

X-RAY CRYSTALLOGRAPHIC STUDIES  
OF  
ESCHERICHIA COLI BRANCHING ENZYME  
IN COMPLEX WITH MALTOOCTAOSE  
AND  
RICE BRANCHING ENZYME I  
IN COMPLEX WITH DODECAOSE

By

Remie Fawaz

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Chemistry—Doctor of Philosophy

2016

## ABSTRACT

### X-RAY CRYSTALLOGRAPHIC STUDIES OF ESCHERICHIA COLI BRANCHING ENZYME IN COMPLEX WITH MALTOOCTAOSE AND RICE BRANCHING ENZYME I IN COMPLEX WITH DODECAOSE

By

Remie Fawaz

Branching enzyme plays a key role in determining the final structure of glycogen or starch; this outcome structure is unique to every species, therefore the diversity of branching enzymes structures. While the biosynthesis of the polymers constitutes of three major steps, the last reaction of the pathway catalyzed by branching enzyme embodies the cleavage of the  $\alpha$ -1,4 glycosidic bond and the transfer of the produced oligosaccharide to the specific  $\alpha$ -1,6 position creating an  $\alpha$ -1,6 branch point. This process results in highly branched polymeric structures, which represent major carbon sources and carbohydrate storage compounds in living organisms, and are essential both in nature and industry.

The structure of *Escherichia coli* glycogen branching enzyme has been solved both in the apo and holo forms. Binding to linear and cyclic oligosaccharides has been studied and showed seven external binding sites, but binding in the active site was not seen. Oligomers longer than previously used were investigated in an attempt to see binding in the catalytic center, only to discover more peripheral binding sites seemingly independent of each other. Examining the binding sites' locations and sugars' orientations indicates that these sites probably cooperate together in order to hold the long sugar polymer

on the surface of the protein during the branching reaction. Hypotheses regarding the mode of binding between *E. coli* branching enzyme and the sugar during the reaction, and the roles for key binding sites and residues in the mechanism are proposed.

Rice branching enzyme I, a starch branching enzyme, was also previously crystallized and the structure of the truncated version was solved. Binding attempts on this protein showed surface binding sites as well. In an effort to better understand the mechanism of action of this enzyme, we crystallized the protein and soaked it in a 12-unit oligomer, which bound to the enzyme hanging over the active site without reaching into the groove. Two new binding sites were discovered for this protein and the residues involved were identified. There are several hydrogen bonds and aromatic stacking interactions within the binding sites. Hypotheses for the binding mode and reaction mechanism are proposed.

In addition, we have worked on ADP-glucose pyrophosphorylases that catalyze the first step in the biosynthesis pathway. The small subunit of the protein expresses as a homotetramer that is highly active. Although the potato tuber form was previously expressed in a collaborating laboratory at extremely small levels, and the protein was crystallized in the inactive form in the Geiger lab, our intensive attempts only succeeded in purifying the protein of interest, but the enzyme would precipitate instead of concentrating.

Copyright by  
REMIE FAWAZ  
2016

This thesis is dedicated to my dear husband, Rabih, my daughter, Anna Maria, and my brother and best friend, Marc, who preceded me to Heaven. Thank you for always believing in me.

## ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, professor James H. Geiger for leading, helping, mentoring, advising, and guiding me throughout this process. He is a role model in perseverance and in thinking of new ways to overcome roadblocks. He encouraged me enormous times and I cannot have accomplished what I have as a graduate student without his patience and help. He has a passion for science in general and the starch projects going on in his lab in particular, which keeps him informed of every new little detail in the field while preserving a unique direction in projects. I sincerely thank him for all his dedicated time and help.

I would also like to thank our collaborator, Dr. Andrew Mort, from Oklahoma State University for performing the HPAEC for the chain length specificity assay on the *E. coli* branching enzyme mutants; this assay represents an important part in understanding the branching enzyme critical residues for determining the branch length.

A special thank to Dr. Stacy Hovde for collaborating in regards to the *E. coli* branching enzyme project, but most importantly for providing laboratory prepared starting enzymes, such as Sumo protease, and special expression cells, to say the least. In addition, Stacy always landed many useful protocols and small tricks for every area in the research. Thank you Stacy!

I would also like to thank my group members Susan Wortas, Meisam Nosrati, Camille Watson and Xiaofei Jia, and especially Rafida Nossoni, who has

become a dear friend. Although our projects were very different, discussing issues with Rafida was very beneficial, and sharing our life circumstances has empowered me with much energy when needed.

Finally, I would like to thank my beloved husband who supported and believed in me regardless of situations. I depended on his encouragement always when I needed it most. I would like to also thank my daughter, who put up with me since she was so little throughout my studies. I would like to give a big shout out to my mother who always motivated me to pursue my dream field of study to the highest degree, so I owe her much thanks for this inspiration. Also, I thank my dear brothers who listened and helped when I called them.

## TABLE OF CONTENTS

LIST OF TABLES	xi
LIST OF FIGURES	xii
KEY TO ABBREVIATIONS	xv
<b>CHAPTER 1: LITERATURE REVIEW</b>	<b>1</b>
1.1 Starch versus Glycogen	1
1.1.1 Natural occurrence and function	1
1.1.2 Structure and organization	2
1.1.3 Biosynthesis	7
1.1.4 Industrial connection and relevance	9
1.2 Branching Enzyme	13
1.2.1 Background	13
1.2.2 Branching enzyme isoforms	16
1.2.3 <i>Escherichia coli</i> branching enzyme (EcBE)	18
1.2.4 Rice branching enzyme I	21
1.2.4.1 Background on starch branching enzyme	21
1.2.4.2 Starch branching isozymes in rice	22
1.2.4.3 Rice branching enzyme 1	29
1.2.5 <i>Mycobacterium tuberculosis</i> branching enzyme (MycoBE)	32
1.2.6 Activity assay	34
1.2.7 Chain length specificity assay	35
REFERENCES	36
<b>CHAPTER 2: ACTIVITY AND CHAIN LENGTH SPECIFICITY ASSAYS</b>	
<b>IN <i>ESCHERICHIA COLI</i> BE</b>	<b>49</b>
2.1 Objective	49
2.2 The Crystal Structure of <i>E. coli</i> BE (EcBE)	50
2.3 EcBE Glucan Binding Sites	51
2.3.1 Binding sites I and II	53
2.3.2 Binding sites III, IV, V and VI	55
2.3.3 Cyclodextrin exclusive binding site VII	58
2.4 What Residues to Mutate Based on Binding Sites Discovered	61
2.5 Results and Discussion	66
2.5.1 N112BE expression vector and purification	66
2.5.2 Activity assay (N112BE)	66
2.5.3 Single mutations	70

2.5.4	Combination mutations	79
2.5.5	Activity assay (full-length EcBE)	82
2.5.6	Chain length specificity assay	86
2.6	Conclusion and Hypothesis	95
2.6.1	Structure-function correlation	95
2.6.2	Hypothesis on enzymatic activity	97
	REFERENCES	99

<b>CHAPTER 3: X-RAY CRYSTALLOGRAPHIC STUDIES OF <i>E. COLI</i> BE IN COMPLEX WITH OCTAOSE</b>	103	
3.1	Research Objective	103
3.2	The Three Dimensional Structure of EcBE Bound to Maltotetraose	107
3.2.1	Structure determination	107
3.2.2	Comparing the molecules in the asymmetric unit	109
3.2.3	Detailed interactions in each binding site	115
3.2.3.1	Binding sites I and II	115
3.2.3.2	Binding sites III and VIII	118
3.2.3.3	Binding sites IV and IX	120
3.2.3.4	Binding site V	121
3.2.3.5	Binding site IX	123
3.2.3.6	Binding site X	124
3.3	Structural Comparison	127
3.3.1	CBM48 domain	127
3.3.2	Comparing EcBE – M8 to EcBE – M7	129
3.3.2.1	Binding sites I and II compared	129
3.3.2.2	Binding sites III compared	131
3.3.2.3	Binding sites V	136
3.3.2.4	Binding site VIII, IX and X	138
3.4	Hypothesis and Conclusion	140
	REFERENCES	145

<b>CHAPTER 4: X-RAY CRYSTALLOGRAPHIC STUDIES OF RICE BEI IN COMPLEX WITH LINEAR OLIGOSACCHARIDES</b>	150	
4.1	Research Objective	150
4.2	Why a Twelve Unit Oligomer?	154
4.3	The Three Dimensional Structure of Rice BE I Bound to Dodecaose	157
4.3.1	Crystallization: attempts and successes	157
4.3.2	Structure determination	159
4.3.3	Overall structure	161
4.3.3.1	Binding site I	164
4.3.3.2	Binding sites IV and V	167

4.3.4	Structural comparison	171
4.3.4.1	To GH13 family enzymes	171
4.3.4.2	Active site residues	173
4.3.4.3	Carbohydrate binding module	174
4.3.4.4	Binding in site I of RBEI-M12 compared with RBEI-M5	180
4.3.4.5	Binding in sites IV and V and comparison to EcBE-M8	184
4.3.4.6	Comparison with EcBE-M8	186
4.4	Conclusion and Future Research Plans	191
	REFERENCES	194
 <b>CHAPTER 5: CLONING, EXPRESSION AND PURIFICATION OF POTATO TUBER ADP-GLUCOSE PYROPHOSPHORYLASE</b>		
		204
5.1	Introduction	204
5.2	Experimental Procedure	208
5.2.1	Potato small subunit ADP-Glc PPase	208
5.2.1.1	Vectors	208
5.2.1.2	Over-expression	209
5.2.1.3	Purification	209
5.3	Discussion and Conclusion	214
	REFERENCES	215
 <b>CHAPTER 6: EXPERIMENTAL PROCEDURES – <i>EcBE</i> AND RBEI</b>		
		218
6.1	Vectors	218
6.2	Overexpression and Purification	220
6.3	Glucan-Iodine Complex Activity Assay	224
6.4	Chain Length Specificity Assay	226
6.5	EcBE Mutants	228
6.6	Proteins Crystallization and Substrate Soaking	231
	REFERENCES	232

## LIST OF TABLES

Table 1.1	Molecular weights and amino acid counts of rice branching enzyme isoforms	24
Table 2.1	Outline of suggested mutations for the seven binding sites discovered based on EcBE / linear saccharides complex	64
Table 2.2	Multiple mutations within a binding site or across sites	65
Table 2.3	Effect of varying protein amount on specific activity	67
Table 2.4	Variation of specific activity of native N112BE with increase in amylose amounts	68
Table 2.5	Specific activity of N112BE at different amylose concentrations	69
Table 2.6	Iodine assay statistics of the N112BE wild type and mutants	72
Table 2.7	Combination mutations tested from all binding sites	81
Table 2.8	Comparing truncated and full-length mutants' percent activity	84
Table 3.1	Data collection and refinement statistics	108
Table 3.2	Summary of binding sites, molecule occupied, glucose chain name and length, and the residues involved in the interactions	114
Table 4.1	Dissociation constants for multiple $\alpha$ -glucans and potato tuber BEI	156
Table 4.2	Data collection and refinement statistics	160
Table 4.3	Residues involved in the interactions in binding sites I, IV and V of RBEI-M12	193
Table 6.1	Mutations in the N112BE protein	228

## LIST OF FIGURES

Figure 1.1	Scanning electron micrograph of starch granules inside a barley grain	1
Figure 1.2	Schematic diagram of starch granule structure	4
Figure 1.3	Chemical structures of the components of starch	5
Figure 1.4	Amylose and glycogen representation	6
Figure 1.5	Starch and glycogen biosynthesis pathway	8
Figure 1.6	Reactions of the members of the $\alpha$ -amylase family	15
Figure 1.7	Sequence alignment showing the four conserved regions and all the conserved amino acids in the catalytic center of the $\alpha$ -amylase family of enzymes	17
Figure 1.8	The crystal structure of N113 truncated EcBE	21
Figure 1.9	Schematic diagram of the amylopectin synthesis in rice	28
Figure 1.10	RBEI, crystal structure and overlay with EcBE	31
Figure 1.11	Overlay of EcBE with MycoBE	33
Figure 2.1	EcBE apo and holo	52
Figure 2.2	Binding sites I and II of the EcBE / maltoheptaose complex	54
Figure 2.3	EcBE – M7 binding at sites III and IV	56
Figure 2.4	Interactions at binding sites V and VI	57
Figure 2.5	The overall structure of EcBE in complex with cyclodextrins	59
Figure 2.6	Binding site VII interactions	60
Figure 2.7	Absorbance versus time for the native N112BE	70
Figure 2.8	Distribution of oligomers	87
Figure 3.1	Crystal structure of EcBE	106

Figure 3.2	Composite representation of EcBE – M8 complex showing the relative positions of the nine binding sites	111
Figure 3.3	Binding sites in each of the four molecules per asymmetric unit	112
Figure 3.4	Binding sites I and II detailed interactions	117
Figure 3.5	Detailed interactions in binding sites III and VIII	120
Figure 3.6	Detailed interactions in binding sites IV	122
Figure 3.7	Detailed interactions in binding site V	122
Figure 3.8	Detailed interactions in binding site IX	125
Figure 3.9	Detailed interactions in binding site X	126
Figure 3.10	CBM48 overlay EcBE, RBEI and AMPK	128
Figure 3.11	Overlay of all oligosaccharides bound in sites I and II in EcBE – M8 and EcBE – M7 complexes	131
Figure 3.12	Comparing EcBE, MycoBE and RBEI around binding site III	135
Figure 3.13	Binding site V compared	137
Figure 3.14	Binding sites IX compared	137
Figure 3.15	Binding sites occupied in one molecule of EcBE – M8	138
Figure 3.16	Binding sites I, IV, and VII relation explained	144
Figure 4.1	Overall structure of RBEI in complex with M12	162
Figure 4.2	Surface depiction of RBEI in complex with M12	163
Figure 4.3	Binding site I: detailed interactions between the sugar and RBEI	165
Figure 4.4	Oligosaccharide binding to RBEI and its symmetry related molecule	166
Figure 4.5	Binding sites IV and V: detailed interactions between the oligosaccharide and RBEI	170
Figure 4.6	Loops around the active site in GH13 family enzymes	172

Figure 4.7	Overlay of active site residues in GH13 enzymes	174
Figure 4.8	Overlay of CBM48 domains of RBEI-M12 and AMPK-beta-CD	179
Figure 4.9	Overlay of RBEI-M12 and RBEI-M5	182
Figure 4.10	Dodecaose molecule modeled in binding site I	183
Figure 4.11	RBEI and CGT overlaid around the active site	185
Figure 4.12	Overlay of RBEI-M12 and EcBE-M8	188
Figure 4.13	Comparison of RBEI and EcBE	189
Figure 5.1	The three step process of starch / glycogen biosynthesis	206
Figure 5.2	ADP-Glc PPase homotetramer	207
Figure 5.3	Sample purification of ADP-Glc PPase	211
Figure 5.4	Anti-MBP and anti-His western blots of ADP-Glc PPase purification	212
Figure 5.5	Gel filtration chromatography run of the elution sample	213
Figure 6.1	RBEI purification (N and C termini truncated)	222
Figure 6.2	N112BE (EcBE truncated at N-terminus)	223

## KEY TO ABBREVIATIONS

Aamy	<i>Aspergillus Oryzae</i> alpha-amylase
ADP-Glc	ADP-glucose
ADP-Glc PPase	ADP-glucose pyrophosphorylase
AMPK	5'-adenosine monophosphate-activated protein kinase
APS	Advanced Photon Source
BE	Branching enzyme
BEI, IIa, IIb	Branching enzyme I, IIa, IIb
BL assay	Branch-linkage assay
CAZy	Carbohydrate active enzymes
CBM48	Carbohydrate binding module 48
CGT	Cyclodextrin glycosyl transferase
<i>E. coli</i>	<i>Escherichia coli</i>
EcBE	<i>Escherichia coli</i> branching enzyme
FL-EcBE	Full-length <i>Escherichia coli</i> branching enzyme
FPLC	Fast protein liquid chromatography
GBE	Glycogen branching enzyme
GH13	Family 13 of glycosyl hydrolase enzymes
Glc, Glc-1-P	Glucose, Glucose-1-phosphate
GS	Glycogen synthase
GSD IV	Glycogen storage disease, type IV

HPAEC	High-performance anion exchange chromatography
isoA	<i>Pseudomonas amyloclavata</i> isoamylase
mBEI, IIa, IIb	Maize branching enzyme I, IIa, IIb
MycoBE	<i>Mycobacterium tuberculosis</i> branching enzyme
N112BE	EcBE with N-terminus 112 amino acids cloning truncation
N113BE	EcBE with N-terminus 111-113 amino acids Proteinase K
RBE-I	Rice branching enzyme-I gene (includes RBE1, 2a and 2b)
RBEI, IIa, IIb	Rice BEI, IIa, IIb
RBE1, 2a, 2b	Rice BE1, 2a, 2b (mean same as in roman letters, used
SBE, SS	Starch branching enzyme, Starch synthase
SBEI, IIa, IIb	Starch branching enzyme I, IIa, IIb
SSI, IIa, IIIa	Starch synthase I, IIa, IIIa

### Units and Other Symbols

Å	Ångström
bp	Base pair
CCP4	Collaborative computational project, number 4
DNA	Deoxyribonucleic acid
DP	Degree of polymerization
Da	Dalton
DTT	Dithiothreitol
hr(s)	Hour(s)

IPTG	Isopropyl-1-thio- $\beta$ -D-galactopyranoside
L	Liter
LB	Luria broth
mL	Milliliter
$\mu$ L	Micro-Liter
M	Molar
M5, M6, M7, M8	Oligosaccharide with the respective number of Glc units
M12	Dodecaose
Min	Minutes
mg	Milligrams
Mol.	Molecule
MW	Molecular weight
MWCO	MW cutoff
nM	Nano molar
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
pM	Pico molar
R-factor	Reliability factor
rmsd	Root mean square deviation
RNA	Ribonucleic acid
rpm	Rotation per minute

SDS-page	Sodium dodecyl sulfate – polyacrylamide gel electrophoresis
Tris	2-Amino-2-(hydroxymethyl)-1,3-propanediol
UV	Ultraviolet light
Vis	Visible light
wt	Wild type

### **Amino Acids**

Ala, A	Alanine
Arg, R	Arginine
Asn, N	Asparagine
Asp, D	Aspartic acid
Cys, C	Cysteine
Gln, Q	Glutamine
Glu, E	Glutamic acid
Gly, G	Glycine
His, H	Histidine
Ile, I	Isoleucine
Leu, L	Leucine
Lys, K	Lysine
Met, M	Methionine
Phe, F	Phenylalanine
Pro, P	Proline

Ser, S

Serine

Thr, T

Threonine

Trp, W

Tryptophan

Tyr, Y

Tyrosine

Val, V

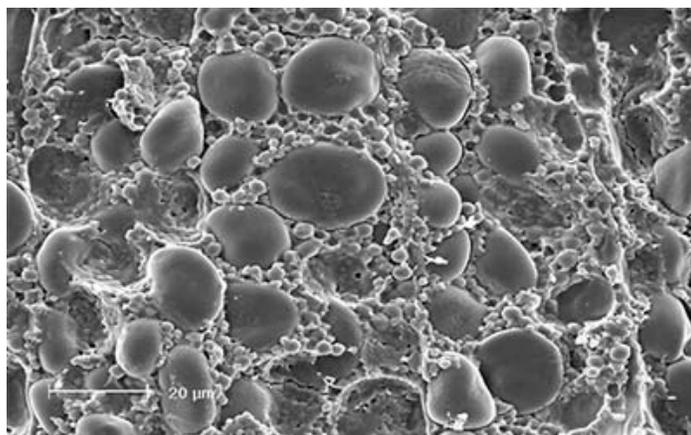
Valine

## CHAPTER 1: LITERATURE REVIEW

### 1.1 Starch versus Glycogen

#### 1.1.1 Natural occurrence and function

Starch, the major polysaccharide storage molecule in plants, is found in both photosynthetic and non-photosynthetic tissues and consists of two main types. The first is the “transitory starch”, whose levels vary in the chloroplasts of plants’ leaves. This form acts as the ongoing source of energy for the plant. The second type, called “storage starch”, mainly accumulates in seeds, tubers, roots, corms, fruits and the rhizomes of plants. This form of starch is essential for plant survival during extreme conditions of heat or cold. <sup>1</sup> **Figure 1.1** depicts a scanning electron micrograph of starch granules inside a barley seed.



**Figure 1.1: Scanning electron micrograph of starch granules inside a barley grain.** <sup>2</sup>

Glycogen is, on the other hand, the major carbohydrate storage reservoir in animals and bacteria. It is synthesized in a number of bacteria during growth and in others during the stationary phase, and can form up to 50% of the total dry weight. In humans and animals, glycogen is synthesized after a meal containing carbohydrates and accumulates mainly in the liver hepatocytes, muscles and red blood cells. <sup>3</sup>

In biochemical classification, glycogen and starch are considered to be carbon and energy-storage compounds because they fulfill three criteria. <sup>3</sup> First, these compounds must be accumulated during times of energy surplus over basic cell needs. Second, they are put to use when exogenous energy supplies are incapable of providing maintenance energy levels for the cell such as for growth, division and viability. Third, the compounds must decompose to an energy form employable by the cell permitting it to endure the environment.

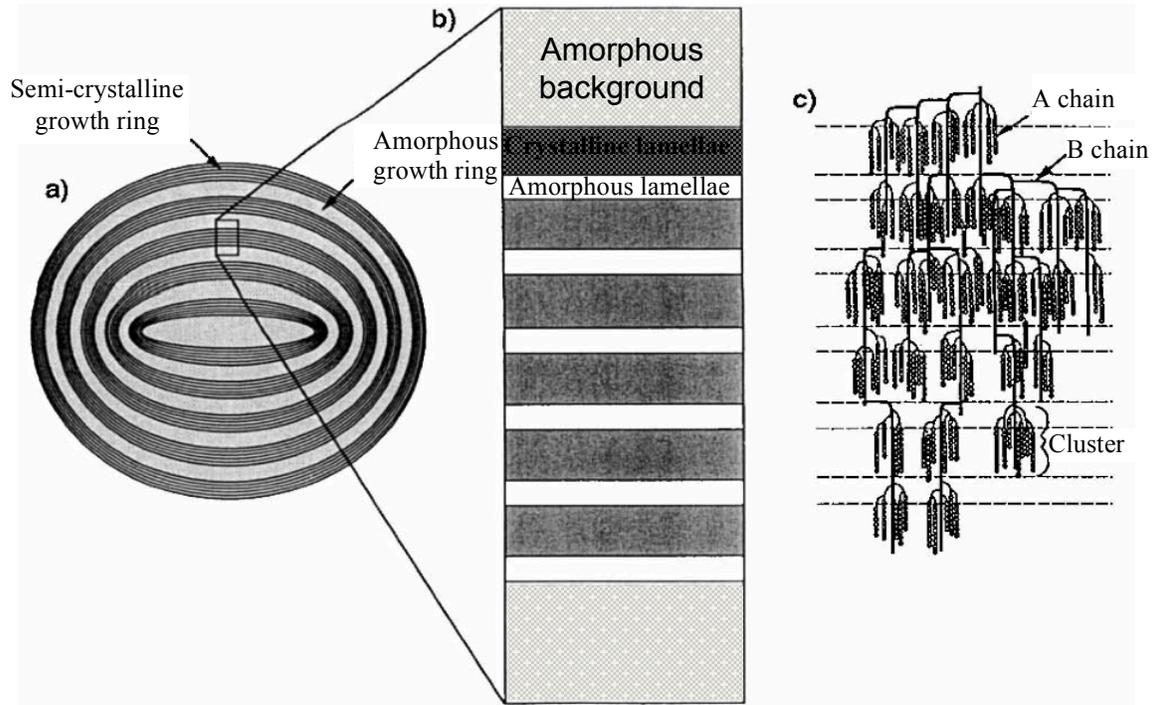
### **1.1.2 Structure and organization**

Starch is a gigantic quaternary complex (ranging from 0.1 to 50  $\mu\text{m}$  in diameter), mainly consisting of insoluble semi-crystalline material (**Figure 1.2.a**). In all species containing starch, the polymer is formed of two major constituents: amylose (18-28%) and amylopectin (72-82%). <sup>1</sup>

Amylose is essentially a linear polymer formed of only glucose units; where the sugar monomers are attached by  $\alpha$ -1,4-D-glycosidic bonds (**Figure 1.3.a**). In the crystalline state, it exists in a left-handed single or double-stranded helix. <sup>4, 5</sup>

Each helical turn contains six glucose units (**Figure 1.4.a**).<sup>6, 7</sup> Amylopectin is also formed of only glucose units attached by the same types of bonds as in amylose, but is a highly branched polymer. The number and lengths of branches differ between organisms.<sup>1</sup> **Figure 1.2.c** represents the model of Meyer,<sup>8</sup> the accepted description of the structure of amylopectin. The inner branches B develop from the main chain C, and lead to the outer chains A; which are more numerous than chains B. According to the Manners review<sup>8</sup> that explains the models of Meyer, Haworth and Staudinger, the ratio of chains A:B ranges from 1.0:0 to 1.5:1.

