27
C793	114.73	6.95		52.87	40.18	4.96	2.87, 3.55	
H794			173.07	56.74	33.83	4.57	2.71, 2.95	
S795				57.32	63.01	4.61	3.79, 3.97	
I796			176.15	63.55	37.95	4.09	2.13	Cy1 26.92 Hy1 1.39, 1.43 Cy2 17.57 Hy2 1.08 C8 14.01 I18 0.97
N797			174.96	53.21	39.01	5.00	2.78, 3.04	
S798				64.32	63.96	4.09	3.26	
N799			176.32	57.67	38.21	4.00	2.68, 2.72	
S800			177.17	61.46	63.11	4.09	3.93, -	Cy1 19.9 Hy1 0.08
V801	121.07	7.96	177.24	66.44	31.61	3.48	1.75	Cy2 23.52 Hy2 0.81

Table chemical shifts of OsBGalI Cter (Continued)

Amino acid residue	NH	H	C=O	Ca	Cβ	Hα	Hβ	Other
L802	119.73	7.06	179.01	55.90	40.07	3.78	1.33, 1.92	Cγ 28.67 Hy 1.8 Cδ1 25.8 Hδ1 22.35 Cδ2 0.86 Hδ2 0.87 Cγ 37.46 Hy 1.99, 2.48
E803	117.84	8.69	179.06	59.54	29.89	3.69	1.88, 2.10	Cγ 27.34 Hy 1.69, - Cδ 43.44 Hδ 3.16, -
R804	117.87	8.00	179.16	58.75	30.01	3.97	1.86, 1.92	Cγ 25.41 Hy 1.36, 1.53 Cδ 28.95 Hδ 1.56, - Cε 42.07 Hε 2.87, -
K805	114.25	7.91	178.25	57.39	33.67	4.18	1.70, 1.81	Cy 1 29.36 Hy 1.07, 1.59 Cy 16.57 Hy 2 0.85 Cδ 11.93 Hδ 0.76
C806	110.70	7.55	174.74	57.12	44.58	4.84	3.02, 3.10	
I807	117.47	7.57	178.03	63.54	35.27	3.66	2.00	
G808	114.09	9.41		44.34		3.64, 4.34		

Table chemical shifts of OsBGal1 Cter (Continued)

Amino acid residue	NH	H	C=O	C α	C β	H α	H β	Other
L809	120.00	7.73	176.69	53.47	43.97	4.76	1.66, 1.80	Cy 26.98 Hy 1.47 C δ 1 26.05 H δ 1 1.04 C δ 2 22.87 H δ 2 0.91 Cy 36.62 Hy 2.14, 2.43
E810	121.24	8.77	174.23	58.50	31.76	4.21	1.92, 2.20	Cy 26.48 Hy 1.51,- C δ 43.71 H δ 3.16, -
R811	111.08	7.28	173.99	53.53	33.58	5.21	1.83, -	Cy1 20.88 Hy1 0.58
C812	118.27	9.12	171.57	54.00	46.43	4.92	2.85, 3.32	Cy1 20.85 Hy1 0.62
V813	124.49	8.45	175.09	60.36	34.02	5.02	1.75	Cy2 21.42 Hy2 0.67
V814	126.10	8.65	174.57	60.62	35.50	4.13	1.67	
A815	129.90	8.57	177.17	52.31	18.56	4.35	1.20	Cy1 26.57 Hy1 1.50, - Cy2 18.04 Hy2 0.52 C δ 1 13.82 H δ 1 0.59
I816	121.82	8.62	173.60	61.58	36.21	3.70	1.75	

Table chemical shifts of OsBGaII Cter (Continued)

Amino acid residue	NH	H	C=O	C α	C β	H α	H β	Other
I816	121.82	8.62	173.60	61.58	36.21	3.70	1.75	Cy1 26.57 Hy1 1.50, - Cy2 18.04 Hy2 0.52 C δ 1 13.82 H δ 1 0.59
S817	118.78	6.99		54.61	64.52	5.07	3.72, 4.08	
P818								
S819			176.97	61.35	62.19	4.1	3.81	
N820	119.35	7.67	175.48	54.14	38.78	4.57	2.17, 2.61	
F821	114.17	7.38	174.74	58.55	40.64	4.39	2.52, 3.66	
G822	108.25	7.72	173.98	46.28		3.97, 3.84		
G823				44.16		3.63, 4.37		
D824								
P825			176.37	63.46	31.97	4.46	2.07	
C826				54.55	44.69	4.76	3.11, 3.27	
P827			176.98	64.02	32.16	4.52	1.81, 2.22	Cy 27.34 Hy 1.80, 2.03 C δ 50.13 H δ 3.08, 3.33
E828	115.19	8.95	175.29	57.57	28.00	3.96	2.12, -	Cy 36.61 Hy 2.17, -
V829	120.41	7.47	175.08	60.74	34.20	4.12	1.99	Hy1 0.79 Cy2 21.48 Hy2 0.76