Unlike starch, glycogen has a highly branched polymeric structure formed of  $\alpha$ -1,4-D-glycopyrannose units attached by 1,6-glycosidic linkages.<sup>9-12</sup> In 1941, the structure of glycogen was for the first time determined to be multiply branched based on enzymatic analysis and in 1951, it was confirmed that the average chain length is 10-14 units of glucose; the polymer consists of more than five-thousand individual chains and with equal numbers of exterior and interior chains.<sup>10, 11, 13, 14</sup> The glycogen granule is therefore very dense. **Figure 1.4.b** depicts the structure of glycogen showing the glucose units in circles emphasizing the compact architecture of the polymer.



**Figure 1.2: Schematic diagram of starch granule structure.**<sup>15</sup> **a.** Whole granule, consisting of alternating amorphous (amylose) and semi-crystalline (amylopectin) rings. **b.** Internal structure enlarged: stacks of amorphous and crystalline lamellae form the semi-crystalline growth rings. **c.** The currently accepted structure of amylopectin within the semi-crystalline growth ring: A chains form double helices regularly packed within the crystalline lamellae, while B chains interconnect A chains; the origin of A and B branches essentially reside in the amorphous lamellae.

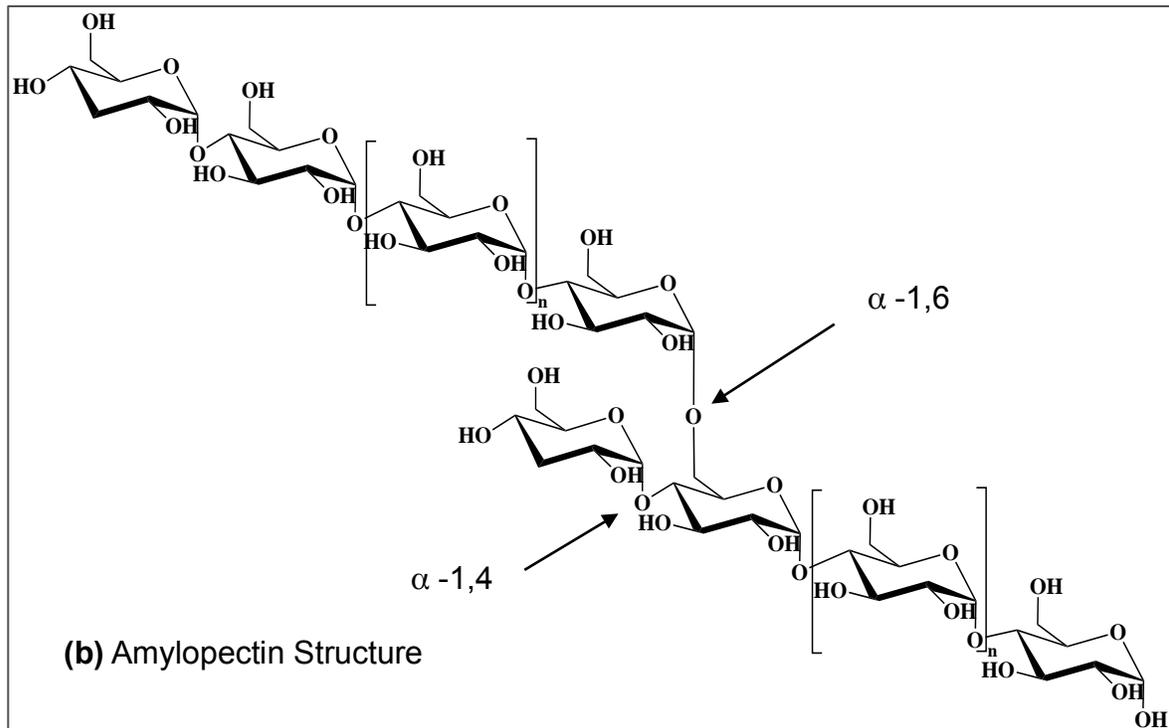
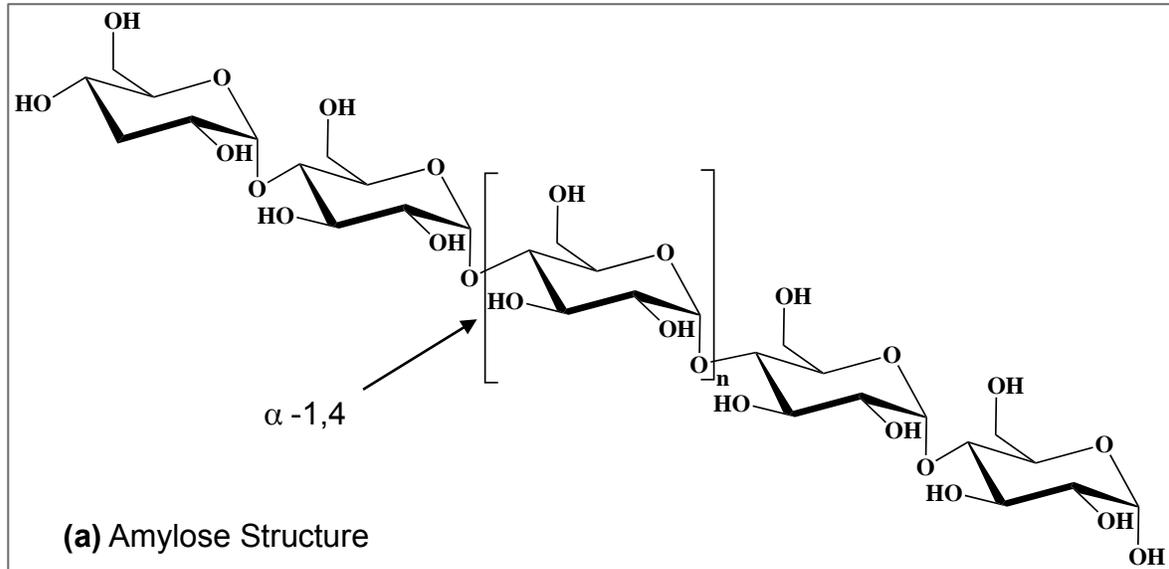
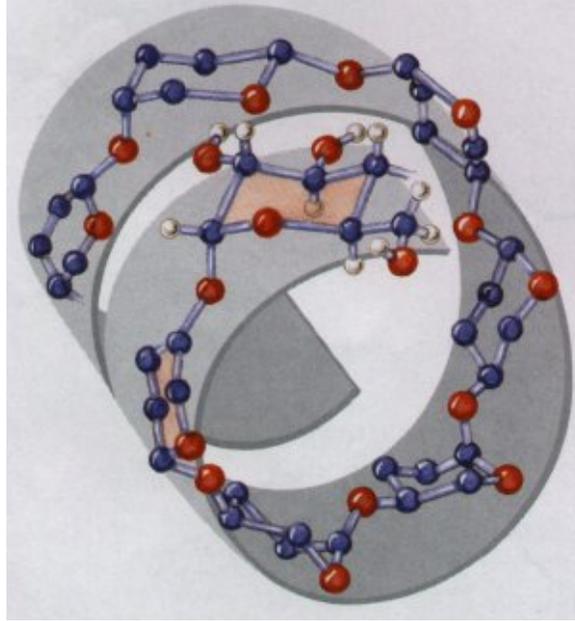
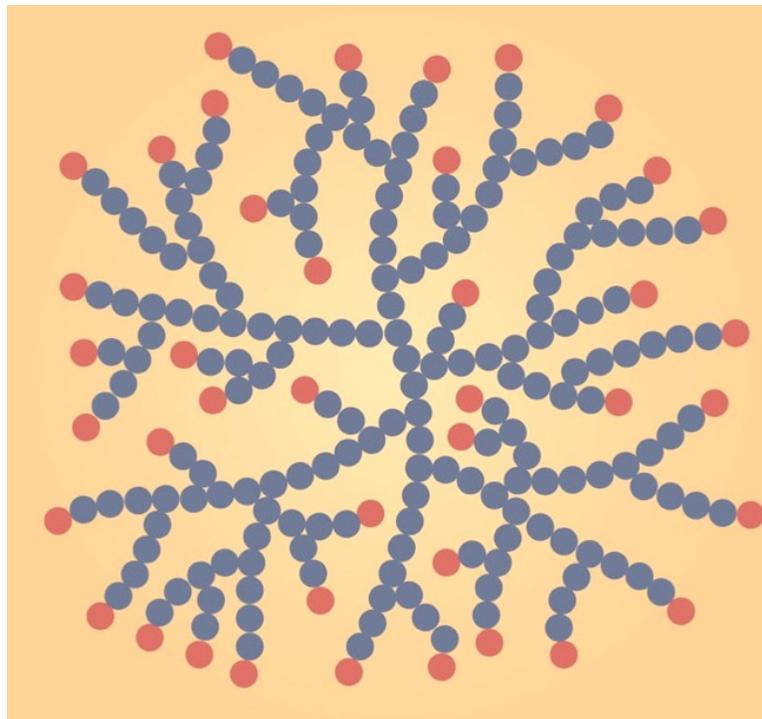


Figure 1.3: Chemical structures of the components of starch. <sup>16</sup>

(a)



(b)



**Figure 1.4: Amylose and glycogen representation.** a. Amylose helix: the  $\alpha$ -1,4 glycosidic linkages cause a tightly coiled structure making dense granules of stored starch or glycogen.<sup>17</sup> b. Sketch representation of glycogen: the glucose units are shown in blue circles except for the terminal units in red.<sup>18</sup>

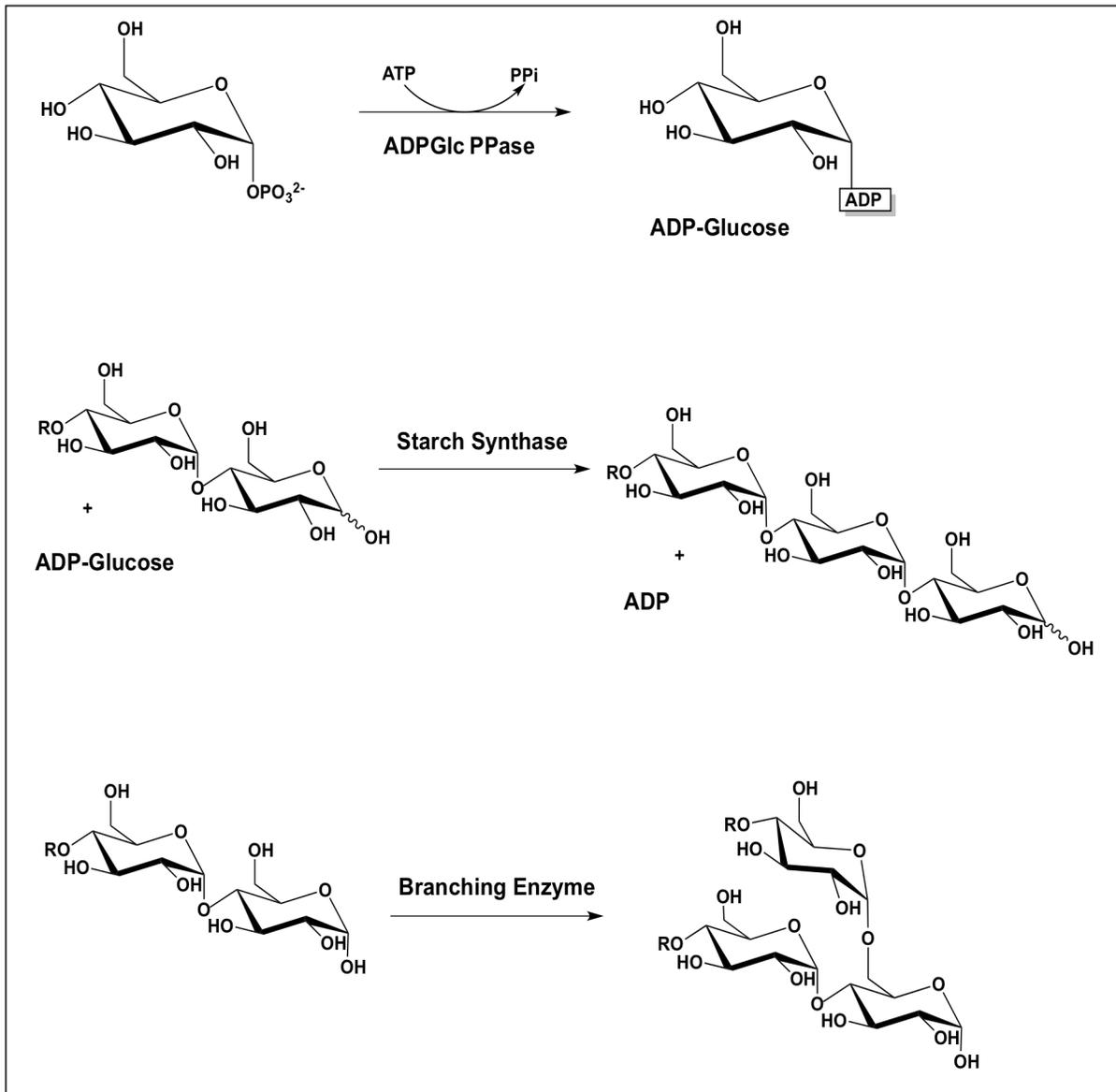
### 1.1.3 Biosynthesis

The biosynthesis of starch / glycogen *in vivo* involves three major reactions **(Figure 1.5)**.<sup>19-22</sup> The first step results in the synthesis of ADP-Glucose from a precursor molecule, Glucose-1-Phosphate, the rate-determining step. This reaction uses one ATP molecule and generates one pyrophosphate. ADP-Glucose Pyrophosphorylase (ADP-Glc PPase) catalyzes this step; it is an allosteric enzyme.<sup>20, 23, 24</sup> ADP-Glc PPase is well characterized in several plants like spinach, maize, and especially potato tuber as well as in *E. coli* and the allosteric regulation has been studied in detail.<sup>23-36</sup> In addition, the crystal structure of the enzyme from potato tuber was obtained by our lab.<sup>37</sup>

The second step of the process results in the elongation of the polysaccharide chain. Starch synthase or glycogen synthase (SS or GS) catalyze this reaction; which builds up the polymer by transferring a glucose unit from adenosine diphosphate glucose (ADP-Glc) to an acceptor glucan chain, retaining the  $\alpha$ -1,4 configuration in the newly added glucose unit. Several SS's, including those from wheat, maize, barley, potato tuber and rice, GS's from *Escherichia coli* (*E. coli*), yeast and several other bacteria, have been studied and characterized in detail.<sup>38-50</sup> The crystal structures of the rice barley, *E. coli*, and *Agrobacterium tumefaciens*, synthases have been solved.<sup>40, 51-53</sup>

The last reaction in the biosynthesis is the branching of the linear polymer; this step is catalyzed by a starch or glycogen branching enzyme (SBE or GBE). BE breaks  $\alpha$ -1,4-glycosidic bonds and forms  $\alpha$ -1,6 linked branches. SBE and GBE

have been characterized in several organisms, including maize, and rice<sup>39, 47, 54-</sup>  
<sup>66</sup> and in *E coli* and a few other bacteria<sup>44, 49, 67-74</sup> respectively.



**Figure 1.5: Starch and glycogen biosynthesis pathway.**<sup>3, 20, 21</sup> The three-step process shows the reactions involving the enzymes ADP-Glucose Pyrophosphorylase, Glycogen or Starch Synthase and Branching Enzyme.

#### 1.1.4 Industrial connection and relevance

Starch from several sources has been used in industrial applications for decades. Although industries also avail from glycogen in bioregenerative sciences; it is definitely still in early development. Starches have been and continue to be extracted from maize, potato, wheat, rice, cassava and tapioca and used or chemically engineered post-harvest and in large scale for the food industries.<sup>75, 76</sup> Potatoes have been transgenically engineered in the last century<sup>76, 77</sup> because of the very high swelling power of potato starch, caused by the high level of covalently bound phosphate, giving it stable-paste properties.<sup>75</sup> More recently, starches other than from potato have begun to be engineered to further optimize starch production.<sup>75, 78</sup>

Starch in the native form is a thickener and a binder. Upon heating in water, starch goes through a three-step process: melting of amylopectin helices, granule swelling and finally, complete disintegration and solubility. Later cooling causes linear chains to aggregate and precipitate as a gel. Controlling this process is key for starch functionality; this currently happens by crosslinking glucan chains and addition of charged groups.<sup>75</sup> Starches then become tolerant of freeze-thawing processes and pH variation.<sup>75</sup>

Modified starches currently in use include waxy and resistant starches. Waxy starch is amylose-deficient; it is easily synthesized because a single gene, the granule bound starch synthase gene, is necessary and sufficient for amylose production. This type of starch gelatinizes readily and is beneficial as a food stabilizer, thickener and emulsifier for salad dressings. Starch can also be

modified to become a partial wax; which offers better pasta quality. Waxy starches provide freeze-thaw stability.<sup>75</sup> Resistant starch, on the other hand, contains 50 - 90% amylose (while the native form contains 20-30%).<sup>75</sup> This type is very useful for producing sweets, leaving crispy coating on fried products, supplying adhesive products for the manufacturing of corrugated board and paper, and the packaging and textile industries.<sup>76</sup> Resistant starch is fermented in the large intestine rather than digested in the small intestine; therefore it is transformed into short chain fatty acids such as butyrate that is beneficial for colon health.<sup>75</sup>

Researchers successfully attempted experiments to increase starch content and yield in transgenic plants by manipulating the enzyme adenylate kinase, responsible for the conversion of ATP and AMP into ADP, an essential precursor for starch biosynthesis.<sup>76, 77</sup> Others have worked on improving existing enzymes involved in starch synthesis and finding or fabricating other enzymes targeting the specific operative needs of starch and to increase the efficiency in the production of sugars from variable starches.<sup>79</sup> More studies continue to explore the molecular events of starch annealing in order to improve this process, which has a marked impact on starch functionality.<sup>80</sup> Yet other studies focus on performing modifications to the biosynthetic pathway, rather than chemically modifying and enzymatically treating starches after harvest, in order to discover novel starches with new properties for modernistic applications, including but not limited to making delivery vehicles that protect pharmaceutically active proteins from digestion, microcapsules for small molecules, and biodegradable films.<sup>75, 78</sup>

More efforts are recently directed towards biofuels production. Professor Ladisch at Purdue University developed a process to convert the fiber left in the corn kernel after the starch is extracted into ethanol; moreover, researcher Ho developed a new genetically modified yeast that converts polysaccharides present in sugarcane bagasse, wheat straw, woodchips, sweet sorghum, corn stover, kernel fiber, grasses and tree leaves, after being broken down to simple sugars, into ethanol.<sup>81</sup> These agricultural materials have the advantage of being inexpensive and renewable resources making them an excellent substitute feedstock for second-generation ethanol production.

Alternative approaches though still under development, constitute the production of transgenic plants or trees (for starch) and cyanobacteria or microalgae (for glycogen).<sup>82-84</sup> Trees contain high levels of cellulose, which scientists are looking into using for ethanol production.<sup>85</sup> Cellulose in trees is covered by lignin, giving them stiffness and resistance to pests. Moderate reduction of lignin gives easier enzyme access to cellulose to break it down into simple sugars.<sup>85-88</sup> In addition, microalgae represent another promising polysaccharide resource for biorefinery. Microalgae and cyanobacteria have several benefits, including rapid rate of reproduction and superior efficiency over plants in converting solar energy into chemical energy. They can also be grown in salt water and other harsh conditions, eliminating the need for fresh water. On the other hand, environmental conditions, such as proper light intensity and wavelength, carbon dioxide concentration, salinity and cultivation systems, need to be optimized for highest efficiency.<sup>83</sup> A rapidly-amplifying cyanobacterium,

*Arthrospira platensis*, for the production of glycogen, was studied for optimal glycogen production; which involved depleting the nitrate source and providing optimal light intensity to maximize. Glycogen yield reached 1.03 g/L, the highest polysaccharide production ever reported in microalgae.<sup>82</sup>

Understanding the starch and glycogen branching process in detail remains an important challenge. A detailed understanding of the branching enzyme structure, mechanism, and specificity, will reveal critical insights in the process of starch granule synthesis.

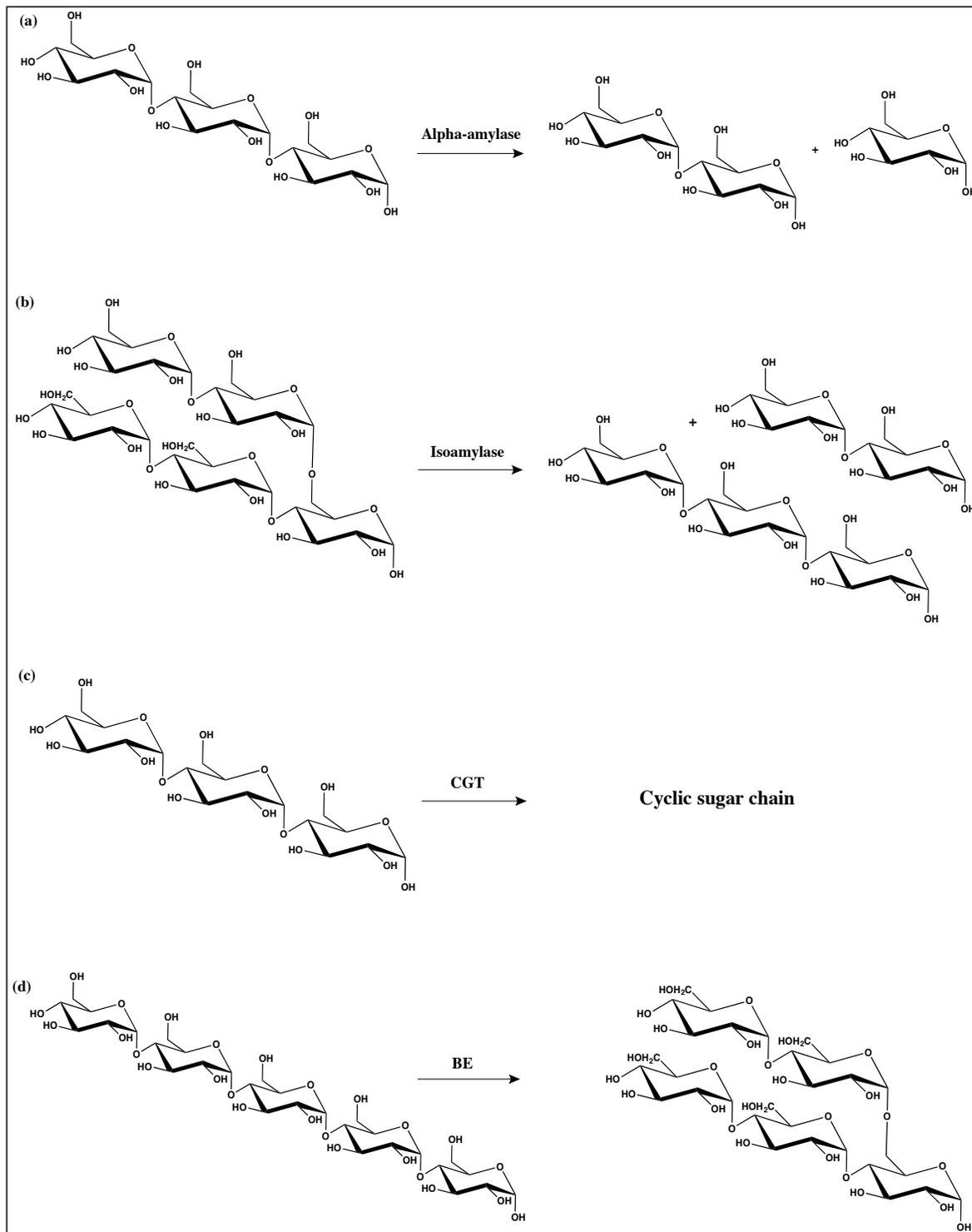
## 1.2 Branching Enzyme

### 1.2.1 Background

Branching Enzyme (1,4- $\alpha$ -D-glucan: 1,4- $\alpha$ -D-glucan 6- $\alpha$ -D-[1,4- $\alpha$ -D-glucano]-transferase; EC 2.4.1.18) (BE) exists in all organisms that make glycogen or starch. BE plays an important role in determining the final structure of these polysaccharides. The enzyme cleaves linear  $\alpha$ -1,4-glycosidic bonds and transfers these cleaved chains to form  $\alpha$ -1,6-branches. The formation of branching increases the non-reducing ends, making the polymer more reactive towards degradation and the synthetic enzymes providing rapid glucose release and rapid polymer rebuilding for storage when needed. Branched carbohydrates are substantially more soluble than linear glycogen or starch; which can precipitate. Mutations in the gene of human BE causes accumulation of insoluble glycogen; which is known as glycogen storage disease type IV (GSD IV).<sup>89, 90</sup> GSD IV affects the liver, skeletal muscle and the central and peripheral nervous system.

BE is a member of the  $\alpha$ -amylase family of enzymes or family 13 glycosyl hydrolases (GH13);<sup>91</sup> which also encompasses  $\alpha$ -amylase, pullulanases, isoamylases, and cyclodextrin glucanotransferases (CGT). All of these enzymes hydrolyze  $\alpha$ -glycosidic bonds to generate  $\alpha$ -anomeric saccharides, produce  $\alpha$ -1,4 or  $\alpha$ -1,6 glycosidic linkages, or perform both functions (**Figure 1.6**).<sup>92, 93</sup> Besides the similarity in activity, members of the  $\alpha$ -amylase enzyme family also share an important structural feature: they possess a central TIM barrel domain containing 8  $\alpha$  helices and 8  $\beta$  sheets; which constitutes the catalytic site.<sup>91</sup>

Moreover, branching enzyme has two other domains not shared with most of the other enzymes of GH13: an amino-terminal domain and a carboxy-terminal domain.<sup>94</sup> Chimeric mutants involving swapping of domains constructed in order to understand the function of each domain of BE revealed that the N-terminal domain is associated with dictating the size of the glucan chain transferred while the carboxy-terminal domain is involved in substrate specificity and catalytic capacity.<sup>64</sup>



**Figure 1.6: Reactions of the members of the  $\alpha$ -amylase family.** **a.**  $\alpha$ -amylase cleaves only  $\alpha$ -1,4 bonds. **b.** Isoamylase saccharifies only  $\alpha$ -1,6 bonds. **c.** Cyclodextrin glucanotransferase enzyme catalyzes the formation of cyclodextrins. **d.** Branching enzyme catalyzes the formation of  $\alpha$ -1,6 bonds.

### 1.2.2 Branching enzyme isoforms

In the majority of glycogen-producing organisms such as bacteria, fungi and higher mammals, branching enzyme exists in a single isoform; however, in most plants, the enzyme is found in three isoforms, starch branching enzyme I (SBEI), starch branching enzyme II a (SBEIIa) and II b (SBEIIb).<sup>55, 59, 65, 72, 95-97</sup> Although all BE isoforms share structural features including the three domains: an N-terminal, a C-terminal and a GH13 ( $\alpha/\beta$ )<sub>8</sub> barrel domain; SBE and GBE have fairly low sequence homology. For instance, the protein sequences of maize SBE (mBE) and *Escherichia coli* GBE (EcBE) are only 25% identical.<sup>72, 98</sup> Despite the low sequence similarity between SBE and GBE, both forms of the enzyme share with members of the  $\alpha$ -amylase family the four highly conserved regions identified by Nakajima et al. in 1986,<sup>98</sup> all confined within the catalytic domain (**Figure 1.7**). Jespersen *et al.* and Svensson *et al.* identified in the early 90s the residues that are responsible for substrate binding and catalysis (**Figure 1.7**).<sup>99-</sup><sup>101</sup> Another resemblance between SBE and GBE is their close molecular mass (84 kDa for GBE; 86 kDa for SBEI, and 84 kDa for SBEII).<sup>72</sup>

These differences in protein sequence are the basis for the unique actions of the branching enzyme isoforms revealed in the specificity for the length of the glucan chain transferred.<sup>94</sup> GBE from *E. coli* has a preference for transferring chains between 5 and 16 glucose units, but most predominantly for 12 units;<sup>102,</sup><sup>103</sup> while SBEI transfers chains up to 30 (in maize) or 40 (in rice) glucose units, preferentially 10-13 units in maize, SBEIIa in rice transfers a wide range of

shorter chains varying from 6 to 15 glucose units, and SBEII (a and b in maize) and SBEIIb (in rice) preferentially transfers 6 to 7 glucose units.<sup>64, 65, 72, 104</sup>

<b>EcBE</b>	297	S	W	G	Y	Q	P	T	
<b>Hum</b>	248	S	F	G	Y	Q	I	T	
<b>mBEI</b>	303	S	F	G	Y	H	V	T	
<b>mBEII</b>	277	S	F	G	Y	H	V	T	
<b>isoA</b>	253	Y	W	G	Y	M	T	E	
<b>Aamy</b>	79	Y	H	G	Y	W	Q	Q	
<b>CGT</b>	97	Y	H	G	Y	W	A	R	
<b>EcBE</b>	331	N	V	I	L	D	W	V	Region 1
<b>Hum</b>	282	I	V	L	L	D	V	V	
<b>mBEI</b>	337	R	V	L	M	D	V	V	
<b>mBEII</b>	311	L	V	L	M	D	V	V	
<b>isoA</b>	288	K	V	Y	M	D	V	V	
<b>Aamy</b>	113	Y	L	M	V	D	V	V	
<b>CGT</b>	131	K	V	I	I	D	F	A	
<b>EcBE</b>	338	P	G	H	F	P	T	D	
<b>Hum</b>	289	H	S	H	A	S	K	N	
<b>mBEI</b>	346	H	S	H	A	S	N	N	
<b>mBEII</b>	318	H	S	H	A	S	S	N	
<b>isoA</b>	301	Y	N	H	T	A	E	G	
<b>Aamy</b>	120	A	N	H	M	G	Y	D	
<b>CGT</b>	138	P	N	H	T	S	P	A	
<b>EcBE</b>	398	I	D	A	L	R	V	D	Region 2
<b>Hum</b>	350	F	D	G	F	R	F	D	
<b>mBEI</b>	408	F	D	G	F	R	F	D	
<b>mBEII</b>	379	F	D	G	F	R	F	D	
<b>isoA</b>	368	V	D	G	F	R	F	D	
<b>Aamy</b>	199	I	D	G	L	R	I	D	
<b>CGT</b>	222	I	D	G	I	R	M	D	
<b>EcBE</b>	455	T	M	A	E	E	S	T	Region 3
<b>Hum</b>	409	T	I	A	E	D	V	S	
<b>mBEI</b>	467	V	V	A	E	D	V	S	
<b>mBEII</b>	438	T	I	G	E	D	V	S	
<b>isoA</b>	480	Y	V	T	E	D	A	N	
<b>Aamy</b>	227	C	I	G	Q	V	L	D	
<b>CGT</b>	254	T	F	G	E	W	F	L	
<b>EcBE</b>	523	L	S	H	D	E	V		Region 4
<b>Hum</b>	478	E	S	H	D	Q	A		
<b>mBEI</b>	478	E	S	H	D	Q	A		
<b>mBEII</b>	506	E	S	H	D	Q	A		
<b>isoA</b>	507	D	V	H	D	G	M		
<b>Aamy</b>	294	E	N	H	D	N	P		
<b>CGT</b>	325	D	N	H	D	M	E		

**Figure 1.7: Sequence alignment showing the four conserved regions and all the conserved amino acids in the catalytic center of the  $\alpha$ -amylase**

**Figure 1.7 (cont'd) family of enzymes.** Four regions in the catalytic domain numbered 1, 2, 3 and 4 shared between the members of the  $\alpha$ -amylase family of enzymes identified in 1986 are highly conserved; the majority of the residues in each region is shown and highlighted in blue boxes.<sup>98</sup> Eight conserved amino acids in the catalytic center are enclosed in black boxes and constitute the residues involved in substrate binding very close to the active site or catalysis.<sup>99-</sup><sup>101</sup> The enzymes abbreviated respectively are: *E. coli* branching enzyme (EcBE), human branching enzyme (Hum), maize BEI (mBEI) and maize BEII (mBEII), isoamylase from *Pseudomonas amyloclavata* (isoA),  $\alpha$ -amylases from *Aspergillus Oryzae* (Amy), and cyclodextrin glucanotransferase from *Bacillus Circulans* (CGT).

### **1.2.3 *Escherichia coli* branching enzyme (EcBE)**

The native EcBE consists of a single polypeptide comprising 728 amino acids and has a molecular weight of 84 KDa.<sup>105, 106</sup> The protein was over-expressed and purified to near homogeneity.<sup>70, 72</sup> The activity of the wild type EcBE was measured using three different assays,<sup>72, 96</sup> the phosphorylase stimulation assay, iodine staining assay, and branching linkage assay, described by Guan and Preiss.<sup>96</sup> In order to better understand the structure, mechanism, and specificity of BE, researchers attempted to crystallize the full-length protein, but all attempts were unsuccessful; which lead to pursuing a stable digest of EcBE. Limited proteolysis employing proteinase K gave a 71.6 KDa product, which retained 40-60% of wild type branching activity depending on the kind of the assay utilized.<sup>68</sup> The sequence of the truncated enzyme obtained lacked the first 111 or 113 amino acids (will be referred to as N113BE). Although the activity was compromised, N113BE had a substrate preference and  $K_m$  value similar to those of the native full-length enzyme.<sup>69</sup> A recombinant branching enzyme was

constructed with a deletion of the first 112 residues. This construct had identical properties to those of the proteolyzed product.<sup>68</sup> As mentioned above, EcBE has a preference for transferring chains between 5 and 16 glucose units, with predominance for 12 units;<sup>102, 103</sup> however, N113BE had a notably altered preference for chain transfer with prevalence for longer chains of 15-20 glucose units.<sup>69, 71</sup> EcBE with shorter deletions at the N-terminus (63- and 83-amino acids) transferred intermediate length chains varying between 10 and 20.<sup>71</sup> These observations prove that progressive shortening of the N-terminal domain result in progressive accession of chain-transferred length. Analysis of these findings elucidates an important role of the wild type protein N-terminus that is affecting the length of the chain transferred by supporting the glucan substrate during the reaction mechanism.<sup>71</sup>

Determining the crystal structure of EcBE in 2002 was a breakthrough in the glycogen and starch biosynthesis area of study because it was the first structure of BE from any organism and the first among the three enzymes involved in the biosynthetic pathway.<sup>107</sup> The structure of the recombinant EcBE with an N-terminal truncation of 112 amino acids (N112BE) shows three major domains, an N-terminal domain formed of seven  $\beta$ -strands making a sandwich-like shape containing 128 amino acids, a C-terminal domain consisting of 116 amino acids, and a central ( $\alpha/\beta$ ) barrel domain comprising 372 residues.<sup>94</sup> **Figure 1.8** depicts the N112BE structure with the N-terminal domain in red, catalytic domain in orange, and C-terminal in blue.<sup>94</sup> An acarbose molecule, “a pseudo-oligosaccharide resembling amylose with a linked hydroxymethylconduritol unit

and a 4-amino-4-deoxy-D-chinovose residue extended by two glucose units”, was modeled into the active site of N112BE in order to shed some light into the interaction between the enzyme and the substrate.<sup>94</sup> Crystals of N112BE were reproduced and soaked in a number of oligosaccharides sequentially increasing in length from maltose to maltoheptaose, and in  $\alpha$ -  $\beta$ - and  $\gamma$ -cyclodextrins.<sup>92</sup> A total of seven binding sites were identified in EcBE, one of them being exclusive for cyclodextrins, by virtue of similarity to the helical structure of amylose.<sup>92</sup> This discovery gave insight into the mode of binding of the substrate to the enzyme.

In the work presented here, we were successful in obtaining crystals soaked with maltooctaose and were able to both confirm seven binding sites identified in the original study, but also identify four new glucan binding sites as well. In the following chapters, we present this work, compare the current structure of EcBE with the previous ones, and propose a hypothesis on the way the enzyme interacts with its substrate throughout the catalysis.



**Figure 1.8: The crystal structure of N113 truncated EcBE.** Crystal structure of N113 truncated *E. coli* BE depicted in a ribbon form with red for the N-terminal domain, orange for the catalytic domain, and blue for the C-terminal domain. Residues involved in catalysis are distinguished in green, with atoms colored by kind: red for oxygen, green for carbon, and blue for nitrogen.

## 1.2.4 Rice branching enzyme I

### 1.2.4.1 Background on starch branching enzyme

Starch branching enzyme is more complex than the glycogen protein. SBE has been isolated and purified or the gene expressed in *E. coli* cells for characterization from several sources including maize,<sup>39, 55, 57, 59, 60, 63, 64, 72, 96, 108</sup> rice,<sup>62, 65, 109-112</sup> wheat,<sup>113-116</sup> barley,<sup>117-119</sup> potato,<sup>120, 121</sup> cassava,<sup>122</sup> pea and soybean.<sup>123-125</sup> In green plants, there exists three different isoforms of SBE,

branching enzyme I (BEI), branching enzyme IIa (BEII), and branching enzyme IIb (BEIIb).<sup>65, 95, 96, 109, 112, 126</sup> These are distinguishable in their reactivity towards experimental substrates such as cyclodextrins,<sup>121</sup> certain chemicals such as inorganic phosphate and phosphorylated compounds,<sup>113</sup> temperature dependence,<sup>108</sup> propensity to associate with or become irreversibly bound to starch granules,<sup>127</sup> and expression levels during growth of reserve tissues.<sup>109, 113, 128</sup> They also differ in substrate preference; in vitro experiments showed that maize BEI mainly acts on amylose with a smaller number of branches than do BEII that preferentially branches amylopectin.<sup>72, 96, 108</sup> Studies on the chain length transferred in maize isozymes of the protein revealed more diversity among them; BEI essentially transfers chains with degree of polymerization (DP) larger than 10, while BEII predominantly transfers shorter chains with DP 3-9 but mostly DP6 and DP7.<sup>72</sup>

#### **1.2.4.2 Starch branching isozymes in rice**

In immature rice seeds, the branching enzyme exists in multiple isoforms: rice branching enzyme 1, 2a, 2b, 3, 4a and 4b termed RBE1, RBE2a, RBE2b, RBE3, RBE4a and RBE4b. They are different in several aspects although some share high areas of resemblance. RBE1 and RBE2b have a molecular weight (MW) of 82 KDa, while RBE2a MW is 85 KDa (**Table 1.1**). Sequence analysis demonstrates that each of these three isoforms has two amino-terminal sequences that are a repeat of each other with the exception that one is lacking

two residues present at the N-terminus of the other, that their catalytic domain is highly homologous, while the carboxy terminal is poorly conserved.<sup>110</sup> In addition, Mizuno *et al.* noticed that latter steps of purification of these three isoforms show significant decrease in the ratio of RBE2a to RBE2b compared to earlier steps.<sup>110</sup> Therefore, they have hypothesized that RBE2a is converted into RBE2b by removal of a carboxy-terminal fragment based on this data and on the expression product of a constructed plasmid carrying rice BE-I fragment (RBE-I, designation given to RBE1, 2a and 2b combined and will be used hereafter) at nucleotides 215-2739, which showed two protein bands immunoreactive with anti-maize BE-I antibody and almost identical in gel electrophoresis mobility (unpublished data of Mizuno *et al.*).<sup>110</sup> In addition, western blot analysis using anti-maize BE-I antibody exhibited greater immunoreactivity of RBE1, RBE2a and RBE2b than the other isozymes of RBE. The three isoforms RBE1, 2a, and 2b are then identical in molecular weight except for an additional 3 KDa for RBE2a, in the amino-terminal sequence and in immunoreactivity towards anti-maize BE-I antibody.<sup>110</sup> Also, their N-terminal sequence matched that deduced from the inserts of cDNA clones encoding rice BE-I.<sup>110</sup> This knowledge strongly indicates that all three isoforms RBE1, 2a and 2b are the product of the same gene.<sup>62, 110, 111</sup> In fact the 820-residue precursor RBE-I gene is later differentiated into RBE1 after transportation into the amyloplast by removal of the 64-residue transit peptide at the N-terminus.<sup>110, 129</sup> It was suggested that the heterogeneity of the three isozymes arise from a post-translational modification of the synthesized precursor gene.<sup>110</sup> The N-terminal leader region is rich in hydroxylic

(23% Ser and Thr) and basic (14% Arg and Lys) residues and contains only one acidic amino acid at position 44.<sup>110</sup> This composition is similar to the transit peptide of potato pyrophosphorylase and shares 53 % sequence identity with that of the maize protein.<sup>110</sup> The overall structure of RBE-I is 86 % identical to the maize protein, although very poor conservation of the C-terminus, 81 % to the potato branching enzyme, and only about 23 % to the *E coli* BE;<sup>110</sup> however, they all share the four regions conserved among the  $\alpha$ -amylase family members identified by Nakajima *et al.*<sup>98</sup>

RBE isoform	MW	Amino acid count	
		Immature form	Mature form
1	82	820	755
2a	85	820	
2b	82	820	
3	87	825	760
4a	93	841	--
4b	83	--	788

**Table 1.1: Molecular weights and amino acid counts of rice branching enzyme isoforms.** While RBE1, 2a and 2b are expressed in the rice endosperm from the same gene as an 820-residue precursor protein, and further differentiated into the three isoforms,<sup>110</sup> RBE3 is an independent gene specifically expressed in immature seeds,<sup>110</sup> and RBE4 exists in two forms, RBE4a and 4b, in the immature seeds, although RBE4b is proven to be a processed form of the recombinant RBE4a.<sup>111</sup>

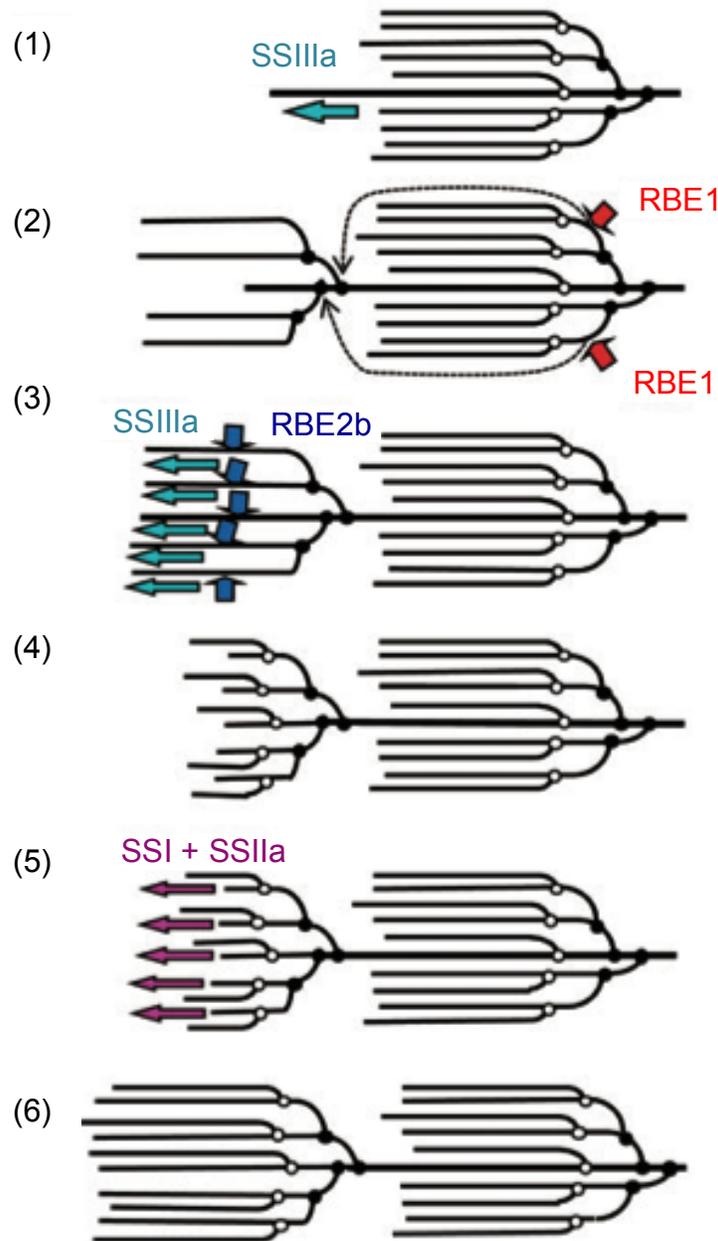
Even though RBE1 and RBE2 are the predominant forms of the enzyme, RBE3 and RBE4 have proven to be distinct enzymes. RBE3 is initially produced as a precursor protein formed of 825 residues with a molecular weight of 87 KDa and is further processed into the mature 760-residue peptide (**Table 1.1**). To compare RBE isoforms, all were subjected to western blot experiments using anti-maize BE-I antibody and all RBE bands showed reactivity. Western blot analysis using affinity-purified anti-RBE3 antibody showed very little reactivity towards RBE1, 2a and 2b but strong reactivity towards RBE3. This data indicates that RBE3 is distinguishable from RBE-I although immunologically related.<sup>110</sup> Furthermore, amino-terminal sequencing experiments showed that the RBE3 amino-terminal amino-acid sequence is neither related to RBE4a nor to RBE-I.<sup>109, 110</sup> The comprehensive sequences of both RBE3 and RBE1 are however significantly identical due to the high conservation of the central domain albeit with low similarity in the N- and C-terminal regions since RBE3 carries an additional 70-residue sequence at the amino-terminus and lacks about 50 residues at the carboxy-terminus when compared to RBE1.<sup>109</sup> The four conserved regions characteristic of the  $\alpha$ -amylase family of enzymes located in the catalytic domain are also conserved in RBE3 and RBE4, indicating that those enzymes are members of this family.<sup>62, 109-111</sup>

RBE4 is another isoform expressed as an 841-residue precursor protein that is further processed into a 788-amino acid peptide by deletion of an N-terminal fragment.<sup>110, 111</sup> These two forms, named RBE4a and RBE4b, have molecular weights of 93 and 83 KDa, respectively (**Table 1.1**). Even when the mature

RBE4b shows up, the peptide with the transit sequence, RBE4a, does not completely vanish. The two forms of RBE4 also present distinct bands on SDS-page gels and different reactivity towards anti-maize BE-I antibody. In terms of amino acid sequence similarity, mature RBE4 is significantly more identical to RBE3 and to maize BEIIa and BEIIb (80%), and to maize BEIIa (89%) than to RBE1 (47%). When the RBE4b sequence is compared to RBE1, an additional 90 amino acid sequence is found at the amino-terminus while the carboxy-terminus lacks 59 residues. Although RBE4 is highly similar to RBE1 and RBE3, it can be distinguished from these isoforms by the pattern of gene expression in tissues and during seed development and in enzymatic functions. RBE4 is expressed in the leaves as well as developing seeds of rice, while RBE3 is exclusively expressed in immature seeds and expression of RBE1 mainly occurs in the developing seeds but was also detected in the leaves in a significantly smaller amount, although considerable.