Table chemical shifts of OsBGal1 Cter (Continued)

Amino acid residue	NH	H	C=O	C α	C β	H α	H β	Other
M830			176.19	56.11	32.02	4.34	1.93, 2.01	Cy 31.98 Hy 2.44, 2.50
K831			175.65	55.27	36.12	4.48	1.50, -	Cy 26.37
R832	117.16	9.14	175.23	53.79	35.07	5.18	1.46, 1.82	Hy 1.44, 1.60 C δ 43.86 H δ 3.10, 3.20
R832								Cy 21.48 Hy 1 0.63
V833	118.19	8.79		61.15	33.26	5.16	1.67	Cy 2 24.15 Hy 2 0.67
A834	128.03	8.09	176.07	49.50	21.41	5.14	0.22	Cy 1 21.51
V835	119.17	8.84	173.72	59.29	36.25	5.11	1.89	Hy 1 1.01 Cy 2 22.06 Hy 2 0.92 Cy 36.78
E836			174.74	55.67	33.07	5.22	2.13, 2.28	Hy 2.15, 2.52 7
A837			174.65	50.07	23.16	5.17	1.40	Cy 1 21.38
V838			175.21	61.86	34.42	4.48	1.91	Hy 1 0.8
C839	124.96	8.83	174.21	55.04	44.89	5.39	2.80, 3.02	
S840	115.29	9.33	174.07	57.22	66.41	5.18	3.80, 3.89	

Table chemical shifts of OsBGalI Cter (Continued)

Amino acid residue	NH	H	C=O	C α	C β	H α	H β	Other
T841	117.32	8.49	173.96	63.10	69.95	4.37	4.18	Cy2 22.09 Hy21.31
A842	128.75	8.5	176.40	52.72	19.32	4.29	1.34	
A843	129.58	7.99		53.75	20.18	4.07	1.29	



Table chemical shifts of OsBGal1 Cter (Continued)

Amino acid residue	NH	H	C=O	C α	C β	H α	H β	Other
A733	124.34	9.04	174.78	51.92	23.89	4.74	1.35	
D734	120.46	8.27	175.62	53.80	41.73	5.34	2.72, 2.92	
V735	123.98	8.65		61.14	35.88	4.47	1.84	C γ 1 21.46 H γ 2 0.97 C γ 1 21.77 H γ 2 1.11
S736	120.98	8.56	174.39	57.7	64.84	4.72	3.89, -	
E737	121.68	8.44	175.47	57.51	31.01	4.09	1.95, 2.10	C γ 36.39 H γ 2.22, 2.38
Y738	120.41	8.14		58.14	39.19	4.56	2.90, -	
H739	125.29	7.99	174.82	54.24	31.26	4.61	2.93, 2.99	
P740				63.55	32.08	4.26	1.82, 2.19	C γ 27.18 H γ 1.82 C δ 50.46 H δ 3.08, 3.34
N741			175.38	53.27	38.62	4.69	2.73, 2.86	
I742	120.73	7.85	176.02	61.53	38.70	4.03	1.79	C γ 1 27.36 H γ 1 1.10, 1.35 C γ 2 17.46 H γ 2 0.80 C δ 13.06 H δ 0.77

Table chemical shifts of OsBGal1 Cter (Continued)

Amino acid residue	NH	H	C=O	C α	C β	H α	H β	Other
K743	124.01	8.24		56.52	32.65	4.18	1.60, -	C γ 24.69 H γ 1.27, 1.29 C δ 28.96 H δ 1.57, - C ϵ 42.09
Q746	121.04	7.97		56.23	29.20	4.14	1.79, 1.93	C γ 33.51 H γ 1.91, 2.02
I747	120.45	7.81	176.43	61.59	38.68	4.01	1.79	C γ 1 27.48 H γ 1 1.40, 1.11 C γ 2 17.47 H γ 2 0.82 C δ 13.08 H δ 0.81
E748	123.13	8.39	176.38	56.88	29.93	4.12	1.94,1.83	C γ 36.37 H γ 2.18, -
S749	115.64	8.03	174.19	58.35	63.76	4.29	3.65	
Y750	121.46	8.07	176.22	58.18	38.49	4.44	2.88, 3.02	
G751	110.12	8.17		44.85		3.80, 3.89		
E752	121.41	8.05	173.6	54.34	29.75	4.53	1.86, 2.00	C γ 36.11 H γ 2.22, -
P753			177.09	63.73	31.98	4.24	1.74, 2.12	C γ 27.44 H γ 1.91, -