Another major feature that differentiates the various rice isozymes is the length of the chains transferred. BE1 formed a variety of short and intermediate chains of  $DP \leq 40$  attacking inner and outer chains of branched glucans with a higher amount of chains of  $DP = 6$  and 10-12.<sup>65, 111</sup> BE2a transferred outer chains of  $DP$  6-15 and was poorly active towards inner chains, while BE2b transferred mainly chains of  $DP$  6 and  $DP$  7 and was completely inactive towards inner chains.<sup>65</sup> The RBE3 chain transfer pattern varied from 10 min and 22 hours incubation between chains of  $DP$  6 and 7, and 3 and 4 respectively, while RBE4 transferred exclusively chains of  $DP$  6.<sup>111</sup>

Based on structural features and reactivity of RBE isoforms, it was established that all of them belong to the alpha-amylase family of enzymes; which supports the entrenched fact that branching enzyme handles two enzymatic functions, cleavage of alpha-1,4-glycosidic linkages and transfer of the newly cut oligosaccharide to another part of the sugar molecule.<sup>109</sup> These two steps must be accomplished simultaneously since branching enzyme was reported to have no starch-hydrolyzing activity.<sup>109, 130</sup> Although specific roles of all BE isoforms are still not completely elucidated, but the function of RBE3 and RBE4 was suggested to be involved in synthesis of different forms of amylopectin in the leaves versus the developing seeds.<sup>111</sup> On the other hand, the functions of RBE1, RBE2a and RBE2b were well studied by Nakamura et al. in 2010.<sup>65</sup> The synthesis of amylopectin was divided into six steps (**Figure 1.9**): (1) after the synthesis of a cluster of amylopectin, starch synthase IIIa (SSIIIa) elongates one chain, (2) RBE1 transfers the inner chains and/or long outer chains (up to DP ≤ 12) to the longest chain, (3) SSIIIa elongates the newly transferred chains and when the length reaches 12, (4) RBE2b interferes, with the specific activity it possesses, to form second branches having a uniform length of DP6 or DP7, (5) SSI and SSIIa elongate these branches now to reach a similar length as the initial cluster, (6) then steps (1) through (5) are repeated to form a new cluster.<sup>65</sup>

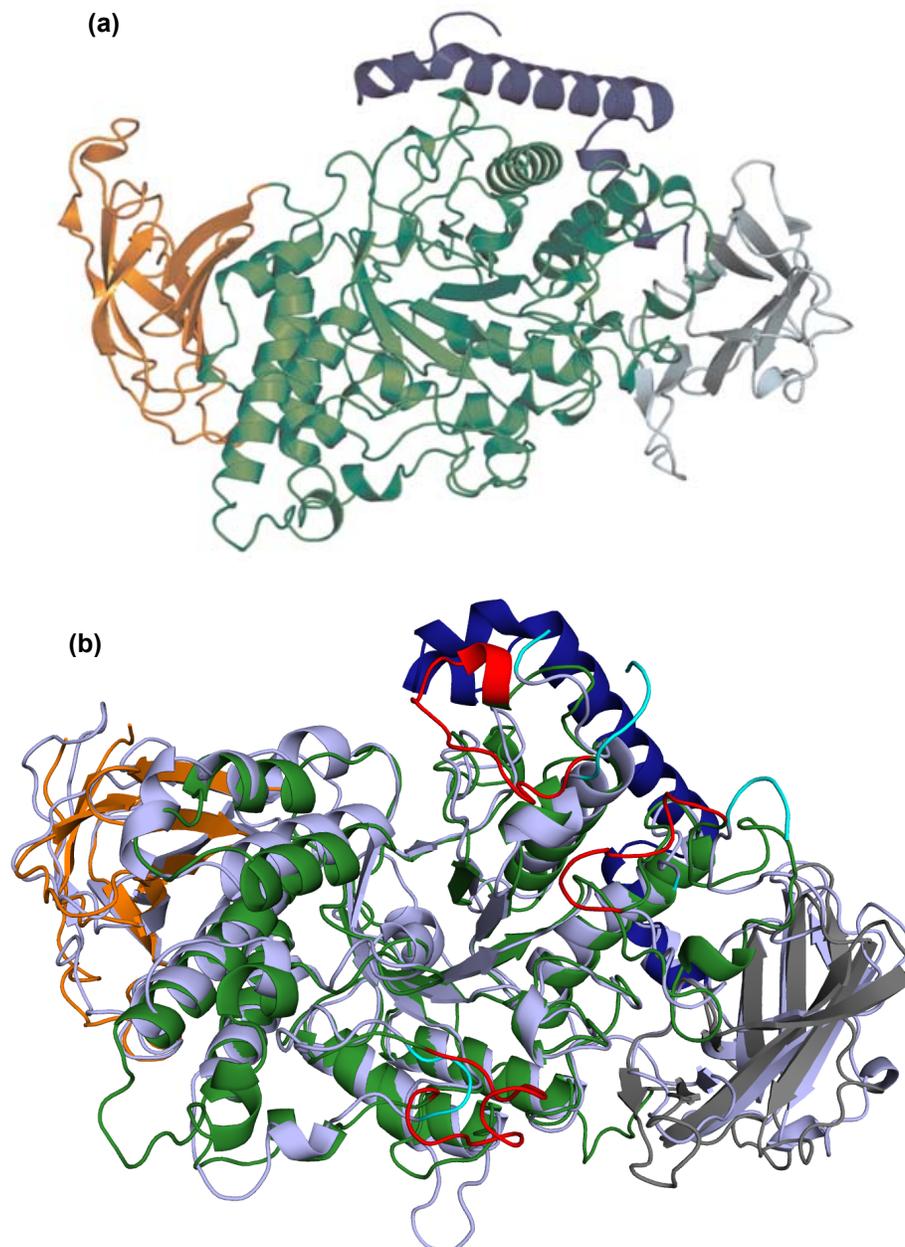


**Figure 1.9: Schematic diagram of the amylopectin synthesis in rice.** The role of BE and SS isoforms is represented in six stages: (1) elongation of one chain of the amylopectin cluster by SSIIIa, (2) transfer of the inner chains and/or long outer chains (up to DP ≤ 12) to the longest chain by RBE1, (3) elongation of the newly transferred chains by SSIIIa, (4) when the chain length reaches 12, RBE2b forms second branches with a uniform length of DP6 or DP7, (5) SSI and SSIIa elongate these branches now to reach a similar length as the initial cluster, (6) then steps (1) through (5) are repeated to form a new cluster.

### 1.2.4.3 Rice branching enzyme 1

SBE cleaves  $\alpha$ -1,4-glycosidic bonds and transfers newly formed chains to assemble  $\alpha$ -1,6 linkages. The branching process is intricate and results in starch that is diverse in architecture among organisms. Understanding the structure of branching enzyme is key to elucidate its function in detail. The crystal structure of rice branching enzyme 1, which will be referred to as RBEI (in roman letters, different than rice BE-I gene) hereafter, was determined in 2011 (**Figure 1.10**), nine years after the first structure of any branching enzyme (EcBE).<sup>66, 129</sup> The structure shows two distinct N-terminal regions, the first is located in a position as to be looking down onto the rest of the molecule (**Figure 1.10**) different from that of *Mycobacterium tuberculosis* BE<sup>131</sup> (MycoBE), a GBE. This domain, containing the first 58 amino acids, is formed of three  $\alpha$  helices, followed by the second N-terminal domain (residues 59-160), which constitutes the carbohydrate binding module 48 (CBM48), a polypeptide sequence shared among all BEs. The CBM48 domain is linked to the central catalytic domain (residues 182-581) that is formed of the  $(\beta/\alpha)_8$  barrel typical for the  $\alpha$ -amylase family of enzymes. Eight parallel inner  $\beta$ -strands are surrounded by eight outer  $\alpha$ -helices forming a groove shape, stabilized by the two N-terminal and the C-terminal domains on three sides. The C-terminus forms a  $\beta$ -sandwich structure consisting of loop linked anti-parallel  $\beta$ -strands. The structure of RBEI resembles that of EcBE in several features, and the two superimpose well especially in the catalytic domain with a few exceptions. One loop in RBEI, residues 537 – 550, is quite longer than its EcBE counterpart, residues 577 – 582 (towards the bottom of **Figure 1.10.b**).

Two other loops, residues 353 – 356, and 301 – 310, extend towards the top of the active site in RBEI and in most other GH13 enzymes, but in EcBE, they are disordered and appear to be pointing away from to active site although not shorter (residues 413 – 427 and 361 – 372). The figure also shows the relative position of the far N-terminal domain in RBEI (dark blue helices); whose structure is yet to be determined in EcBE (only far-N-terminal truncated structure is available so far). In addition, CBM48 domains of both structures overlap quite well, although some differences in their secondary structures. Finally, the C-terminal domain is very well superimposable between the two structures.

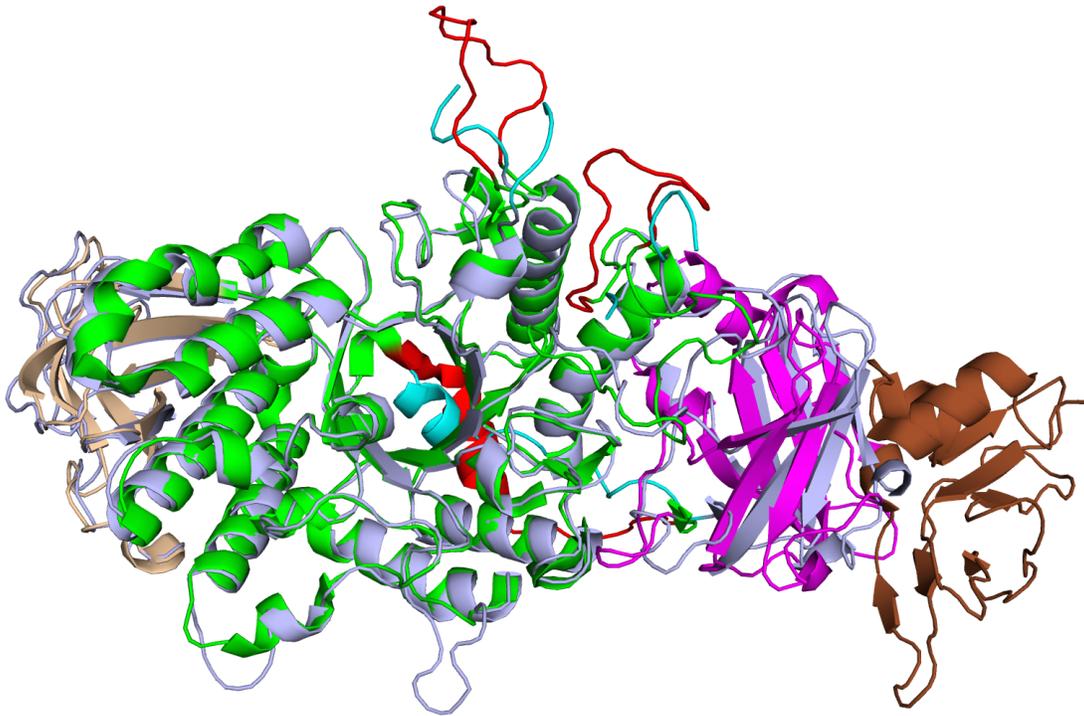


**Figure 1.10: RBEI, crystal structure and overlay with EcBE.** **a.** Crystal structure of RBEI. <sup>129</sup> The first N-terminal domain is displayed in dark blue, the second in grey, the catalytic domain in green and the C-terminus in orange. **b.** Overlay of RBEI and EcBE. The two proteins are mostly similar in the catalytic domain with a few loops differences highlighted in red for RBEI and cyan for EcBE, which cartoon is in light blue. Rest of colors is as above. Structures are those of our present work.

### 1.2.5 *Mycobacterium tuberculosis* branching enzyme (*MycoBE*)

The crystal structure of the full-length *MycoBE* was determined in 2010.<sup>131</sup> The protein is formed of four domains, just like *RBEI*, the N-terminal domain formed of a far domain (residues 1 – 105, colored brown), called N1, and a closer N2 domain (residues 106 – 226, magenta), the central domain (residues 227 – 631, bright green) and the C-terminal domain (residues 631 – 731, wheat).<sup>131</sup> This is a GBE protein, whose domains superimpose well with those of *EcBE* (**Figure 1.11**). The minimal exceptions to this are loops highlighted in the figure in red (*MycoBE*) and cyan (*EcBE*). These loops correspond in *MycoBE* from top to bottom to residues 422 – 435, 366 – 380, and 230 – 252. The first 2 loops are those disordered in *EcBE*; the only few residues of these loops that have a clear electron density superimpose well with *MycoBE*; they span residues 414 – 430 and 360 – 374. The last loop is the one that is most misaligned between the two structures; it is formed of residues 221 – 238 in *EcBE*. It is especially obvious that these structures share an entirely open active site, deficient of any long loop encumbering the polymer site of entry. Even the loops that are not ordered in *EcBE* (cyan) show tendency to emanate similarly to those in *MycoBE* (red). This is explained by the high similarity in substrate. *RBEI*, although not overwhelmed with loops that narrow the active site, has two loops discussed above that do not overlap with *EcBE* (red loops in **Figure 1.10.b**). It is reasonable to relate this to the difference in the ligands; glycogen has a very dense structure while amylopectin of starch though very branched is less dense than glycogen. What is most intriguing is the location of the far N-terminal domain, quite different than in

RBEI. We expect EcBE to have this domain in a similar location, but we can only conclude once the full-length structure is determined.



**Figure 1.11: Overlay of EcBE with MycoBE.** <sup>131</sup> MycoBE is colored according to the domains: the N-terminal domain formed of a far N1 domain (residues 1 – 105, brown), and a closer N2 domain (residues 106 – 226, magenta), the central domain (residues 227 – 631, bright green) and the C-terminal domain (residues 631 – 731, wheat). The cartoon of EcBE is colored in light blue.

In our work, we have obtained the structure of RBE1 bound to a twelve-glucose-unit oligomer; which revealed new binding sites. Based on this structure, we propose a hypothesis on the mode of binding between RBE1 and its substrate. All the results and the proposal are explained in the following chapters.

### 1.2.6 Activity assay

The activity of the branching enzyme was assayed by three different methods.

(1) *Branch-linkage (BL) assay*: it is based on the process of branching amylose, hydrolyzing the  $\alpha$ -1,6 linkages formed by isoamylase, then reading the absorbance of the product to determine the concentration of the reducing terminals formed per min; which is equivalent to the concentration of the branch linkages formed per min. <sup>64, 96, 108</sup>

(2) *Phosphorylase-stimulation assay*: the basis of the assay is the stimulation of synthesis of  $\alpha$ -D-glucan from [<sup>14</sup>C] labeled  $\alpha$ -D-Glucose-1-phosphate (Glc-1-P) by branching enzyme. Aliquots are taken at different times and incorporation of radioactive label into the glucan was checked. The activity is proportional to the concentration of labeled Glc-1-P incorporated into the glucan per min. <sup>64, 95, 96</sup>

(3) *Decrease of absorption of amylose-iodine complex*: amylose is dissolved in sodium hydroxide and branching enzyme added. Aliquots of the mixture were taken at timed intervals and iodine reagent added immediately. The iodine binds in the available space to the linear helix of amylose and therefore prevents further branching enzyme activity. The decrease in absorbance at 660 nm indicates less iodine binding to amylose due to more branches in the polysaccharide. <sup>59, 64, 95, 96</sup>

### **1.2.7 Chain length specificity assay**

Another essential feature that is unique to each branching enzyme from various organisms is the specificity in determining the lengths of chains transferred; making therefore starches and glycogen from different sources so diverse. Analyzing the chain transfer pattern was part of almost every study on branching enzyme and was reported by many research groups.<sup>60, 69, 71, 72, 74, 108</sup>

The idea of this assay is to branch amylose by incubating with BE over a time course, aliquoting out at specific times, then after halting the reaction, to reduce the  $\alpha$ -1,6-linkages by the action of isoamylase. The length of the hydrolyzed branches is then analyzed by HPAEC and gel filtration.

## REFERENCES

## REFERENCES

1. Buleon, A., Colonna, P., Planchot, V., and Ball, S. (1998) Starch granules: structure and biosynthesis, *International Journal of Biological Macromolecules* 23, 85-112.
2. Richards, N. (October 2010) Feed Cooked Grains, *FeedXL: Nutrition Makes a Difference Newsletter #18*, <http://feedxl.com/newsletters/18-feed-cooked-grains.html>.
3. Iglesias, A. A., and Preiss, J. (1992) Bacterial Glycogen and Plant Starch Biosynthesis, *Biochemical Education* 20, 196-203.
4. Imberty, A., Chanzy, H., and Perez, S. (1988) The Double-helical Nature of the Crystalline Part of A-Starch, *J. Mol. Biol.* 201, 365-378.
5. Imberty, A., and Perez, S. (1988) A Revisit to the Three-Dimensional Structure of B-Type Starch, *Biopolymers* 27, 1205-1221.
6. Immel, S. (1995) The Hydrophobic Topography of Amylose, *Clemens-Schöpf-Institut: Chapter 11*, <http://csi.chemie.tu-darmstadt.de/ak/immel/publications/phdthesis/chapter11.pdf>, 249-262.
7. Schulz, W., Sklenar, H., Hinrichs, W., and Saenger, W. (1993) The Structure of the Left-Handed Antiparallel Amylose Double Helix: Theoretical Studies, *Biopolymers* 33, 363-375.
8. Manners, D. J. (1989) Recent Developments in Our Understanding of Amylopectin Structure, *Carbohydrate Polymers* 11, 87-112.
9. Halsall, T. G., Hirst, E. L., and Jones, J. K. N. (1947) The Structure of Glycogen. Ratio of Non-terminal to Terminal Glucose Residues., *J. Chem. Soc.*, 1399-1400.
10. Haworth, W. N., Hirst, E. L., and Isherwood, E. (1937) Polysaccharides. Part XXIII. Determination of the Chain Length of Glycogen, *J. Chem. Soc.*, 577-581.
11. Manners, D. J. (1991) Recent Developments in Our Understanding of Glycogen Structure, *Carbohydrate Polymers* 16, 37-82.
12. Wolfrom, M. L., Lassetre, E. N., and O'Neill, A. N. (1951) Degradation of Glycogen to Isomaltose, *JACS* 73, 595-599.
13. Larnier, J., Illingworth, B., Cori, G. T., Cori, C. F., 199, and 641-51. (1952) Structure of glycogens and amylopectins - II. Analysis by stepwise enzymatic degradation. , *Journal of Biological Chemistry* 199, 641-651.

14. Meyer, K. H., and Fuld, M. (1941) Recherches sur L'amidon XII. L'arrangement des restes de glucose dans le glycogene., *Helv. Chim. Acta* 24, 375-378.
15. Hamley, I. W. (2010) Liquid crystal phase formation by biopolymers, *Soft Matter* 6, 1863.
16. Cornuéjols, D., and Pérez, S. (2010) Starch: a structural mystery, *Science In School*, <http://www.scienceinschool.org>.
17. Lehninger, A., Nelson, D., and Cox, M. (1993) Carbohydrates, *Principles of Biochemistry: Chapter 11*. <http://www.bioinfo.org.cn/book/biochemistry/chapt11/sim2.htm>.
18. Ahern, K., and Rajagopal, I. (2012) Biochemistry Free & Easy, *Oregon State University, Storage Carbohydrate Metabolism, Chapter 18*, 519. <http://oregonstate.edu/instruct/bb350/textmaterials/ch18.html>
19. Dickinson, D. B., and Preiss, J. (1969) Presence of ADP-Glucose Pyrophosphorylase in Shrunken-2 and Brittle-2 Mutants of Maize Endosperm, *Plant Physiology* 44, 1058-1062.
20. Preiss, J. (1984) Bacterial Glycogen Synthesis and Its Regulation, *Annu Rev Microbiol* 38, 419-458.
21. Preiss, J., Ball, K., Smith-White, B., Iglesias, A. A., Kakefuda, G., and Li, L. (1991) Starch Biosynthesis and its Regulation, *Biochem Soc Trans* 19, 539-547.
22. Tsai, C. Y., and Nelson, O. E. (1966) Starch Deficient Maize Mutant Lacking Adenosine Diphosphate Glucose Pyrophosphorylase Activity, *Science* 151, 341-343.
23. Ballicora, M. A., Iglesias, A. A., and Preiss, J. (2003) ADP-glucose pyrophosphorylase- a regulatory enzyme for plant starch synthesis, *Microbiol Mol Biol Rev* 67, 213-225.
24. Ballicora, M. A., Iglesias, A. A., and Preiss, J. (2004) ADP-glucose pyrophosphorylase- a regulatory enzyme for plant starch synthesis, *Photosynthesis Research* 79, 1-24.
25. Ball, K., and Preiss, J. (1994) Allosteric sites of the large subunit of the spinach leaf ADP-glucose pyrophosphorylase. , *Journal of Biological Chemistry* 269, 24706-24711.

26. Ballicora, M. A., Frueauf, J. B., Fu, Y., Schurmann, P., and Preiss, J. (2000) Activation of the Potato Tuber ADP-glucose Pyrophosphorylase by Thioredoxin, *Journal of Biological Chemistry* 275, 1315-1320.
27. Ballicora, M. A., Laughlin, M. J., Fu, Y., Okita, T. W., Barry, C. F., and Preiss, J. (1995) Adenosine 5'-Diphosphate-Glucose Pyrophosphorylase from Potato Tuber : Significance of the N Terminus of the Small Subunit for Catalytic Properties and Heat Stability, *Plant Physiology* 109, 245-251.
28. Bejar, C. M., Ballicora, M. A., Gomez-Casati, D. F., Iglesias, A. A., and Preiss, J. (2004) The ADP-glucose pyrophosphorylase from *Escherichia coli* comprises two tightly bound distinct domains, *FEBS letters* 573, 99-104.
29. Bejar, C. M., Ballicora, M. A., Iglesias, A. A., and Preiss, J. (2006) ADPglucose pyrophosphorylase's N-terminus: structural role in allosteric regulation, *Biochemical and biophysical research communications* 343, 216-221.
30. Frueauf, J. B., Ballicora, M. A., and Preiss, J. (2002) Alteration of inhibitor selectivity by site-directed mutagenesis of Arg294 in the ADP-glucose pyrophosphorylase from *Anabaena* PCC 7120, *Archives of biochemistry and biophysics* 400, 208-214.
31. Frueauf, J. B., Ballicora, M. A., and Preiss, J. (2003) ADP-glucose pyrophosphorylase from potato tuber: site-directed mutagenesis of homologous aspartic acid residues in the small and large subunits, *Plant Journal* 33, 503-511.
32. Greene, T. W., Woodbury, R. L., and Okita, T. W. (1996) Aspartic Acid 413 is Important for the Normal Allosteric Functioning of ADP-Glucose Pyrophosphorylase, *Plant Physiology* 112, 1315-1320.
33. Morell, M. K., Bloom, M., and Knowles, V. L. (1987) Subunit structure of spinach leaf ADP-glucose pyrophorylase. , *Plant Physiology* 85, 182-187.
34. Plaxton, W. C., and Preiss, J. (1987) Purification and properties of nonproteolytic degraded ADP-Glucose pyrophorylase from maize endosperm. , *Plant Physiology* 83, 105-112.
35. Preiss, J., and Greenberg, E. (1981) Biosynthesis of bacterial glycogen - activator specificity of the adenosine-diphosphate glucose pyrophosphorylases from the genus *Rhodospirillum*. , *Journal of bacteriology* 147, 711-719.

36. Tuncel, A., Kavakli, I. H., and Keskin, O. (2008) Insights into subunit interactions in the heterotetrameric structure of potato ADP-glucose pyrophosphorylase, *Biophysical Journal* 95, 3628-3639.
37. Jin, X., Ballicora, M. A., Preiss, J., and Geiger, J. H. (2005) Crystal structure of potato tuber ADP-glucose pyrophosphorylase, *EMBO J.* 24, 694-704.
38. Ainsworth, C., Clark, J., and Balsdon, J. (1993) Expression, organisation and structure of the genes encoding the waxy protein (granule-bound starch synthase) in wheat, *Plant Molecular Biology* 22, 67-82.
39. Boyer, C. D., and Preiss, J. (1981) Evidence for Independent Genetic Control of the Multiple Forms of Maize Endosperm Branching Enzymes and Starch Synthases, *Plant Physiology* 67, 1141-1145.
40. Cuesta-Seijo, J. A., Nielsen, M. M., Marri, L., Tanaka, H., Beeren, S. R., and Palcic, M. M. (2013) Structure of starch synthase I from barley: insight into regulatory mechanisms of starch synthase activity, *Acta Crystallographica D* 69, 1013-1011-1025.
41. Edwards, A., Fulton, D. C., Hylton, C. M., Jobling, S. A., Gidley, M., Roßner, U., Martin, C., and Smith, A. M. (1999) A combined reduction in activity of starch synthases II and III of potato has novel effects on the starch of tubers, *The Plant Journal* 17, 251-261.
42. Farkas, I., Hardy, T. A., Depaoliroach, A. A., and Roach, P. J. (1990) Isolation of the eGSY1 gene glycogen-synthase and evidence for the existence of a 2nd gene. , *Journal of Biological Chemistry* 265, 20879-20886.
43. Gao, M., Wanat, J., Stinard, P. S., James, M. G., and Myers, A. M. (1998) Characterization of dull1, a Maize Gene Coding for a Novel Starch Synthase, *The Plant Cell* 10, 399-412.
44. Holmes, E., Boyer, C. D., and Preiss, J. (1982) Immunological Characterization of *Escherichia coli* B Glycogen Synthase and Branching Enzyme and Comparison with Enzymes from Other Bacteria, *Journal of bacteriology* 151, 1444-1453.
45. Mason-Gamer, R. J., Weil, C. F., and Kellogg, E. A. (1998) Granule-Bound Starch Synthase: Structure, Function, and Phylogenetic Utility, *Mol Biol Evol* 15, 1658-1673.
46. Nakai, H., Baumann, M. J., Petersen, B. O., Westphal, Y., Hachem, M. A., Dilokpimol, A., Duus, J. O., Schols, H. A., and Svensson, B. (2010) *Aspergillus nidulans* alpha-galactosidase of glycoside hydrolase family 36

- catalyses the formation of alpha-galacto-oligosaccharides by transglycosylation, *The FEBS journal* 277, 3538-3551.
47. Nakamura, Y., Aihara, S., Crofts, N., Sawada, T., and Fujita, N. (2014) In vitro studies of enzymatic properties of starch synthases and interactions between starch synthase I and Starch Branching Enzymes from rice, *Plant science : an international journal of experimental plant biology* 224, 1-8.
  48. Pollock, C., and Preiss, J. (1980) The Citrate-Stimulated Starch Synthase of Starchy Maize Kernels- Purification and Properties, *Archives of biochemistry and biophysics* 204, 578-588.
  49. Seibold, G. M., Breiting, K. J., Kempkes, R., Both, L., Kramer, M., Dempf, S., and Eikmanns, B. J. (2011) The glgB-encoded Glycogen Branching Enzyme is essential for glycogen accumulation in *Corynebacterium glutamicum*, *Microbiology* 157, 3243-3251.
  50. Sheng, F., Yep, A., Feng, L., Preiss, J., and Geiger, J. H. (2009) Oligosaccharide binding in Escherichia coli glycogen synthase, *Biochemistry* 48, 10089-10097.
  51. Buschiazzo, A., Ugalde, J. E., Guerin, M. E., Shepard, W., Ugalde, R. A., and Alzari, P. M. (2004) Crystal structure of glycogen synthase: homologous enzymes catalyze glycogen synthesis and degradation., *EMBO J.* 23, 3196-3205.
  52. Momma, M., and Fujimoto, Z. (2014) Interdomain Disulfide Bridge in the Rice Granule Bound Starch Synthase I Catalytic Domain as Elucidated by X-Ray Structure Analysis, *Bioscience, Biotechnology and Biochemistry* 76, 1591-1595.
  53. Sheng, F., Jia, X., Yep, A., Preiss, J., and Geiger, J. H. (2009) The crystal structures of the open and catalytically competent closed conformation of Escherichia coli glycogen synthase, *The Journal of biological chemistry* 284, 17796-17807.
  54. Akasaka, T., Vu, N. T., Chaen, K., Nishi, A., Satoh, H., Ida, H., Omori, T., and Kimura, M. (2009) The Action of Rice Branching Enzyme I (BEI) on Starches, *Bioscience, Biotechnology, and Biochemistry* 73, 2516-2518.
  55. Baba, T., Kimura, K., Mizuno, K., Etoh, H., Ishida, Y., Shida, O., and Arai, Y. (1991) Sequence conservation of the catalytic regions of amylolytic enzymes in maize branching enzyme-I *Biochem. Biophys. Res. Commun.* 181, 87-94.
  56. Funane, K., Libessart, N., Stewart, D., Michishita, T., and Preiss, J. (1998) Analysis of Essential Histidine Residues of Maize Branching Enzymes by

- Chemical Modification and Site-Directed Mutagenesis, *J Protein Chem* 17, 579-590.
57. Gao, M., Fisher, D. K., Kim, K. N., Shannon, J. C., and Guiltinan, M. J. (1996) Evolutionary conservation and expression patterns of maize Starch Branching Enzyme I and IIb genes suggests isoform specialization, *Plant Molecular Biology* 30, 1223-1232.
  58. Gao, M., Fisher, D. K., Kim, K. N., Shannon, J. C., and Guiltinan, M. J. (1997) Independent Genetic Control of Maize Starch-Branching Enzymes IIa and IIb: Isolation and characterization of SBEIIa cDNA, *Plant Physiology* 114, 69-78.
  59. Guan, H. P., Baba, T., and Preiss, J. (1994) Expression of Branching Enzyme I of Maize Endosperm in *Escherichia coli*, *Plant Physiology* 104, 1449-1453.
  60. Guan, H. P., Kuriki, T., Sivak, M., and Preiss, J. (1995) Maize Branching Enzyme Catalyzes synthesis of Glycogen-like Polysaccharide in GlgB-Deficient *Escherichia coli*, *Proceedings of the National Academy of Sciences in the United States of America* 92, 964-967.
  61. Hong, S., Mikkelsen, R., and Preiss, J. (2001) Analysis of the amino terminus of maize branching enzyme II by polymerase chain reaction random mutagenesis, *Archives of biochemistry and biophysics* 386, 62-68.
  62. Kawasaki, T., Mizumo, K., Baba, T., and Shimada, H. (1993) Molecular analysis of the gene encoding a rice starch branching enzyme, *Molecular and General Genetics* 237, 10-16.
  63. Kuriki, T., Guan, H., Sivak, M., and Preiss, J. (1996) Analysis of the Active Center of Branching Enzyme II from Maize Endosperm, *J Protein Chem* 15, 305-313.
  64. Kuriki, T., Stewart, D. C., and Preiss, J. (1997) Construction of Chimeric Enzymes out of Maize Endosperm Branching Enzymes I and II: Activity and Properties, *Journal of Biological Chemistry* 272, 28999-29004.
  65. Nakamura, Y., Utsumi, Y., Sawada, T., Aihara, S., Utsumi, C., Yoshida, M., and Kitamura, S. (2010) Characterization of the reactions of Starch Branching Enzymes from rice endosperm, *Plant and Cell Physiology* 51, 776-794.
  66. Vu, N. T., Shimada, H., Kakuta, Y., Nakashima, T., Ida, H., Omori, T., Nishi, A., Satoh, H., and Kimura, M. (2008) Biochemical and Crystallographic Characterization of the Starch Branching Enzyme I (BEI) from *Oryz sativa* L, *Biosci. Biotechnol. Biochem.* 72, 2858-2866.

67. Binderup, K., Libessart, N., and Preiss, J. (2000) Slow-binding inhibition of Branching Enzyme by the pseudooligosaccharide BAY e4609, *Archives of biochemistry and biophysics* 374, 73-78.
68. Binderup, K., Mikkelsen, R., and Preiss, J. (2000) Limited proteolysis of Branching Enzyme from *Escherichia coli*, *Archives of biochemistry and biophysics* 377, 366-371.
69. Binderup, K., Mikkelsen, R., and Preiss, J. (2002) Truncation of the amino terminus of Branching Enzyme changes its chain transfer pattern, *Archives of biochemistry and biophysics* 397, 279-285.
70. Binderup, K., and Preiss, J. (1998) Glutamate-459 Is Important for *Escherichia coli* Branching Enzyme Activity, *Biochemistry* 37, 9033-9037.
71. Devillers, C. H., Piper, M. E., Ballicora, M. A., and Preiss, J. (2003) Characterization of the branching patterns of Glycogen Branching Enzyme truncated on the N-terminus, *Archives of biochemistry and biophysics* 418, 34-38.
72. Guan, H., Li, P., Imparl-Radosevich, J., Preiss, J., and Keeling, P. (1997) Comparing the Properties of *Escherichia coli* Branching Enzyme and Maize Branching Enzyme, *Archives of biochemistry and biophysics* 342, 92-98.
73. Mikkelsen, R., Binderup, K., and Preiss, J. (2001) Tyrosine residue 300 is important for activity and stability of branching enzyme from *Escherichia coli*, *Archives of biochemistry and biophysics* 385, 372-377.
74. Palomo, M., Kralj, S., van der Maarel, M. J., and Dijkhuizen, L. (2009) The unique branching patterns of *Deinococcus* Glycogen Branching Enzymes are determined by their N-terminal domains, *Applied and environmental microbiology* 75, 1355-1362.
75. Jobling, S. (2004) Improving starch for food and industrial applications, *Current opinion in plant biology* 7, 210-218.
76. Shewmaker, C. K., and Stalker, D. M. (1992) Modifying Starch Biosynthesis with Transgenes in Potatoes, *Plant Physiology* 100, 1083-1086.
77. Regierer, B., Fernie, A. R., Springer, F., Perez-Melis, A., Leisse, A., Koehl, K., Willmitzer, L., Geigenberger, P., and Kossmann, J. (2002) Starch content and yield increase as a result of altering adenylate pools in transgenic plants, *Nature biotechnology* 20, 1256-1260.
78. Slattery, C. J., Kavakli, I. H., and Okita, T. W. (2000) Engineering starch for increased quantity and quality, *Trends in Plant Science* 5, 291-298.

79. Crabb, W. D., and Shetty, J. K. (1999) Commodity scale production of sugars from starches, *Curr Opin Microbiol* 2, 252-256.
80. Tester, R. F., and Debon, S. J. J. (2000) Annealing of starch — a review, *International Journal of Biological Macromolecules* 27, 1-12.
81. Tally, S. (2002) Making biofuel from corn stover, *BioCycle* 43, 62.
82. Aikawa, S., Izumi, Y., Matsuda, F., Hasunuma, T., Chang, J. S., and Kondo, A. (2012) Synergistic enhancement of glycogen production in *Arthrospira platensis* by optimization of light intensity and nitrate supply, *Bioresour Technol* 108, 211-215.
83. Singh, N. K., and Dhar, D. W. (2011) Microalgae as second generation biofuel. A review, *Agronomy for Sustainable Development* 31, 605-629.
84. Tang, W., and Tang, A. Y. (2014) Transgenic woody plants for biofuel, *Journal of Forestry Research* 25, 225-236.
85. Demain, A. L., Newcomb, M., and Wu, J. H. (2005) Cellulase, clostridia, and ethanol, *Microbiol Mol Biol Rev* 69, 124-154.
86. Ruegger, M., Meyer, K., Cusumano, J. C., and Chapple, C. (1999) Regulation of ferulate-5-hydroxylase expression in *Arabidopsis* in the context of sinapate ester biosynthesis, *Plant Physiology* 119, 101-110.
87. Sewalt, V., Ni, W., Blount, J. W., Jung, H. G., Masoud, S. A., Howles, P. A., Lamb, C., and Dixon, R. A. (1997) Reduced lignin content and altered lignin composition in transgenic tobacco down-regulated in expression of l-phenylalanine, *Plant Physiology* 115, 41-50.
88. Vignols, F., Rigau, J., Torres, M. A., Capellades, M., and Puigdomenech, P. (1995) The brown midrib3 (bm3) mutation in maize occurs in the gene encoding caffeic acid O-methyltransferase, *Plant Cell* 7, 407-416.
89. Chen, Y. T., and Burchell, A. (1995) Glycogen Storage Diseases, *The Metabolic and Molecular Basis of Inherited Diseases*. C.R. Scriver, A.L. Beaudet, W.S. Sly, and D. Valle, eds. (New York, McGraw-Hill, Inc., Health Professions Division) 1, 935-965.
90. DiMauro, S., and Tsujino, S. (1994) Non-Lysosomal Glycogenoses, *Myology* (Engel, A. G. and Franzini-Armstrong, C., eds) 2, 1554-1576.
91. Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities., *Biochemical Journal* 280, 309-316.

92. Feng, L. (2009) X-Ray Crystallographic Studies of Branching Enzyme/Polysaccharide Complex and RNA Polymerase III Transcription Factor TFIIIB Complex, *Dissertation*, p. 10.
93. van der Maarel, M. J. E. C., van der Veen, B., Uitdehaag, J. C. M., Leemhuis, H., and Dijkhuizen, L. (2002) Properties and applications of starch-converting enzymes of the  $\alpha$ -amylase family, *Journal of Biotechnology* 94, 137–155.
94. Abad, M. C., Binderup, K., Rios-Steiner, J., Arni, R. K., Preiss, J., and Geiger, J. H. (2002) The X-ray crystallographic structure of *Escherichia coli* Branching Enzyme, *Journal of Biological Chemistry* 277, 42164-42170.
95. Boyer, C. D., and Preiss, J. (1978) Multiple Forms of (1,4)- $\alpha$ -D-Glucan, (1,4)- $\alpha$ -D-Glucan-6-Glycosyl Transferase From Developing *Zea mays* L. Kernels, *Carbohydrate Research* 61, 321-334.
96. Guan, H. P., and Preiss, J. (1993) Differentiation of the Properties of the Branching Isozymes from Maize (*Zea mays*), *Plant Physiology* 102, 1269-1273.
97. Romeo, T., Kumar, A., and Preiss, J. (1988) Analysis of the *Escherichia coli* glycogen gene cluster suggests that catabolic enzymes are encoded among the biosynthetic genes, *Gene* 70, 363-376.
98. Nakajima, R., Imanaka, T., and Aiba, S. (1986) Comparison of amino acid sequences of eleven different  $\alpha$ -amylases, *Appl. Microbiol. Biotechnol.* 23, 355--360.
99. Jespersen, H. M., MacGregor, E. A., Henrissat, B., Sierks, M. R., and Svensson, B. (1993) Starch- and Glycogen-Debranching and Branching Enzymes: Prediction of Structural Features of the Catalytic ( $\alpha/\beta$ )8-Barrel Domain and Evolutionary Relationship to Other Amylolytic Enzymes, *Journal of Protein Chemistry* 12, 791-805.
100. Jespersen, H. M., MacGregor, E. A., Sierks, M. R., and Svensson, B. (1991) Comparison of the domain-level organization of starch hydrolases and related enzymes, *Biochemistry J.* 280, 51-55.
101. Svensson, B. (1994) Protein engineering in the  $\alpha$ -amylase family- catalytic mechanism, substrate specificity, and stability, *Plant Molecular Biology* 25, 141-157.
102. Matsumoto, A., Nakajima, T., and Matsuda, K. (1990) A Kinetic Study of the Interaction between Glycogen and *Neurospora crassa* Branching Enzyme, *Journal of Biochemistry* 107, 123-126.

103. Matsumoto, A., Nakajima, T., and Matsuda, K. (1990) Mode of Action of Glycogen Branching Enzyme from *Neurospora crassa*, *Journal of Biochemistry* 107, 118-122.
104. Takeda, Y., Guan, H. P., and Preiss, J. (1993) Branching of amylose by the branching isoenzymes of maize endosperm *Carbohydrate Research* 240, 253-263.
105. Baecker, P. A., Greenberg, E., and Preiss, J. (1986) Biosynthesis of Bacterial Glycogen: Primary Structure of *Escherichia coli* 1,4-alpha-D-glucan:1,4-alpha-D-glucan 6-alpha-D-(1,4-alpha-D-glucano)-transferase as deduced from the nucleotide sequence of the Glg B gene, *The Journal of biological chemistry* 261, 8738-8743.
106. Okita, T. W., Rodriguez, R. L., and Preiss, J. (1981) Biosynthesis of Bacterial Glycogen: Cloning of the Glycogen Biosynthetic Enzyme Structural Genes of *Escherichia Coli*, *Journal of Biological Chemistry* 256, 6944-4952.
107. Abad, M. C., Binderup, K., Preiss, J., and Geiger, J. H. (2001) Crystallization and preliminary X-ray diffraction studies of *Escherichia coli* branching enzyme, *Acta Crystallographica, D* 58, 359-361.
108. Takeda, Y., Guan, H.-P., and Preiss, J. (1993) Branching of amylose by the Branching Isoenzymes of maize endosperm, *Carbohydr. Res.* 240, 253-263.
109. Mizuno, K., Kawasaki, T., Shimada, H., Satoh, H., Kobayashi, E., Okumura, S., Arai, Y., and Baba, T. (1993) Alteration of the Structural Properties of Starch Components by the Lack of an Isoform of Starch Branching Enzyme in Rice Seeds, *The Journal of biological chemistry* 268, 19084-19091.
110. Mizuno, K., Kimura, K., Arai, Y., Kawasaki, T., Shimada, H., and Baba, T. (1992) Starch Branching Enzymes from Immature Rice Seeds, *Journal of Biochemistry* 112, 643-651.
111. Mizuno, K., Kobayashi, E., Tachibana, M., Tsutomu, K., Fujimura, T., Funane, K., Kobayashi, M., and Baba, T. (2001) Characterization of an isoform of rice starch branching enzyme, RBE4, in developing seeds, *Plant Cell Physiol.* 42, 349-357.
112. Nakamura, Y., Takeichi, T., Kawaguchi, K., and Yamanouchi, H. (1992) Purification of two forms of starch Branching Enzyme (Q-enzyme) from developing rice endosperm, *Plant Physiology* 84, 329 –335.

113. Morell, M. K., Blennow, A., Kosar-Hashemi, B., and Samuel, M. (1997) Differential Expression and Properties of Starch Branching Enzyme Isoforms in Developing Wheat Endosperm, *Plant Physiol.* 113, 201-208.
114. Rahman, S., Li, Z., Abrahams, S., Abbott, D., Appels, R., and Morell, M. K. (1999) Characterisation of a gene encoding wheat endosperm starch branching enzyme-I, *Theor. Appl. Genet.* 98, 156-163.
115. Repellin, A., Båga, M., and Chibbar, R. N. (2001) Characterization of a cDNA encoding a type I starch branching enzyme produced in developing wheat (*Triticum aestivum* L.) kernels, *Journal of Plant Physiology* 158, 91-100.
116. Sadequr Rahman\*, A. R., Zhongyi Li, Yasuhiko Mukai, Maki Yamamoto, Behjat Kosar-Hashemi, Sharon Abrahams, and Matthew K. Morell. (2001) Comparison of Starch-Branching Enzyme Genes Reveals Evolutionary Relationships Among Isoforms. Characterization of a Gene for Starch-Branching Enzyme IIa from the Wheat D Genome Donor *Aegilops tauschii*1, *Plant Physiology* 125.
117. Sun, C., Sathish, P., Ahlandsberg, S., Deiber, A., and Jansson, C. (1997) Identification of four starch-branching enzymes in barley endosperm-partial purification of forms I, IIa and IIb, *New Phytol.* 137, 215-222.
118. ChuanXin, S., Sathish, P., Deiber, A., and Jansson, C. (1995) Multiple starch branching enzymes in barley endosperm, *Photosynthesis: from light to biosphere Volume V. Proceedings of the Xth International Photosynthesis Congress, Montpellier, France*, 317-320.
119. Sun, C., Sathish, P., Ahlandsberg, S., and Jansson, C. (1998) The Two Genes Encoding Starch-Branching Enzymes IIa and IIb Are Differentially Expressed in Barley, *Plant Physiol.* 118, 37-49.
120. Rydberg, U., Andersson, L., Andersson, R., Åman2, P., and Larsson, H. (2001) Comparison of starch branching enzyme I and II from potato, *Eur. J. Biochem.* 268, 6140-6145.
121. Visko-Nielsen, A., and Blennow, A. (1998) Isolation of starch branching enzyme I from potato using gamma-cyclodextrin affinity chromatography, *J. Chromatogr.* 800, 382-385.
122. Salehuzzaman, S. N. I. M., Jacobsen, E., and Visser, R. G. F. (1992) Cloning, partial sequencing and expression of a cDNA coding for branching enzyme in cassava, *Plant Molecular Biology* 20, 809-819.
123. Amith, A. M. (1988) Major differences in isoforms of starch-branching enzyme between developing embryos of round- and wrinkled-seeded peas (*Pisum sativum* L.), *Planta* 175, 270-279.

124. Tomlinson, K. L., Lloyd, J. R., and Smith, A. M. (1997) Importance of isoforms of starch-branching enzyme in determining the structure of starch in pea leaves, *The Plant Journal* 11, 31-43.
125. Zhu, Z. P., and Chen, S. Q. (1992, 20-24 July) Starch accumulation and branching enzyme gene expression during seed development of soybean and pea, *Proceedings of the Fourth International Workshop on Seeds: basic and applied aspects of seed biology, Angers, France 1*, 127-133.
126. Nakamura, Y. (2002) Towards a better understanding of the metabolic system for amylopectin biosynthesis in plants, *Plant Cell Physiol.* 43, 718-725.
127. Mu-Forster, C., Huang, R., Powers, J. R., Harriman, R. W., Knight, M., Singletary, G. W., Keeling, P., and Wasserman, B. P. (1996) Physical Association of Starch Biosynthetic Enzymes with Starch Granules of Maize Endosperm, *Plant Physiol.* 111, 821-829.
128. Ohdan, T., Francisco, P. B., Jr., Sawada, T., Hirose, T., Terao, T., Satoh, H., and Nakamura, Y. (2005) Expression profiling of genes involved in starch synthesis in sink and source organs of rice, *Journal of experimental botany* 56, 3229-3244.
129. Noguchi, J., Chaen, K., Vu, N. T., Akasaka, T., Shimada, H., Nakashima, T., Nishi, A., Satoh, H., Omori, T., Kakuta, Y., and Kimura, M. (2011) Crystal structure of the Branching Enzyme I (BEI) from *Oryza sativa* L with implications for catalysis and substrate binding, *Glycobiology* 21, 1108-1116.
130. Borovsky, D., Smith, E. E., and Whelan, W. J. (1975) Purification and Properties of Potato 1,4-alpha-D-Glucan: 1,4-alpha-D-Glucan 6-alpha-(1,4-alpha-Glucano)-Transferase, *Eur. J. Biochem.* 59, 615-625.
131. Pal, K., Kumar, S., Sharma, S., Garg, S. K., Alam, M. S., Xu, H. E., Agrawal, P., and Swaminathan, K. (2010) Crystal structure of full-length *Mycobacterium tuberculosis* H37Rv Glycogen Branching Enzyme: insights of N-terminal beta-sandwich in substrate specificity and enzymatic activity, *The Journal of biological chemistry* 285, 20897-20903.