Table chemical shifts of OsBGal1 Cter (Continued)

Amino acid residue	NH	H	C=O	C α	C β	H α	H β	Other
P753								C δ 50.48 H δ 3.64, 3.37
E754			176.32	56.7	29.86	4.12	1.84, -	C γ 36.2 H γ 2.05, 2.08
F755				58.13	39.49	4.45	2.96, -	
H756			174.63	56.15	30.93	4.48	2.84, 2.91	
T757	114.45	7.73	174.10	61.44	70.79	4.35	4.05	C γ 2 21.91 H γ 2 1.17
A758	128.29	8.79	176.47	52.24	20.62	4.35	1.37	
K759	118.05	8.33	175.18	54.67	35.94	4.9	1.26, 1.45	C γ 24.9 H γ 1.28, 1.11 C δ 29.21 H δ 1.42, - C ϵ 42.42 H ϵ 2.82, -
V760	120.05	8.70	173.77	61.35	33.48	4.07	1.9	C γ 1 21.19 H γ 2 0.72 C γ 2 21.66 H γ 2 0.81
H761	126.27	8.57	174.71	54.78	32.12	5.21	2.92, 2.95	
L762	123.85	8.48	174.96	53.50	45.30	4.60	1.12, 1.75	C γ 26.56

Table chemical shifts of OsBGal1 Cter (Continued)

Amino acid residue	NH	H	C=O	C α	C β	H α	H β	Other
L762								H γ 1.36 C δ 1 25.79 H δ 1 0.7 C δ 2 24.03 H δ 2 0.69
K763	121.54	8.15	174.98	55.29	35.64	4.97	1.67, 1.72	C γ 24.43 H γ 1.28, 1.39 C δ 29.79 H δ 1.54, 1.59 C ϵ 42.08 H ϵ 2.87, -
C764	122.00	8.53	173.69	52.88	40.54	4.82	2.67, 3.28	
A765	123.33	7.99		50.99	17.14	4.45	1.22	
P766			178.16	63.82	31.49	4.25	2.28	C γ 27.86 C γ 1.98, 2.10 C δ 50.06 H δ 3.50,
G767	112.92	8.55	174.15	45.01		3.65, 4.29		
Q768	119.35	8.23	175.48	54.10	31.97	4.89	1.75, -	C γ 33.65 2.14, 1.93
T769	108.02	8.60	173.91	59.68	72.87	4.44	3.55	C γ 22.59 H γ 0.92

Table chemical shifts of OsBGal1 Cter (Continued)

Amino acid residue	NH	H	C=O	C α	C β	H α	H β	Other
I770	122.8	8.65	176.81	62.38	36.64	4.47	2.19	C γ 1 29.92 H γ 1 2.02, 0.74 C γ 2 17.49 H γ 2 0.65 C δ 15.01 H δ 0.65
S771	124.88	9.50	174.83	57.79	64.92	4.67	3.67, 3.91	
A772	121.68	7.65	174.87	52.21	21.71	4.38	1.29	
I773	121.99	9.28	175.22	57.93	35.35	4.39	2.23	C γ 1 26.14 H γ 1 1.03, 1.46 C γ 2 17.86 H γ 2 0.75 C δ 8.64 H δ 0.52
K774	128.96	8.58	176.40	57.71	33.66	4.24	1.64, -	C γ 24.82
F775	117.88	8.17	173.18	58.10	43.49	4.58	2.72, 3.31	H γ 1.37, 1.43 C δ 28.97 H δ 1.59, - C ϵ 41.92 H ϵ 2.92, -

Table chemical shifts of OsBGal1 Cter (Continued)

Amino acid residue	NH	H	C=O	C α	C β	H α	H β	Other
A776	127.68	7.41	175.44	51.36	22.28	4.94	0.99	
S777	112.09	8.34		58.91	65.15	4.37	3.39, 3.81	
F778	128.03	8.09	174.07	53.21	38.97	5.45	1.75, 2.28	
G779	113.53	8.56	171.27	46.80		4.10, 3.10		
T780	113.01	8.52		56.86	69.89	4.96	4.54	C γ 2 21.47 H γ 2 1.11
P781			175.63	63.43	32.06	4.36	1.91, 2.15	C γ 26.55 H γ 1.67, - C δ 50.71 C δ 3.13, 3.49
L782			177.26	53.5	44.86	4.72	1.51, 1.59	C γ 27.34 H γ 1.58 C δ 1 25.35 H δ 1 0.87 C δ 2 23.35 H δ 2 0.91
G783	105.35	8.42	173.42	44.6		3.63, 4.22		
T784	109.16	6.83	171.73	58.91	71.34	4.28	4.04	C γ 20.62 H γ 0.96
C785	117.33	7.96	175.27	57.08	37.83	3.46	2.78, 2.93	
G786	118.89	7.69	175.45	45.64		3.76, 4.33		
T787	112.55	8.61	174.18	59.67	69.78	4.45	4.4	C γ 2 21.45 H γ 2 0.94
F788	116.98	8.40	176.95	57.79	38.29	4.96	1.44, 1.49	
Q789	116.91	8.41	174.73	53.97	33.72	4.49	1.73, 2.05	C γ 33.75 H γ 2.26, -

Table chemical shifts of OsBGal1 Cter (Continued)

Amino acid residue	NH	H	C=O	C α	C β	H α	H β	Other
Q790	123.01	9.07	176.11	58.41	29.18	3.94	1.79, 1.89	C γ 33.77 H γ 1.90, 2.09
G791	114.42	9.37	174.1	42.84		3.99, 4.42		
E792	118.04	8.72	175.72	58.75	29.89	3.89	1.95, -	C γ 36.41 H γ 2.24, 2.27
C793	114.73	6.95		52.87	40.18	4.96	2.87, 3.55	
H794			173.07	56.74	33.83	4.57	2.71, 2.95	
S795				57.32	63.01	4.61	3.79, 3.97	
I796			176.15	63.55	37.95	4.09	2.13	C γ 1 26.92 H γ 1 1.39, 1.43 C γ 2 17.57 H γ 2 1.08 C δ 14.01 H δ 0.97
N797			174.96	53.21	39.01	5.00	2.78, 3.04	
S798				64.32	63.96	4.09	3.26	
N799			176.32	57.67	38.21	4.00	2.68, 2.72	
S800			177.17	61.46	63.11	4.09	3.93, -	
V801	121.07	7.96	177.24	66.44	31.61	3.48	1.75	C γ 1 19.9 H γ 1 0.08 C γ 2 23.52 H γ 2 0.81

Table chemical shifts of OsBGal1 Cter (Continued)

Amino acid residue	NH	H	C=O	C α	C β	H α	H β	Other
L802	119.73	7.06	179.01	55.90	40.07	3.78	1.33, 1.92	C γ 28.67 H γ 1.8 C δ 1 25.8 H δ 1 22.35 C δ 2 0.86 H δ 2 0.87
E803	117.84	8.69	179.06	59.54	29.89	3.69	1.88, 2.10	C γ 37.46 H γ 1.99, 2.48
R804	117.87	8.00	179.16	58.75	30.01	3.97	1.86, 1.92	C γ 27.34 H γ 1.69, - C δ 43.44 H δ 3.16, -
K805	114.23	7.91	178.25	57.39	33.67	4.18	1.70, 1.81	C γ 25.41 H γ 1.36, 1.53 C δ 28.95 H δ 1.56, - C ϵ 42.07 H ϵ 2.87, -
C806	110.70	7.55	174.74	57.12	44.58	4.84	3.02, 3.10	
I807	117.47	7.57	178.03	63.54	35.27	3.66	2.00	C γ 1 29.36 H γ 1.07, 1.59 C γ 16.57 H γ 2 0.85 C δ 11.93 H δ 0.76
G808	114.09	9.41		44.34		3.64, 4.34		

Table chemical shifts of OsBGal1 Cter (Continued)

Amino acid residue	NH	H	C=O	C α	C β	H α	H β	Other
L809	120.00	7.73	176.69	53.47	43.97	4.76	1.66, 1.80	C γ 26.98 H γ 1.47 C δ 1 26.05 H δ 1 1.04 C δ 2 22.87 H δ 2 0.91
E810	121.24	8.77	174.23	58.50	31.76	4.21	1.92, 2.20	C γ 36.62 H γ 2.14, 2.43
R811	111.08	7.28	173.99	53.53	33.58	5.21	1.83, -	C γ 26.48 H γ 1.51,- C δ 43.71 H δ 3.16, -
C812	118.27	9.12	171.57	54.00	46.43	4.92	2.85, 3.32	
V813	124.49	8.45	175.09	60.36	34.02	5.02	1.75	C γ 1 20.88 H γ 1 0.58
V814	126.10	8.65	174.57	60.62	35.50	4.13	1.67	C γ 1 20.85 H γ 1 0.62 C γ 2 21.42 H γ 2 0.67
A815	129.90	8.57	177.17	52.31	18.56	4.35	1.20	
I816	121.82	8.62	173.60	61.58	36.21	3.70	1.75	C γ 1 26.57 H γ 1 1.50, - C γ 2 18.04 H γ 2 0.52 C δ 1 13.82 H δ 1 0.59