## **CHAPTER 2: ACTIVITY AND CHAIN LENGTH SPECIFICITY ASSAYS IN *ESCHERICHIA COLI* BE**

### **2.1 Objective**

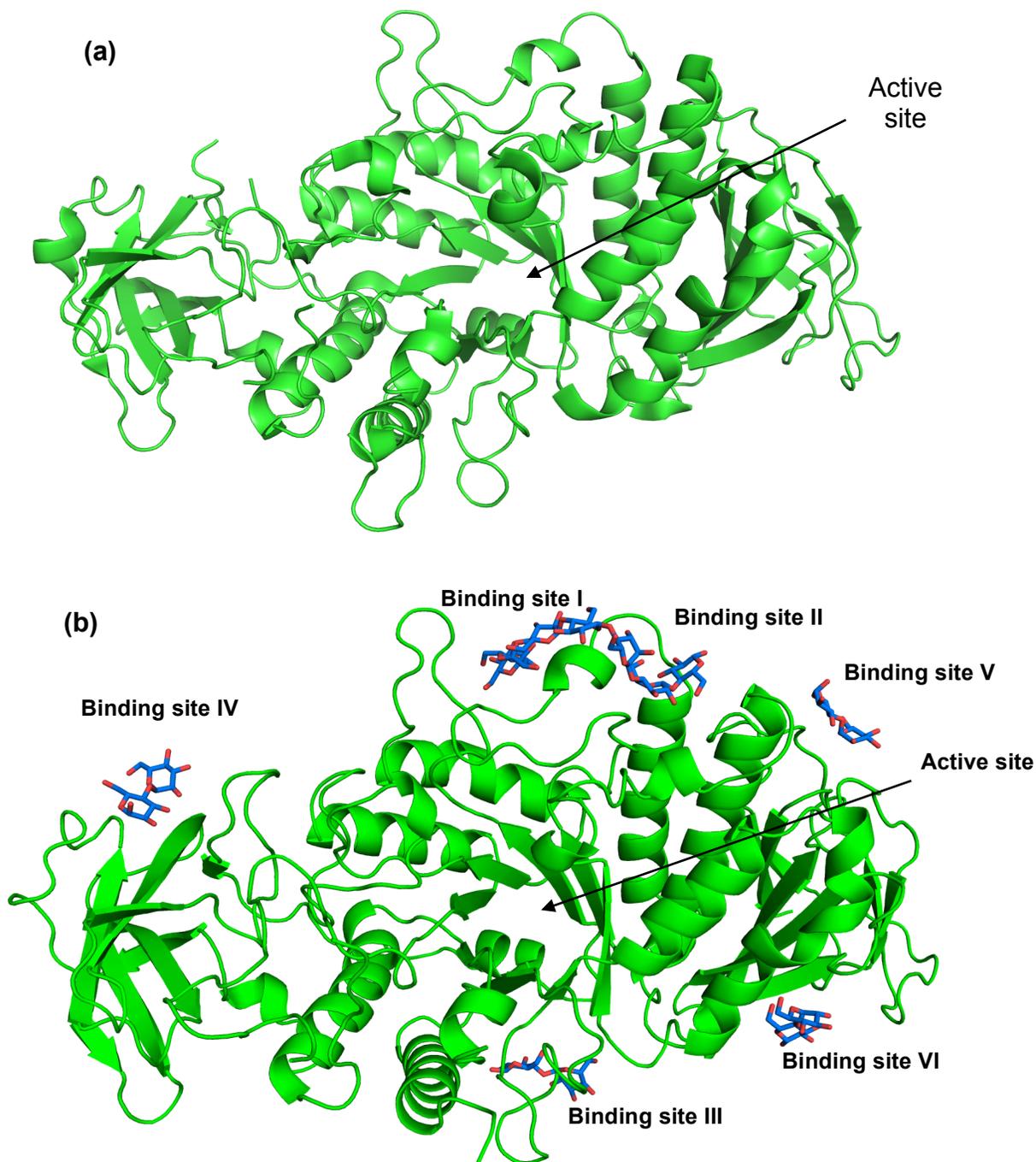
Branching enzyme (BE) belongs to the GH13 family of enzymes and shares several structural and functional features with the members of this family, including regions of conserved residues.<sup>1, 2</sup> Even though BE is related to this group of enzymes, it has a unique reaction, consisting of two steps: the cleavage of an  $\alpha$ -1,4 glycosidic bond followed by the formation of an  $\alpha$ -1,6 linkage.<sup>3, 4</sup> The residues directly assisting in catalysis have been identified.<sup>5-9</sup> After solving the structure of EcBE soaked in oligosaccharides, seven external binding sites were discovered (unpublished data, Feng, L.).<sup>10</sup> The question then became: how do these residues explicitly involved in binding the carbohydrates affect protein activity? In this chapter, we describe these seven binding sites and investigate the importance of each binding site by constructing mutants within or across sites and assaying their activity. We also report the results of transfer chain assays for some of these mutants.

## 2.2 The Crystal Structure of *E. coli* BE (EcBE)

EcBE crystal structure solved in 2002 by Abad and coworkers was the first structure of any branching enzyme reported.<sup>11</sup> This protein has three domains: an NH<sub>2</sub>- and COOH- terminal domain, and a central catalytic domain common to the  $\alpha$ -amylase family of enzymes.<sup>11</sup> Attempts at crystallizing the full-length protein were unsuccessful, which led researchers to gear towards a truncated stable version.<sup>12</sup> The structure is represented in **Figure 2.1.a**. The same crystal form of EcBE was soaked in several linear malto-oligosaccharides ranging in length from maltose to maltoheptaose and in  $\alpha$ ,  $\beta$  and  $\gamma$  cyclodextrins.<sup>10</sup> While EcBE soaking in shorter sugars resulted in weak oligosaccharide density, probably caused by their higher tendency to dissociate, binding of EcBE to maltohexaose (M6), maltoheptaose (M7) and  $\alpha$ ,  $\beta$ , and  $\gamma$  cyclodextrins, in contrast, resulted in interpretable electron density maps. These complexes were used to study binding sites in EcBE<sup>10</sup>.

### 2.3 *EcBE* Glucan Binding Sites

There are four molecules in the asymmetric unit of *EcBE* crystal structure,<sup>10</sup>  
<sup>11</sup>. In consequence, crystal packing prevents identical binding of oligosaccharides  
in all four molecules. The overall structure resulting from superposition of the  
four entities shows six linear oligosaccharide binding sites, numbered I through  
VI (**Figure 2.1**). In addition a unique binding site for cyclic sugars was identified  
in a-, b- and g-cyclodextrin-bound *EcBE*<sup>10</sup> This was numbered binding site VII.

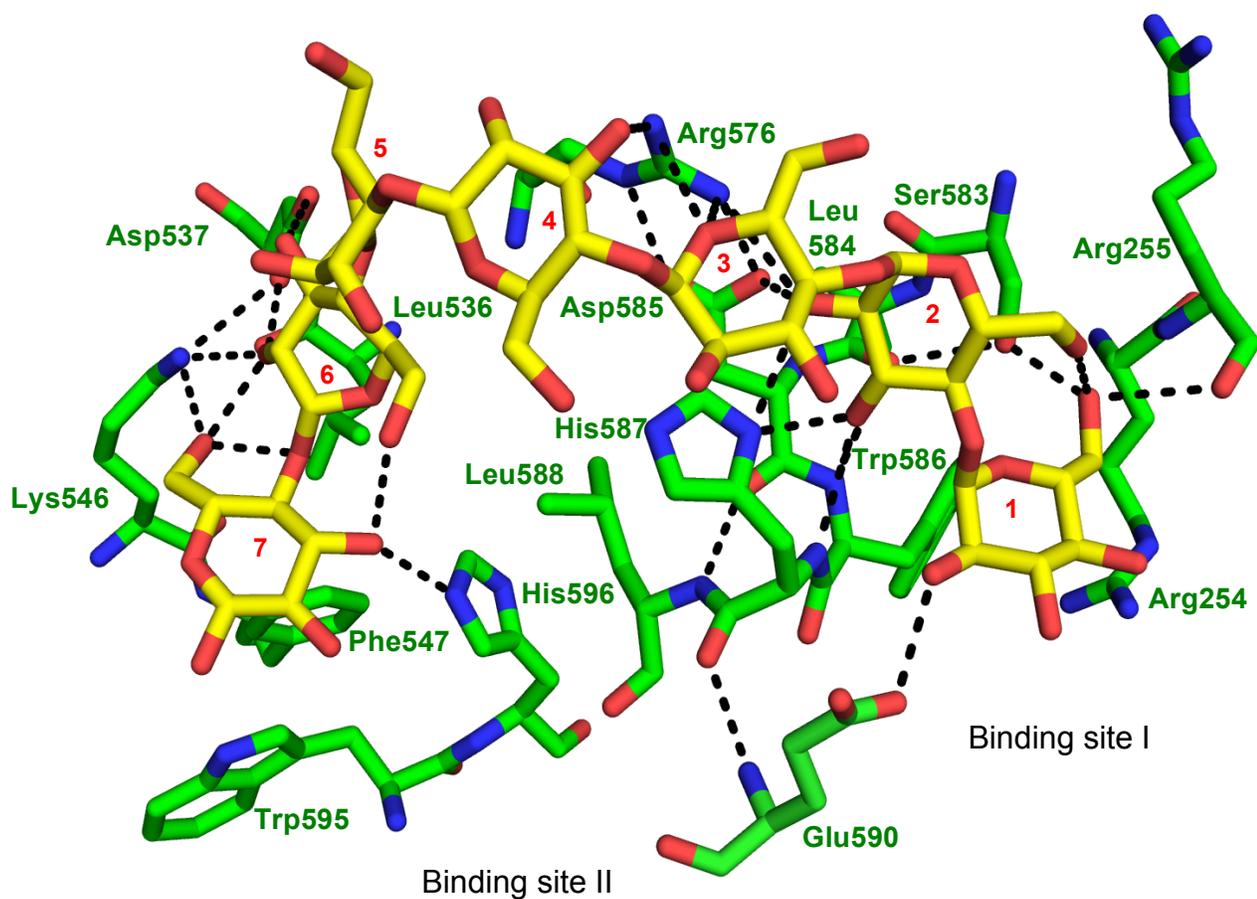


**Figure 2.1: EcBE apo and holo.** a. EcBE overall structure (PDB # 1M7X).<sup>11</sup> b. Overall EcBE structure (molecule B in green cartoon) in complex with heptaose (colored by atom, with C blue and O red).<sup>10</sup> Seven binding sites are seen. Note that there is no binding in the active site.

### 2.3.1 Binding sites I and II

A single M7 molecule spans binding sites I and II, which are closest to the active site (around 20 Å away). Two glucose (Glc) units interact directly with the protein in each site, sugars 1 and 2 on the non-reducing end bind in site I, and sugars 6 and 7 on the reducing end in site II. Several hydrogen bonds are seen in these sites in addition to a unique hydrophobic interaction. Glc-1 makes H-bonds with Arg254, Arg255, Ser583 and Glu590, and Glc-2 with each of Arg576, Asp585, and His587 (**Figure 2.2**). One end-on hydrophobic interaction is observed between Trp586 and Glc-1. Most of these residues are conserved among Enterobacteria, indicating the importance of this binding site.

In binding site II, the protein interacts with the last two sugars of the heptaose. Oxygens O2 and O3 of Glc-6 make hydrogen bonds with Lys 546 and Asp537 side chains, respectively; while oxygens O3 and O6 of Glc-7 hydrogen binds to His596 side chain and Lys546, respectively. These few interactions define binding site II.

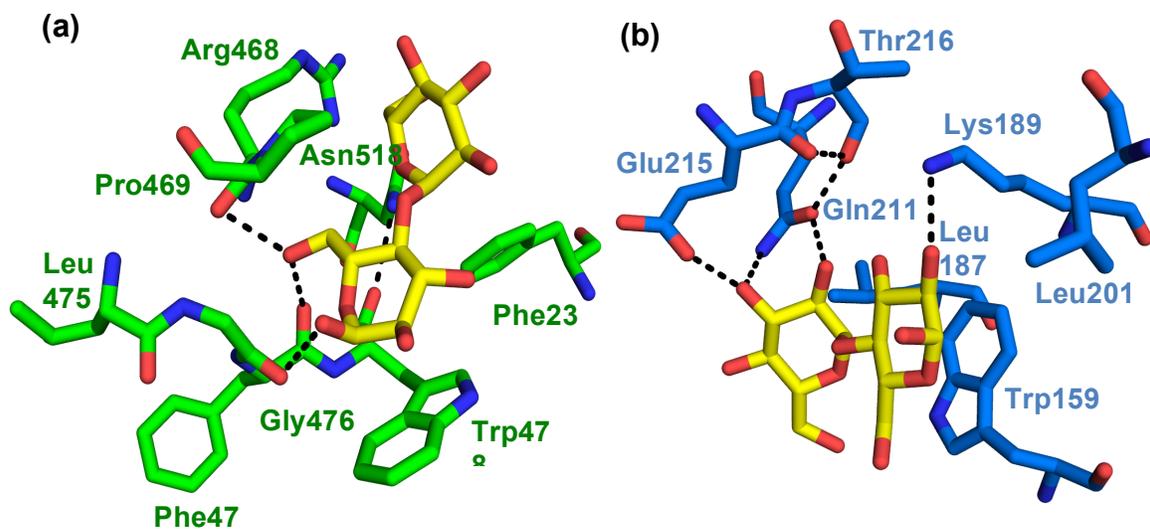


**Figure 2.2: Binding sites I and II of the EcBE / maltoheptaose complex.** The non-reducing end of M7 binds to binding site I, and the reducing end to II. Protein atoms are colored by atom type, with C, green, N, blue and O, red; and M7 atoms are colored in the same way except C is in yellow. Dotted black lines show hydrogen bonds. Glucose units of M7 are numbered in red.<sup>10</sup>

### 2.3.2 Binding sites III, IV, V and VI

Each of the four other binding sites identified in both complexes of EcBE with M6 and M7 show only two ordered-glucose units. Binding sites III and VI are located at the bottom of the groove-like active site of EcBE. This is the area where the reducing end of a GH13 enzyme substrate typically exits the protein <sup>13</sup>. In binding site III, only two glucose units are visible and they make a total of three hydrogen bonds with EcBE, all to the reducing end sugar. Two are made between the main chain carbonyls of Gly476 and Phe477, and O5 and O6 of the sugar, and one between Arg468 main chain nitrogen and O6 of the sugar (**Figure 2.3.a**). This site also shows several hydrophobic interactions such as between Pro469 and the maltose unit and typical sugar / aromatic stacking interactions between Phe239 and Trp478 and the sugar.

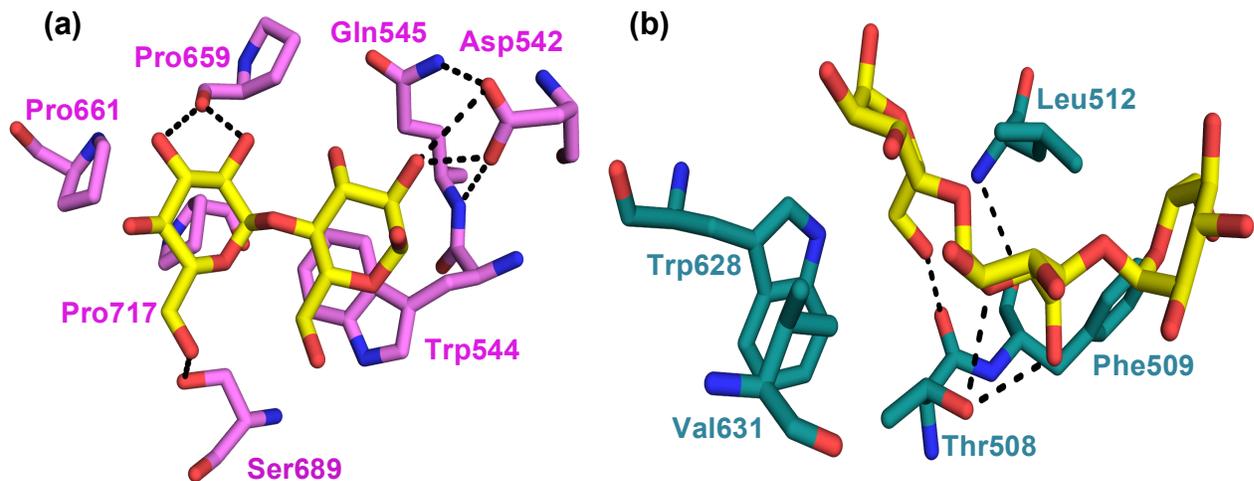
Binding site IV is located in the N-terminal domain CBM48, carbohydrate-binding-module 48 (explained further in chapters 3 and 4), facing the same side of the enzyme as binding sites I and II (**Figure 2.1**). In binding site IV, four hydrogen bonds are visible between the sugar and EcBE. To the non-reducing end sugar, the oxygen and nitrogen side chains of Gln211 and Glu215 respectively make two H-bonds to O3, and the oxygen side chain of Gln211 makes a second H-bond to O2. Another hydrogen bond is made between Lys189 and O2 of the reducing end sugar. Trp159 stacks against the reducing-end sugar making strong hydrophobic interactions; which are also established between the sugar and two other residues, Leu187 and Leu201 located at close proximity of the maltose (**Figure 2.3.b**).



**Figure 2.3: EcBE – M7 binding at sites III and IV.** **a.** Oligosaccharide binding in sites III and IV. Hydrophobic interactions and hydrogen bonds between maltoheptaose (M7) and EcBE at binding site III are shown. Coloring is as in figure 2.2. **b.** Interactions at binding site IV. Binding in this site is visible in molecule D of EcBE (C in blue, O in red and N in deep blue). Hydrophobic interactions are established through stacking of Trp159 against the sugar; and through two other residues, Leu187 and Leu201 located at close proximity of the sugar. Hydrogen bonds are shown in dotted black lines.

Binding sites V and VI bridge the catalytic and C-terminal domains. Binding site V is located on the same face of the enzyme as binding sites I, II and IV (**Figure 2.1**). Stacking hydrophobic interactions between Trp544 and the reducing end sugar is visible in binding site V. Several H-bonds are also visible in this site, two exist between the Asp542 side chain and O2 of the reducing end sugar, and three to the non-reducing end sugar, between Pro659 and O2 and O3 of the sugar, and the last one between the side chain of Ser689 and O6 of the sugar (**Figure 2.4.a**).

Binding site VI, located on the opposite face of the molecule when compared to I, II, IV and V, bridges the catalytic and COOH-terminal domains across from binding site V (**Figure 2.1**). In molecule A, three glucose units are visible. Hydrogen bonds are seen between the glucan and the main chain and side chain of Thr508. The glucan falls among three hydrophobic residues, Leu512, Trp628 and Phe509 making close interactions. It is notable that every binding site involves a hydrophobic stacking interaction between a Tryptophan residue and a sugar (**Figure 2.4.b**).



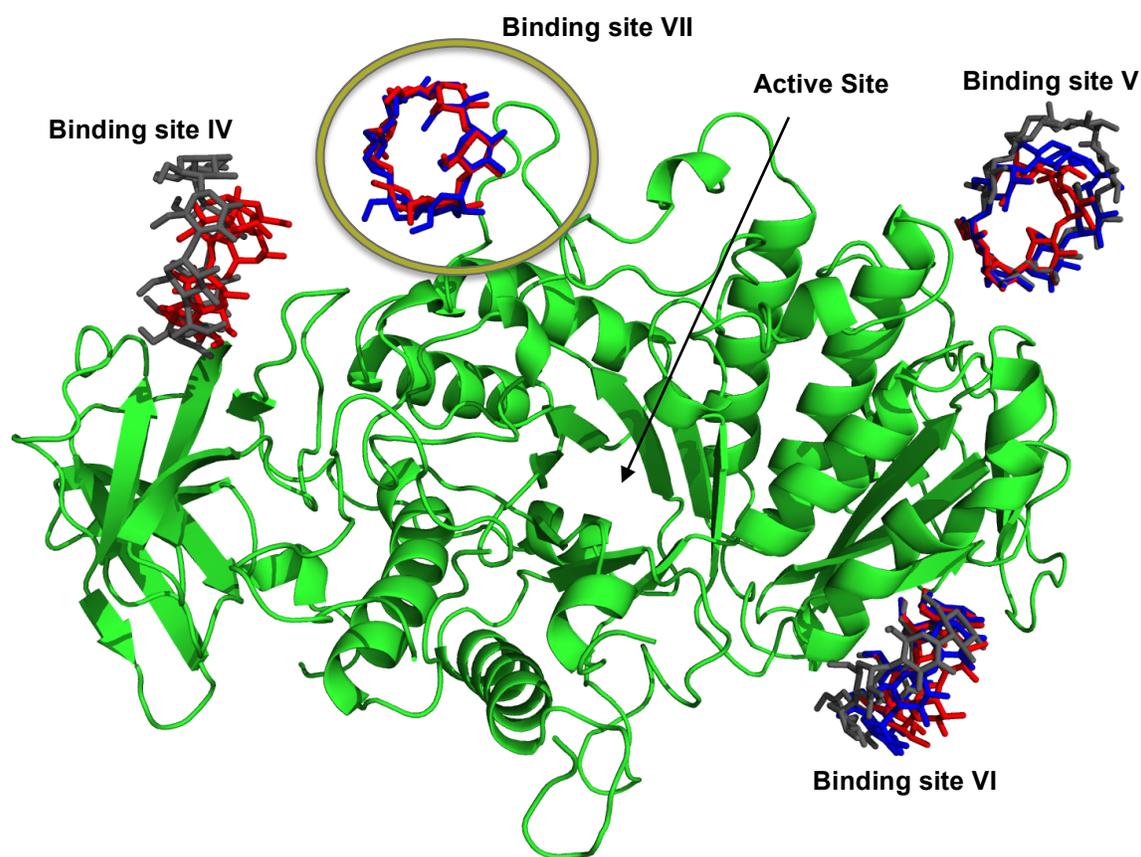
**Figure 2.4: Interactions at binding sites V and VI. a.** Detailed interactions in binding sites V. Glucan carbons are yellow and EcBE (molecule C) carbons are violet and the rest colored as in figure 2.2. Black dotted lines indicate hydrogen bonds. **b.** Detailed interactions in binding site VI. EcBE (molecule A) carbons are colored in deep teal and the rest is as above.

### 2.3.3 Cyclodextrin exclusive binding site VII

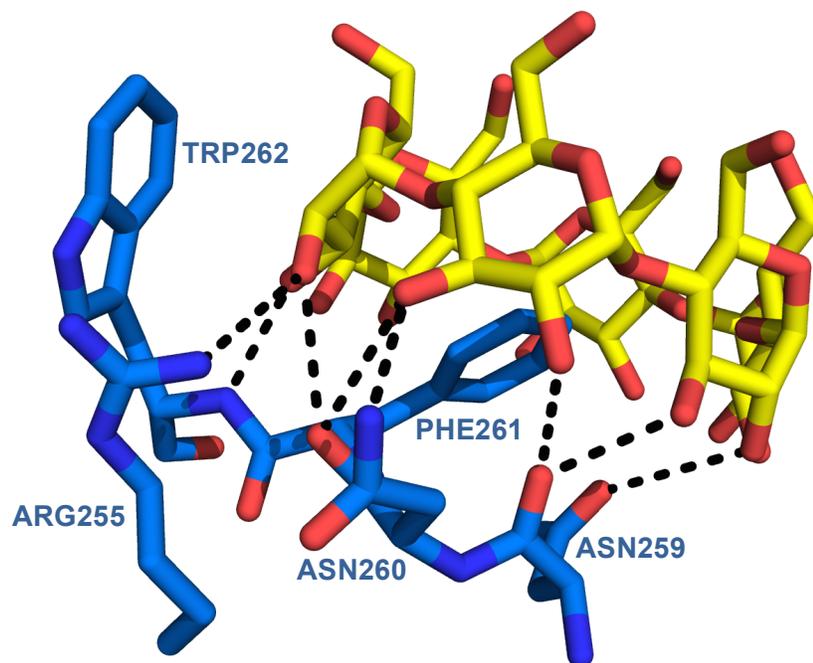
As mentioned before, EcBE soaking in cyclodextrins revealed a unique binding site not seen in the linear polysaccharide structures (**Figure 2.5**). This depiction represents the composite structure of EcBE in complex with  $\alpha$  (red),  $\beta$  (blue), and  $\gamma$  (grey) cyclodextrins. Obviously cyclic sugars containing at least six glucose units per turn will bind differently than linear saccharides; in fact, we see cyclodextrins binding only in sites IV, V, and VI. It is worth mentioning that crystal packing tremendously restricts binding of cyclodextrins in all the sites. In each binding site, a Tryptophan makes strong interactions with the carbons of a glucan, stabilizing sugar-protein binding. Furthermore, we see in each binding site at least two glucose units directly interacting with the enzyme through hydrogen bonds. Within EcBE/cyclodextrins complex, only interactions in binding site VII will be explained in details hereafter in order to conclude the presentation of all the protein residues that interact with the sugars.