Table chemical shifts of OsBGal1 Cter (Continued)

Amino acid residue	NH	H	C=O	C α	C β	H α	H β	Other
I816	121.82	8.62	173.60	61.58	36.21	3.70	1.75	C γ 1 26.57 H γ 1 1.50, - C γ 2 18.04 H γ 2 0.52 C δ 1 13.82 H δ 1 0.59
S817	118.78	6.99		54.61	64.52	5.07	3.72, 4.08	
P818								
S819			176.97	61.35	62.19	4.1	3.81	
N820	119.35	7.67	175.48	54.14	38.78	4.57	2.17, 2.61	
F821	114.17	7.38	174.74	58.55	40.64	4.39	2.52, 3.66	
G822	108.25	7.72	173.98	46.28		3.97, 3.84		
G823				44.16		3.63, 4.37		
D824								
P825			176.37	63.46	31.97	4.46	2.07	
C826				54.55	44.69	4.76	3.11, 3.27	
P827			176.98	64.02	32.16	4.52	1.81, 2.22	C γ 27.34 H γ 1.80, 2.03 C δ 50.13 H δ 3.08, 3.33
E828	115.19	8.95	175.29	57.57	28.00	3.96	2.12, -	C γ 36.61 H γ 2.17, -
V829	120.41	7.47	175.08	60.74	34.20	4.12	1.99	H γ 1 0.79 C γ 2 21.48 H γ 2 0.76

Table chemical shifts of OsBGal1 Cter (Continued)

Amino acid residue	NH	H	C=O	C α	C β	H α	H β	Other
M830			176.19	56.11	32.02	4.34	1.93, 2.01	C γ 31.98 H γ 2.44, 2.50
K831			175.65	55.27	36.12	4.48	1.50, -	
R832	117.16	9.14	175.23	53.79	35.07	5.18	1.46, 1.82	C γ 26.37 H γ 1.44, 1.60
R832								C δ 43.86 H δ 3.10, 3.20
V833	118.19	8.79		61.15	33.26	5.16	1.67	C γ 1 21.48 H γ 1 0.63 C γ 2 24.15 H γ 2 0.67
A834	128.03	8.09	176.07	49.50	21.41	5.14	0.22	
V835	119.17	8.84	173.72	59.29	36.25	5.11	1.89	C γ 1 21.51 H γ 1 1.01 C γ 2 22.06 H γ 2 0.92
E836			174.74	55.67	33.07	5.22	2.13, 2.28	C γ 36.78 H γ 2.15, 2.52
A837			174.65	50.07	23.16	5.17	1.40	7
V838			175.21	61.86	34.42	4.48	1.91	C γ 1 21.38 H γ 1 0.8
C839	124.96	8.83	174.21	55.04	44.89	5.39	2.80, 3.02	
S840	115.29	9.33	174.07	57.22	66.41	5.18	3.80, 3.89	

Table chemical shifts of OsBGal1 Cter (Continued)

Amino acid residue	NH	H	C=O	Cα	Cβ	Hα	Hβ	Other
T841	117.32	8.49	173.96	63.10	69.95	4.37	4.18	C γ 2 22.09 H γ 21.31
A842	128.75	8.5	176.40	52.72	19.32	4.29	1.34	
A843	129.58	7.99		53.75	20.18	4.07	1.29	

CURRICULUM VITAE

Name	Miss Thipwarin Rimlumduan
Date of Birth	8 January 1979
Place of Birth	Nakhon Ratchasima, Thailand
Education	
1996-2000	Biochemistry Department, Khon Kaen University, Khon Kaen, Thailand B.Sc. in Biochemistry
2001-2004	Biochemistry Department, Khon Kaen University, Khon Kaen, Thailand M.Sc. in Biochemistry
2007-2013	School of Biochemistry, Suranaree University of Technology, Nakhon Ratchasima, Thailand Ph.D. in Biochemistry
Grants and Fellowships	
2007-2011	The Commission on Higher Education, Thailand
April-July 2010	Asian Young Researchers in the Field of Life Sciences and Bioengineering, University of Tsukuba, Ibaraki, Japan