Binding site VII is located on the same face of the molecule as sites IV and V. Similar to binding site I, it is about 20 Å away from the active site (**Figure 2.5**). Several hydrogen bonds are formed between protein and cyclodextrin in this site (**Figure 2.6**); an Arg255 side chain nitrogen makes one H-bond with an O2, Trp262 main chain nitrogen makes one H-bond with O3 of the same glucose unit, this same O3 makes another hydrogen bond with the main chain carbonyl of Asn260. The side chain nitrogen of Asn260 makes an H-bond with O3 of a neighboring glucose, and O2 of the same glucose makes a hydrogen bond to the side chain carbonyl of Asn259; which also binds to O3 of a third glucose. Finally,

O2 of the third glucose makes a hydrogen bond to the side chain oxygen of Asn259. In addition, the cyclic sugar surrounds an exposed phenylalanine residue in this site (Phe261) making a strong hydrophobic synergy. Another residue, Trp262, stacks against the sugar securing the association between the sugar and the protein in this site.



**Figure 2.5: The overall structure of EcBE in complex with cyclodextrins.** Cartoon of molecule B is displayed in green, cyclodextrins are represented by sticks and are colored as follows:  $\alpha$  (red),  $\beta$  (blue), and  $\gamma$  (grey). In binding sites V and VI, we see all three cyclodextrins bound, but in site IV, only  $\alpha$ , and  $\gamma$  bind and in site VII, only  $\alpha$ ,  $\beta$ . Binding site VII, unique to cyclodextrins, is encircled in yellow and its interactions are explained in details in figure 2.6.



**Figure 2.6: Binding site VII interactions.** Residues from molecule D of EcBE are shown in blue with C in marine, O in red and N in dark blue. The  $\alpha$ -cyclodextrin is also colored by atom with C in yellow and O in red. Hydrogen bonds are dotted black lines. Note that Trp262 and Phe261 stack against the sugar equidistant from the saccharide stabilizing therefore the binding in this site.

## 2.4 What Residues to Mutate Based on Binding Sites Discovered

Seven binding sites were discovered in all complexes of EcBE with hexose and heptaose, and with cyclodextrins. All binding sites were exposed on the surface of the protein and the active site lacked any binding. Previous biochemical studies indicated that *E. coli* branching enzyme truncated at the amino-terminal residue numbered 112 (N112BE) predominantly transfers chains of twelve glucose units<sup>14-16</sup>. In addition, the active site residues seem to have less affinity towards the sugar residues than the external binding sites. Also, the active site was unobstructed by crystal packing. Still we did not see binding in the active site. All of these reasons perhaps provide an explanation to the lack of binding in the main groove of the enzyme.

**Table 2.1** lists the molecules occupied by each binding site, the residues involved in the interactions within the site, and the suggested mutations for each site. Every residue side chain directly interacting with a sugar were mutated to Alanine, a commonly used screen mutation with a methyl group as side chain, for being non-bulky, chemically inert, and able to conserve the protein's secondary structure. In addition to neutralizing the residue by mutating to Ala, we also chose to test the effect of binding in each site on protein activity by reversing the charge of any charged or partially charged amino acid. Examples of such mutations are R255D, R576D, D585K, and E590R in binding site I; D537K in II; K189D and Q211R in IV, D542K and Q545K in V; D505K in VI, and N260K in VII. Mutations are defined by a one-letter code representing the original residue, followed by a number showing the residue position in the protein sequence, then by a one-

letter code symbolizing the target mutation. In the case of hydrophobic residues, we aimed at changing the size and/or eliminating the aromaticity and/or polarity of the amino acid, deleting the stacking against the sugar. Instances of these mutations include H587F (binding site I), W595L (II), W159L (IV), W544K (V), L512W and W628R (VI), F261W, W262L, and W262K (VII).

Two unique mutations are E215W in binding site IV and S689V in site V. If we carefully examine the binding sites, we will see that there are *at least* two aromatic amino acids making close hydrophobic interactions in each site except for binding site IV. This site contains only one aromatic residue, Trp159, stacking against the sugar. Moreover, Glu215 contributes only to one among several hydrogen bonds present in this site. Both of these reasons lead us to believe that mutating Glu215 to a tryptophan might actually increase the activity of the protein. In binding site V, Ser689 makes one H-bond with the sugar. Mutating this residue to a Valine destroys the hydrogen bond but conserves an almost similar size of residue around the sugar.

Alongside the single mutations laid out above, we also planned for several double or triple mutants within the same site; however, some of these were hard to express or to purify. **Table 2.2** lays out the successful multiple mutations, which happened to be in binding sites II and IV. After assaying the activity of all single mutants, we therefore designed combination mutants to test either an increase or decrease in activity depending on the single mutation assay result. In the following paragraphs, we outline the activities of the mutants relative to the native protein's data. We then classify binding sites according to relevance for

protein activity, and suggest a scenario for mode of action of *E. coli* branching enzyme.

<b>Binding site</b>	<b>Molecules occupied</b>	<b>Residues Involved in binding</b>	<b>Suggested mutations</b>
<b>I</b>	A, B, C and D	R255 R576 S583 D585 H587 E590	R255A, R255S, R255D R576A, R576D S583A D585A, D585K H587A, H587F E590A, E590R
<b>II</b>	B and D	D537 K546 W595 H596	D537A, D537K K546A W595A, W595L H596A
<b>III</b>	A and B	F239 R468 P469 F477 W478	F239A R468A P469A F477A W478A
<b>IV</b>	B and D	W159 K189 Q211 E215	W159A, W159L K189A, K189D Q211A, Q211R E215A, E215W
<b>V</b>	Only C	D542 W544 Q545 S689 P659	D542A, D542K W544A, W544K Q545A, Q545K S689A, S689V P659A
<b>VI</b>	A, B, and D	L512 D505 W628	L512A, L512W D505A, D505K W628A, W628R
<b>VII</b>	A, C, and D	N260 F261 W262	N260A, N260K F261A, F261W W262A, W262L, W262K

**Table 2.1: Outline of suggested mutations for the seven binding sites discovered based on EcBE / linear saccharides complex.**

Binding Site	Mutations
II	H596A + W595L + K546A
IV	K189A + Q211A
IV + VII	W262A + K189D
III + VI	W478A + W628A
III + VI	F477A + L512W

**Table 2.2: Multiple mutations within a binding site or across sites.**

## **2.5 Results and Discussion**

### **2.5.1 N112BE expression vector and purification** (more details in chapter 6)

Truncated EcBE was cloned into a modified pet28a vector (Novagen), and therefore contained a His-tag. Purification was performed using a Ni-NTA Agarose (Qiagen) column, which gave pure protein, with a single minimal impurity band. The native protein was further purified on a gel-filtration chromatography column for purposes of crystallization. This step gave a highly pure protein with a single visible band on the graph. All the mutants used in the activity assays, on the other hand, were purified in one step, using a His-Tag affinity column.

### **2.5.2 Activity assay (N112BE)**

Branching enzymes have been tested for activity since the 1970s<sup>4, 17, 18</sup>. The most commonly used assay, namely the decrease in absorption of a glucan-iodine complex, was developed by Boyer, C. and Preiss, J. in the Biochemistry department of Michigan State University (MSU)<sup>4</sup>. Purified branching enzyme is incubated with amylose, and aliquots are withdrawn in at various times. In order to cease the reaction, diluted iodine-iodide solution<sup>19</sup> is added and samples are monitored by the change in absorbance of the glucan-iodine complex at 660 nm wavelength<sup>20</sup>. Please refer to Chapter 6 for full details on the assay protocol.

In order to determine the most appropriate amounts of amylose and protein to use, we carried out several trials using an amylose concentration of 100 µg in a

total reaction volume of 0.1 mL (which equals 1mg/mL)<sup>4</sup> as a starting point.<sup>4, 12</sup> Since the protein concentration in the assay was not reported previously, we assayed the truncated wild type (wt) protein using a range of protein concentration (**Table 2.3**). At 1 µg of protein, the iodine-glucan absorbance was stable throughout the 30 min reaction, indicating the quantity of protein to be insufficient for detection; while at 5 µg, only the first two data points were linear (**Figure 2.7**), with a higher specific activity at lower amylose concentration. When 20 µg of protein was used, the specific activity was much lower than the literature value. Only when 30 µg of protein was used did we reproduce the 40% difference in activity between wt (FL-EcBE) and truncated protein.<sup>12, 14</sup> Parallel to the experiments reported here, Hovde, S. performed assays on the FL-EcBE and mutants in Dr. Henry's laboratory in the Biochemistry department of MSU. Assays show that N112BE specific activity is 4.18E-01 versus 5.67E-01 for the full-length, N112BE representing about 73% of full-length.

<b>Protein Amount per Reaction (µg)</b>	<b>Result or Specific Activity (U/mg)</b>
1	≈ 0
5	Non-reliable
20	7.40 E-02
30	4.18 E-01

**Table 2.3: Effect of varying protein amount on specific activity.**

The wt N112BE (purified over Nickel resin only) was also tested for activity in the presence of varying amounts of amylose, 1, 2, 3, and 4 mg. This experiment shows that the increase in specific activity is correlated with the increase in amylose (**Table 2.4**). When the reaction sample contains 1 mg of amylose, a specific activity of 0.418 U/mg results. It reaches a value of 0.927 U/mg for 4 mg of amylose. As amylose is not very soluble even in basic solution, a concentration of amylose more than 4 mg/0.1 mL is oversaturated; we used 1mg of amylose per reaction sample in all succeeding assays. This gave reproducible results over a relatively large range of activities.

<b>Amylose Amount per Reaction (mg)</b>	<b>N112BE Specific Activity (U/mg)</b>
1	4.18 E-01
2	5.26 E-01
3	7.61 E-01
4	9.27 E-01

**Table 2.4: Variation of specific activity of native N112BE with increase in amylose amounts.**

We also assayed N112BE purified by both affinity and gel filtration chromatography. In this experiment, we varied the amylose concentrations while keeping the protein at 20 or 30  $\mu\text{g}$  (**Table 2.5**). This highly pure protein shows increased activity with increasing amylose concentration. For 20  $\mu\text{g}$  of protein, the activity for a reaction containing 1.6 mg of amylose was 0.87 U/mg and 1.09 U/mg for 2.4 mg amylose. The latter is comparable to 30  $\mu\text{g}$  of protein and 1.6

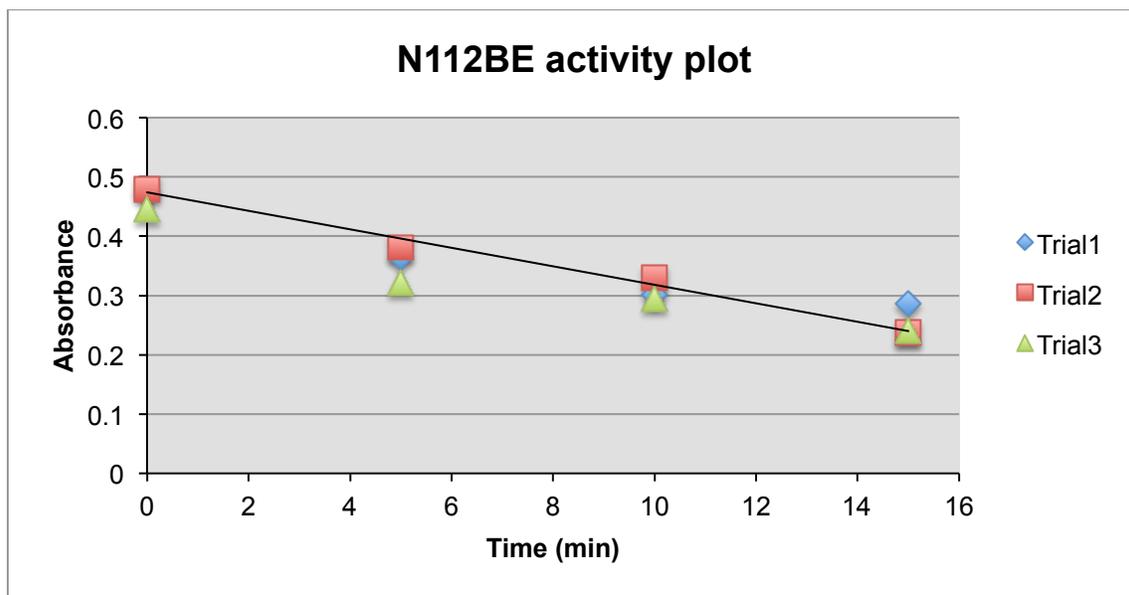
mg of amylose and the activity reaches 1.31 U/mg when we use 2.4 mg amylose per reaction tube. These are substantially higher than that seen for the protein that is only affinity purified.

Protein ( $\mu\text{g}$ )	20		30	
Amylose (mg)	1.6	2.4	1.6	2.4
Specific Activity (U/mg)	0.87	1.09	1.08	1.31

**Table 2.5: Specific activity of N112BE at different amylose concentrations.**

After running the iodine-glucan complex assay, we plot a graph of absorbance versus time. **Figure 2.7** shows an example for the native N112BE. We assayed each protein sample three times simultaneously in the interest of minimizing variables in substrates or instrument and therefore obtained three data sets per protein. The change in absorbance (y-axis) at a wavelength of 660 nm is recorded in 5-minute intervals (min) (x-axis) over a 30-minute total reaction time. The plot of absorbance versus time is linear for the first 10-15 min, but then deviates significantly, reaching a plateau indicating the depletion of unreacted amylose in the reaction sample. We used the linear part of the graph to calculate the slope. The standard deviation of the three slopes represents the error of the three trials for each sample. One unit of activity is defined as a decrease in

absorbance of 1.0 absorbance unit per min at 660 nm<sup>4</sup> and is measured in U/mg of protein.<sup>4, 21</sup>



**Figure 2.7: Absorbance versus time for the native N112BE.** Three trials are shown. The decrease in absorbance is linear for the first 10-15 min then becomes a curve and reaches a plateau.

### 2.5.3 Single mutations

We report in **table 2.6** the statistics for the wt and mutants of N112BE for the iodine stain assay. The activity per milligram (mg) of protein, activity error propagation, percent activity and percent activity error propagation are calculated as follows:

$$\text{Activity (U / mg)} = \frac{\text{Absorbance / min}}{\text{Protein Amount (mg)}} = \frac{\text{slope}}{0.03 \text{ mg}}$$

$$\begin{aligned} \text{Activity Error Propagation} &= \text{Activity} \times \sqrt{\left(\frac{\text{StDev of slope}}{\text{slope}}\right)^2 + \left(\frac{\text{Concentration Error}}{\text{Protein Concentration}}\right)^2} \\ &= \text{Activity} \times \sqrt{\left(\frac{\text{StDev of slope}}{\text{slope}}\right)^2 + (\approx 0)} \\ &= \text{Activity} \times \sqrt{\left(\frac{\text{StDev of slope}}{\text{slope}}\right)^2} \end{aligned}$$

$$\text{Percent Activity} = \frac{\text{Protein Activity (U/mg)}}{\text{WT Activity (U/mg)}} \times 100$$

$$\text{Percent Activity E.P.} = \text{Percent Activity} \times \sqrt{\left(\frac{\text{Activity E.P.}}{\text{Activity}}\right)^2 + \left(\frac{\text{WT Activity E.P.}}{\text{WT Activity}}\right)^2}$$

(E.P. = error propagation; WT = wild type)

$$= \text{Percent Activity} \times \sqrt{\left(\frac{\text{Activity E.P.}}{\text{Activity}}\right)^2 + \left(\frac{6.30\text{E-}03}{4.18\text{E-}01}\right)^2}$$

$$= \text{Percent Activity} \times \sqrt{\left(\frac{\text{Activity E.P.}}{\text{Activity}}\right)^2 + 2.27\text{E-}04}$$

Binding Site	Mutation	Activity per mg of protein (U/mg)	Activity Error Propagation	Percent Activity (%)	Percent Activity Error Propagation
	<b>WT Truncated</b>	<b>4.18E-01</b>	<b>6.30E-03</b>	<b>100</b>	<b>2.1</b>
<b>I</b>	E590A	5.15E-02	3.74E-04	12	0.2
	E590R	3.63E-01	2.50E-02	87	6.1
	S583A	8.57E-01	2.89E-02	205	7.6
	D585A	5.07E-01	4.71E-04	121	1.8
	D585K	1.33E-01	4.90E-03	32	1.3
	R255A	1.96E-04	5.37E-03	2	1.7
	R255D	1.55E-01	1.69E-02	37	4.1
	R255S	4.51E-01	2.40E-02	108	6.0
	R576A	3.70E-01	4.71E-03	89	1.8
	R576D	5.77E-01	9.90E-02	138	23.8
	H587A	4.54E-01	2.78E-02	109	6.9
	H587F	2.02E-01	1.75E-03	48	0.8
<b>II</b>	W595A	9.50E-02	4.18E-03	23	1.1
	W595L	2.44E-01	3.64E-03	58	1.2
	D537A	2.23E-01	2.83E-02	53	6.8
	D537K	6.70E-01	2.83E-02	160	7.2
	K546A	6.33E-01	1.89E-02	152	5.1
<b>III</b>	R468A	4.89E-01	6.67E-03	117	2.4
	W478A	2.17E-01	1.89E-03	52	0.9
	F239A	6.89E-01	2.83E-02	165	7.2
	F477A	9.92E-01	1.65E-02	237	5.3
	P469A	2.08E-01	3.03E-03	50	1.0
<b>IV</b>	K189A	2.97E-01	6.60E-03	71	1.9
	K189D	1.11E-01	8.42E-04	26	0.4
	Q211A	2.65E-01	2.60E-02	63	6.3
	Q211R	5.18E-02	4.55E-03	12	1.1

**Table 2.6: Iodine assay statistics of the N112BE wild type and mutants.** WT represents wild type truncated protein; CD is the catalytic domain.

**Table 2.6 (cont'd)**

<b>Binding Site</b>	<b>Mutation</b>	<b>Activity per mg of protein (U/mg)</b>	<b>Activity Error Propagation</b>	<b>Percent Activity (%)</b>	<b>Percent Activity Error Propagation</b>
<b>IV</b>	W159A	1.23E-01	6.28E-03	30	1.6
	W159L	9.73E-02	1.70E-02	23	4.0
	E215A	6.43E-01	2.36E-02	154	6.1
	E215W	1.36E-01	1.64E-02	33	4.0
<b>V</b>	P659A	3.95E-01	1.89E-03	95	1.5
	S689A	1.36E-01	8.08E-04	33	0.5
	S689V	5.47E-01	9.43E-03	131	3.0
	D542A	2.84E-01	5.19E-03	68	1.6
	D542K	4.68E-01	9.43E-03	112	2.8
	W544A	1.04E-01	8.75E-03	25	2.1
	W544K	1.00E-01	2.43E-03	24	0.7
	Q545A	1.46E+00	1.41E-02	349	6.3
Q545K	6.09E-01	2.50E-02	146	6.4	
<b>VI</b>	L512A	4.11E-01	1.32E-02	98	3.5
	L512W	1.49E+00	9.43E-03	356	5.8
	D505A	4.04E-01	2.36E-03	97	1.6
	D505K	3.87E-01	1.70E-02	93	4.3
	W628A	3.96E-02	6.20E-03	9	1.4
	W628R	1.88E-02	1.11E-03	5	0.3
<b>VII</b>	W262A	3.30E-02	1.68E-04	8	0.1
	W262L	4.00E-01	1.23E-02	96	3.3
	W262K	6.00E-02	5.57E-03	14	1.3
	N260A	2.65E-01	1.98E-02	63	4.8
	N260K	3.75E-01	1.91E-02	90	4.8
	F261A	4.19E-01	1.51E-02	100	3.9
	F261W	6.18E-01	1.65E-02	148	4.5
<b>CD</b>	<b>D405A</b>	<b>2.38E-03</b>	<b>1.35E-02</b>	<b>5.7E-03</b>	<b>3.2E-02</b>

Clearly, the mutants from different binding sites have diverse effects on the protein's activity, some decrease while others increase the activity by small or great amounts. For easier comparison, we will rely on percent activity from **table 2.5** in the following discussion. In binding site I, two mutants show very low activity, E590A at 12% and R255A at 2%. Each one of these single mutations can knock down the branching activity tremendously, especially R255A that resembles the active site mutation D405A (1% activity), indicating the necessity of this binding site to the enzyme's reaction. On the contrary, several mutants in this site increase the activity of the protein, some slightly, such as H587A (109%) and R255S (108%), and others more effectively, like R576D (138%) and D585A (121%), while S583A elevates the activity by more than two fold (205%). In **Figure 2.2** representing the detailed interactions between EcBE and heptaose in binding site I and II, Arg255 and Glu590 each make a single H-bond with the sugar using the main and side chain, respectively, and are not therefore expected to have a shocking effect on the activity as such. This low activity however could be explained by the fact that when the protein binds with long linear saccharides these two residues interact more closely with the substrate than what we see with a heptaose molecule. On the other hand, R255S (108%) and H587A (109%) seem to not interfere with the binding ability of the protein to the sugar. Although the polarity is reversed in a mutant like R576D or deleted in D585A, the activity still increased, probably due to the position of the residue relative to the binding site and to the ability of aspartate (R576D) to make hydrogen bonds with the sugar as Arg did. S583A is surprisingly very active

since the Serine makes a single hydrogen bond with the sugar using its side chain and the Alanine is just slightly smaller than Ser. For both D585A and S583A, the increase in activity could be the result of other binding made only in the mutant protein and prevented in the presence of the native residue.

Binding sites II and III seem to be less essential sites. In binding site II, there is no significant decrease in activity, except for W595A that is at 23%. In every binding site, there is a Tryptophan that stacks against the sugar and is essential for the binding. Mutating Trp595 to a Leucine only partially fulfills the binding necessity in this site, as W595L is 58% active. In site III, hydrophobic interactions seem to be more important, since W478A is 52% active, and P469A is 50%. In these two sites, we also see an increase in activity for D537K (160%) and K546A (152%) in site II, and R468A (117%), F239A (165%), and F477A (237%) in site III. **Figure 2.2** shows that D537 and K546 are the only two residues making direct hydrogen bonding interactions with the sugar in site II. Mutating these amino acids increases the activity, indicating that the sugar does not necessarily prefer to bind in this site. Similarly in site III, F239A and F477A increase the activity while inhibiting the hydrophobic interactions with the sugar that existed in the native protein. Only two glucose units are visible in this site (**Figure 2.3.a**), F239 and F477 might be enforcing a certain direction in binding, which could have a larger effect on a longer glucan.

Binding site IV contains several mutations that cause a decrease in activity, such as W159L (23%), W159A (30%), K189D (26%), E215W (33%), one other that has a larger negative effect, Q211R (12%), and another that increases the

activity of the native protein, E215A (154%). **Figure 2.3.b** shows the stacking of W159 against the sugar, which explains why disrupting this interaction greatly affects the activity. Reversing the charge in Lys189 by the K189D mutation decreases the activity. It is possible that an Aspartate in position 189 is very close to Asn211 and Glu215 that together make a large area of negative charge; this environment stabilizes a Lys more than an Asp. One other residue, E215, binds with the same glucose unit, specifically to the same oxygen, to which binds Q211. When Glu215 is mutated to a Trp, its size probably blocks binding in this site reducing protein activity to 30%. On the opposite side, an Ala in position 215 is shorter than Glu allowing for binding without interaction with the sugar and therefore increases the activity (154%).

In binding site V, mutation of Trp544 to a Lys or Ala decreases the activity to 24 and 25% respectively, showing this residue to be important for the site and therefore the protein's activity. Ser689 causes an increase in activity when mutated to a Valine (131%) and a decrease when changed to Ala (33%). In **Figure 2.4.a**, we see S689 making a single side chain H-bond with the sugar; an Ala is too short for this position to interact with the saccharide while a Val could be making some hydrophobic interactions increasing binding in this site. In position 542, a positive charge looks more favorable than a negative one, giving D542K a little higher activity (112%) when compared to the native protein. One unique residue is Q545, because it does not directly bind to the sugar but to a neighboring residue, D542, yet mutation to Ala increases the activity three fold and mutation to a Lys also increases the activity (146%). Replacing a Gln by Ala,

a very small residue, leaves a larger available space for the sugar to bind more freely, while a floppy Lys could also increase the activity just for the flexibility of binding in several directions.

Binding site VI is similar to binding site I in that it happens to have mutations that both increase and decrease the activity greatly. For one essential residue, W628, the mutation to Ala or Arg brings the activity down to 9 and 5% respectively. As in all other sites, Trp is important for substrate binding but the effect is greater in this site. Another mutation increases the activity by more than three fold, L512W (356%). This residue makes a hydrophobic interaction with the sugar and the position 512 may provide a stacking interaction when mutated to a Trp, in which case this site will have a Trp residue on both sides of the sugar stabilizing the binding, and thus increasing the activity.

Lastly, binding site VII, which only happens with cyclodextrins, has another critical residue for binding, W262. In **Figure 2.6**, we see Trp262 stacking against the sugar on one side of the cyclodextrin ring rather than on the inside such as Phe261. Mutation of position 262 to Ala lowers the activity to 8% and mutation to Lys to 14%. This indicates both the importance of this residue to the binding and this binding site to the activity. Surprisingly, F261A is 100% active indicating that the stacking against the sugar inside the ring is not essential while mutation of this residue to Trp, a larger completely exposed aromatic side chain, allows for more hydrophobic interactions and hence higher activity.

**Table 2.6** also shows, aside from the activity and percent activity for each sample, the error propagation on the activity. To better appreciate the amount of

error in the calculation of each sample's activity, we obtained percentage of error by dividing error propagation by activity and we present here the most relevant results for interpreting. The highest errors were for E215W ( $(1.64E-02/1.36E-01) \times 100 = 12\%$ ) and Q211A ( $(2.60E-02/2.65E-01) \times 100 = 10\%$ ). It is worth mentioning that the error on D405A is high but the activity is almost negligible. On the other hand, some mutants had the three trials of activity assay very close in values making the error really low, like Q545A (0.97%), E590A (0.73%), S689A (0.60%) and D585A (0.09%). The second error propagation reported here is on the percent activity in the last column; this value varies for each mutant but is reasonable overall, except for R576D ( $138 \% \pm 23.8$ ), however, one can not deny the high activity of this mutant despite the large error propagation. A few other mutants also have some relatively high error propagation, such as R255A ( $2 \% \pm 1.7$ ), D537A ( $53 \% \pm 6.8$ ), W159L ( $23 \% \pm 4$ ), and W628A ( $9 \% \pm 1.4$ ). It is worth mentioning that purifying this mutant failed three times (no protein on SDS-gel) before getting one batch that was less pure than other mutants, judged by SDS-page (about 80% versus over 95% for others).

## 2.5.4 Combination mutations

In addition to the single mutations, we also assayed some combined mutations. **Table 2.7** summarizes all multiple mutants. In binding site II, we designed a triple mutant H596A+W595L+K546A with the idea that this might have a greater impact on the activity than single mutations in this site. We discovered however that this mutant is 95% active although one of the mutations gives a 58% active protein. This is consistent with the idea that this site is not essential for branching activity.

In binding site IV, two single mutations K189A and Q211A are 71 and 63% active, respectively. When combined in one protein, K189A+Q211A is 55% active. This double mutant caused a moderate effect on activity. Site IV is indeed more affected by other single mutations listed above and can be considered an essential site for substrate binding.

Mutations W262A (8%, binding site VII) and K189D (26%, binding site IV) were shown to have a very large effect on the activity. In order to determine whether a combined mutation will have a greater impact on the activity than each individual mutation, we tested the double mutant and detected no measurable activity for the enzyme. We then decided to test the individual and double mutant proteins at a concentration 10 times higher than the first test. At 300 $\mu$ g instead of 30 $\mu$ g protein per sample, W262A showed 24% activity and K189D 15%, while the double mutant was only 3% active, verifying our proposal that the combination of mutants in both binding sites has a much larger effect than mutation of a single site. Since depressing both binding sites IV and VII produced

an inactive protein, we can conclude that both of these sites are important for activity.

In the same way, we also tested two mutants in both binding sites III (W478A), and VI (W628A). At 300 $\mu$ g, these mutations are 46% and 4% active when compared to the native protein, while they were at 52% and 9% with 30 $\mu$ g protein per sample. The percent activities retained close values between one protein concentration and another ten times greater. The double mutant here shows an almost dead protein (2%) indicating the necessity of these binding sites to the protein's activity.

One more double mutant also in binding sites III and VI, F477A+L512W, showed high activity percentage for each separate mutation, 237% for F477A and 356% for L512W. When both mutations are combined within one protein, they do not show an increase in activity (289% for the double mutant).

Based on all the information produced from the iodine-stain assay, we can confirm that all the binding sites, save site II, identified structurally are important for the activity of the enzyme.

Binding Site	Mutation	Activity / mg of protein (U/mg)	Activity Error Propagation	Percent Activity (%)	Percent Activity Error Propagation
II	H596A+W595L +K546A-30µg	3.97E-01	1.23E-02	95	3.3
IV	K189A+Q211A-30µg	2.31E-01	6.60E-03	55	1.8
VII	W262A-300µg	9.82E-02	4.01E-03	24	1.0
IV	K189D-300µg	6.13E-02	1.27E-02	15	0.2
IV+VII	W262A+K189D	1.28E-02	2.07E-03	3	0.5
III	W478A-300µg	1.92E-01	8.49E-03	46	2.1
VI	W628A-300µg	1.51E-02	1.89E-04	4	0.1
III+VI	W478A+W628A	6.67E-03	--	2	3.0E-02
III	F477A-30µg	9.92E-01	1.65E-02	237	5.3
VI	L512W-30µg	1.49	9.43E-03	356	5.8
III+VI	F477A+L512W	1.21	9.43E-03	289	4.4

**Table 2.7: Combination mutations tested from all binding sites.** We do not report an error for the double mutant W478A+W628A because only one trial gave a negative slope (the absorbance increased in the other trials for this mutant, which also indicates it is absolutely inactive).

### 2.5.5 Activity assay (full-length EcBE)

As mentioned earlier, Stacy Hovde from Dr. Henry's laboratory constructed several mutants of the full-length protein. We report in **table 2.8** the percent activity of these mutations and compare them to the truncated N112BE mutants.

Overall, we clearly notice a more dramatic effect in N112BE mutants versus the full-length protein. In FL-EcBE, the highest percent activity reaches 155% for mutants Q545A and W595L, while in N112BE, L512W hits 356% and Q545A, 349%. These percentages are relative to the corresponding wt full-length or truncated respectively. The actual activity of the full-length native protein obtained is 0.5667 U/mg, N112BE – L512W is 1.49 U/mg and Q545A is 1.46 U/mg. These mutations lead actually to regaining the native full-length protein's activity. This is very important result because we know that the truncated protein loses about 40% of its activity compared to the full-length, so the far N-terminal must be assisting the protein in the activity somehow. Mutations Q545A and L512W actually overcome this deficiency in activity. Within the common mutations between the 2 proteins, there are however, in FL-EcBE, 23 mutants that are over 100% active compared to 17 in N112BE, acknowledging the fact that the very active mutants are not the same in both proteins. In addition, there are 6 mutants besides the catalytic domain one that are less than 12% active in N112BE, in opposition to 4 in FL-EcBE.

Moreover, the majority of the mutations do not have the same effect on the activity when examining the full-length and the truncated proteins. In binding site I, three mutations drive the activity in opposite directions, E590A (12% in

N112BE versus 103% in FL-EcBE), R255D (37% in N112BE versus 112% in FL-EcBE), and H587F (48% in N112BE versus 107% in FL-EcBE). Other mutations are shockingly different between the two proteins, those include, W595A (23% in N112BE versus 106% in FL-EcBE) in binding site II, F477A (237% in N112BE versus 9% in FL-EcBE) in site III, E215W (3% in N112BE versus 98% in FL-EcBE) in site IV, S689A, W544A, and W544K (33, 25, and 24% in N112BE versus 99, 113, and 127% in FL-EcBE respectively) in site V, and finally L512A (98% in N112BE versus 5% in FL-EcBE) in site VI. Binding site VII is uniquely fairly uniform in this regard.

The deviation and sometimes contradiction in the effect mutants have on the protein's activity may be related, at least partially, to the nature of the polypeptide's tertiary structure. An extra N-terminal domain consisting of 112 amino acids is missing in N112BE. This distinct functional unit appears to assist the protein in the binding mode during the branching reaction. The specific interactions of this domain remain obscure until the crystal structure of the full-length protein bound to the substrate is determined. Nevertheless, we can use the knowledge thus acquired to test the chain length specificity of *E. coli* branching enzyme's mutants. In each binding site, we chose a few mutants that give similar or very different data to test by chain length specificity (last column in **table 2.8**).

Binding Site	Mutation	Percent Activity (%)		Test in Branch Assay
		N112BE	FL-EcBE	
	<b>Native</b>	<b>100</b>	<b>100</b>	<b>X</b>
<b>I</b>	E590A	12	103	
	E590R	87	83	X
	D585A	121	68	X
	D585K	32	16	X
	R255A	2	111	
	R255D	37	112	
	R255S	108	118	X
	R576A	89	109	
	R576D	138	151	
	H587A	109	95	
H587F	48	107	X	
<b>II</b>	W595A	23	106	
	W595L	58	155	
	D537A	53	76	X
	D537K	160	126	
	K546A	152	121	
<b>III</b>	R468A	117	145	X
	W478A	52	5	
	F239A	165	148	
	F477A	237	9	
	P469A	50	132	
<b>IV</b>	K189A	71	47	X
	K189D	26	26	X
	Q211A	63	48	X
	Q211R	12	31	
	W159A	30	13	
	W159L	23	55	X
	E215A	154	136	
	E215W	33	98	

**Table 2.8: Comparing truncated and full-length mutants' percent activity. Which mutations to test for Chain Length Specificity assay?**

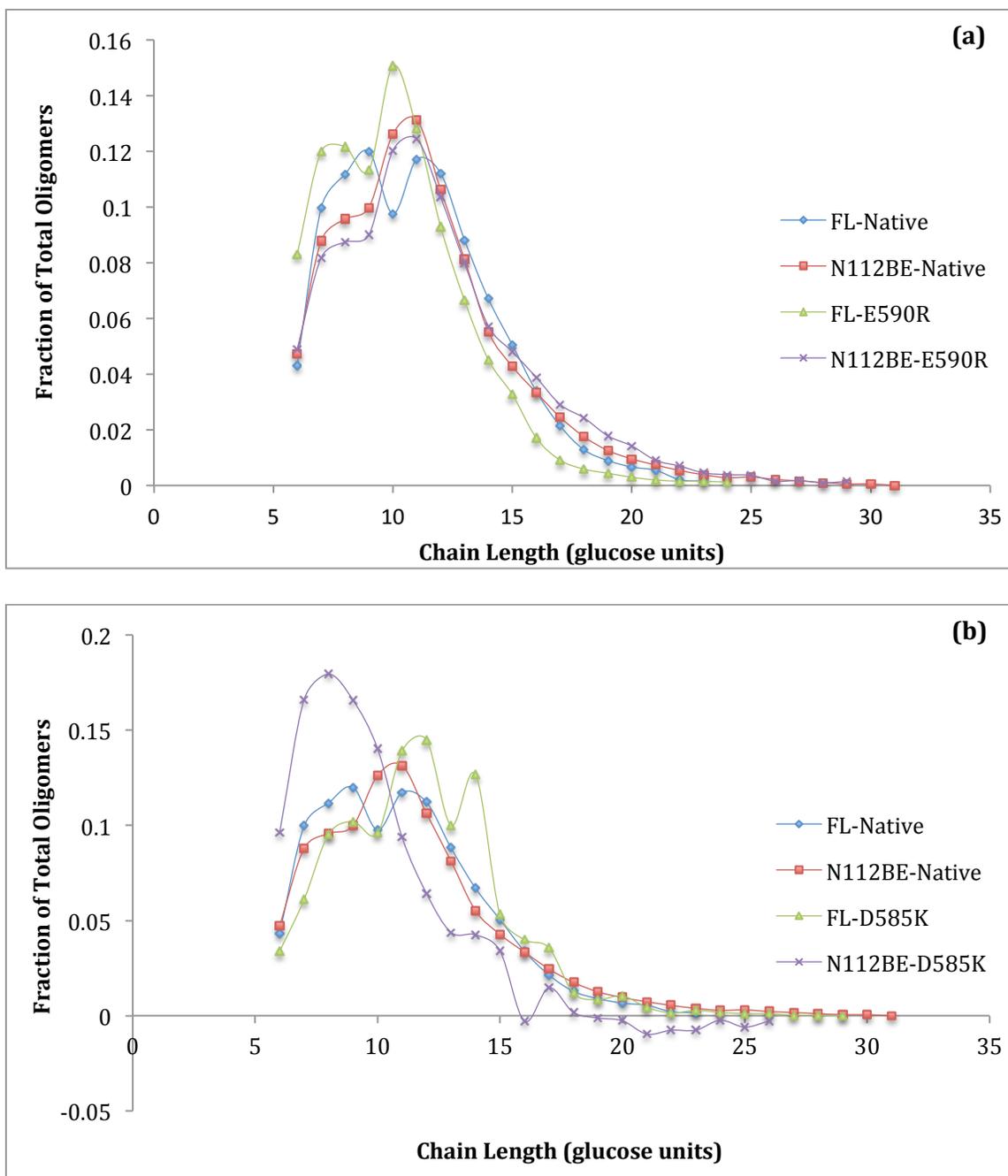
**Table 2.8 (cont'd)**

Binding Site	Mutation	Percent Activity (%)		Test in Branch Assay
		N112BE	FL-EcBE	
V	P659A	95	143	
	S689A	33	99	
	S689V	131	130	
	D542A	68	104	
	D542K	112	121	
	W544A	25	113	
	W544K	24	127	
	Q545A	349	155	X
	Q545K	146	91	
VI	L512A	98	5	
	L512W	356	139	X
	D505K	93	97	
	W628A	9	30	
	W628R	5	14	
VII	W262A	8	3	
	W262L	96	99	
	N260A	63	84	X
	N260K	90	55	
	F261A	100	82	
	F261W	148	75	X
<b>CD</b>	<b>D405A</b>	<b>5.7E-03</b>	<b>2</b>	

### 2.5.6 Chain length specificity assay

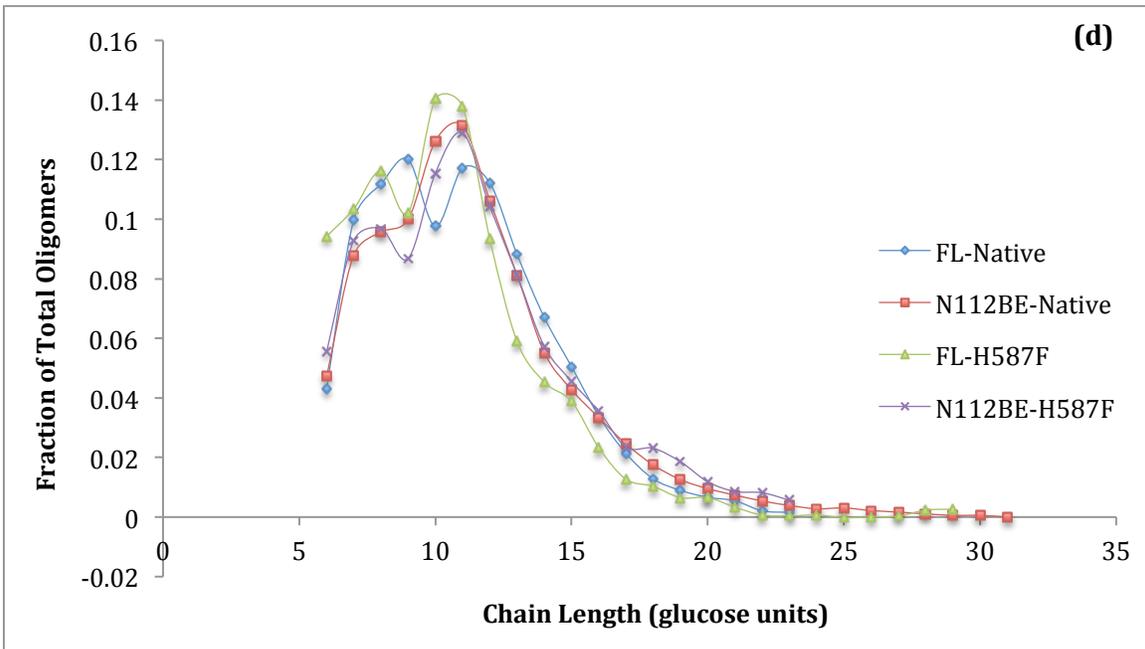
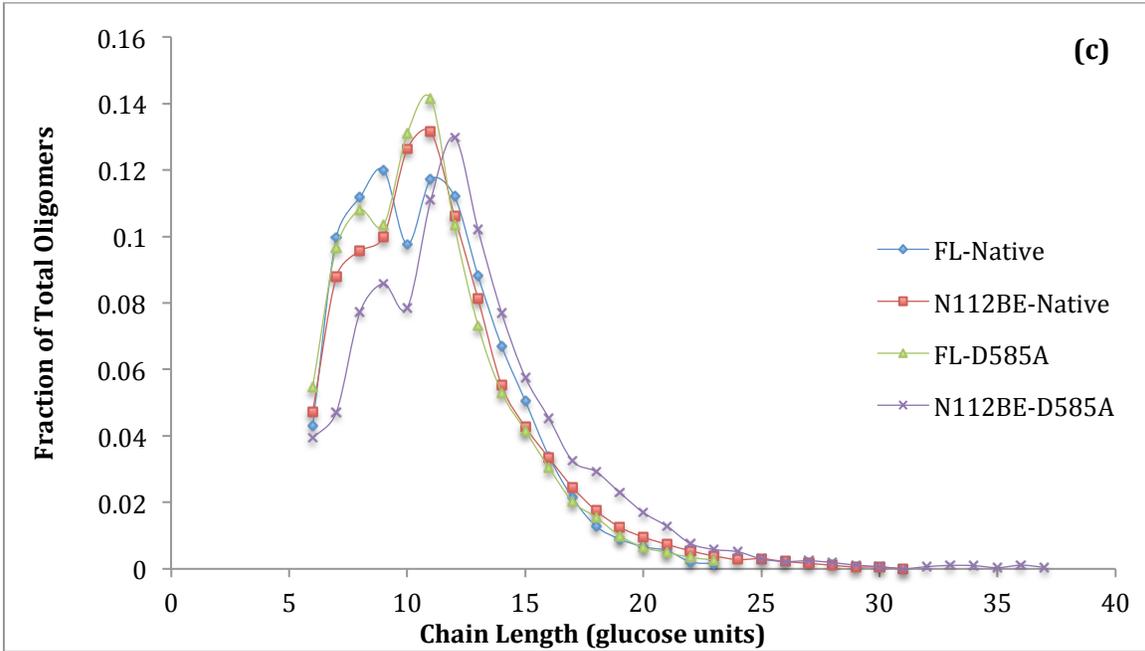
The N-terminal of wild type EcBE was proposed to influence the chain transfer pattern of the protein based on the construction of a hybrid enzyme from maize BE isoforms;<sup>22</sup> it was later shown that an N-terminal 112 amino acid truncation in EcBE,<sup>12</sup> alters the branching pattern.<sup>14</sup> Branches transferred by the truncated and wild type *E. coli* BE were analyzed by gel filtration and high performance anion exchange chromatography (HPAEC). N112BE was shown, by HPAEC, to transfer fewer chains with degree of polymerization (DP) 4-11 and a greater number of chains with DP  $\geq$  12.<sup>14</sup> It is reasonable to conclude that the mutants of both full-length and truncated EcBE may have different impacts on branch specificity.

The last column of **table 2.8** marks the mutants we decided to assay the chain length specificity for. We picked these mutants aiming at explaining the role of each site in determining the branch size of chains transferred. For this, we chose some mutations that showed similar results between N112BE and FL-EcBE and others with contradicting values, leaving inactive or almost inactive mutants out. We tested a total of 16 mutants including the native full-length and truncated proteins. We thank Dr. Andrew Mort from the department of Biochemistry and Molecular Biology at Oklahoma State University, for collaborating with us to complete the second part of this assay, HPAEC.



**Figure 2.8: Distribution of oligomers.** Nine graphs showing the chain distribution for the native full-length and truncated proteins in comparison with: **a.** E590R, **b.** D585A, **c.** D585K, **d.** H587F, **e.** R255S, **f.** K189A, **g.** L512W, **h.** N260A, and **i.** F261W. Each graph shows the individual fractions of oligomers produced (y-axis) versus the length of the chain produced in glucose units (x-axis). For each sample the fractions total 1. The highest chain length varies between mutants; with the longest chain reported on the graphs reaching 40 glucose units, while the fraction of oligomers with DP 22-40 for any protein is lower than 0.01.

Figure 2.8 (cont'd)



All graphs also share the fact that branching of oligomers below six glucose units does not occur.

Figure 2.8 (cont'd)

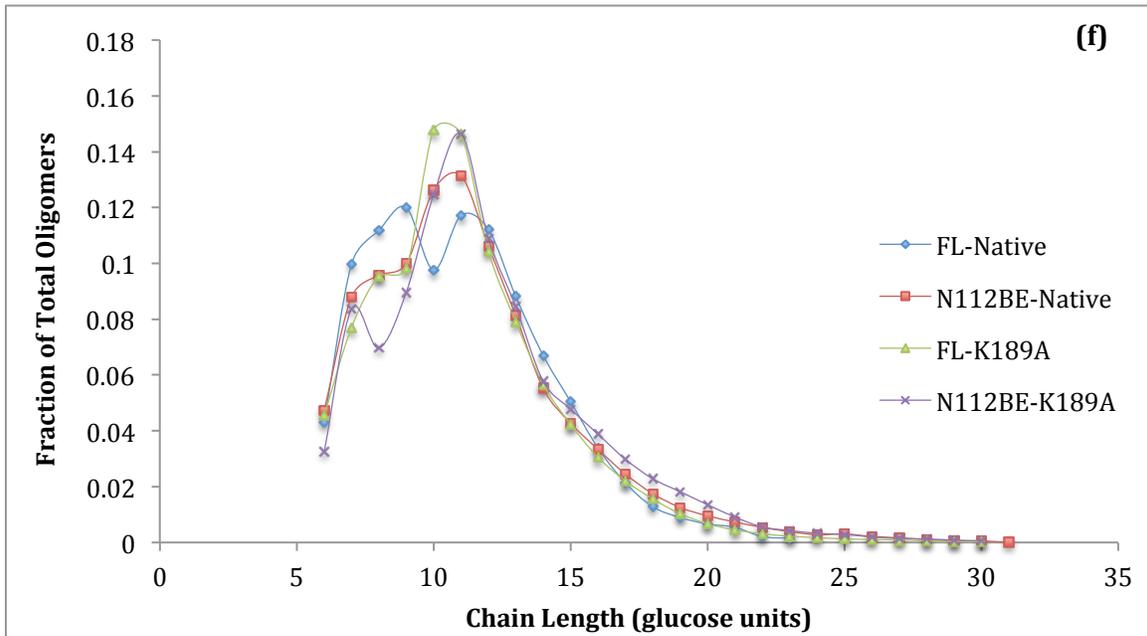
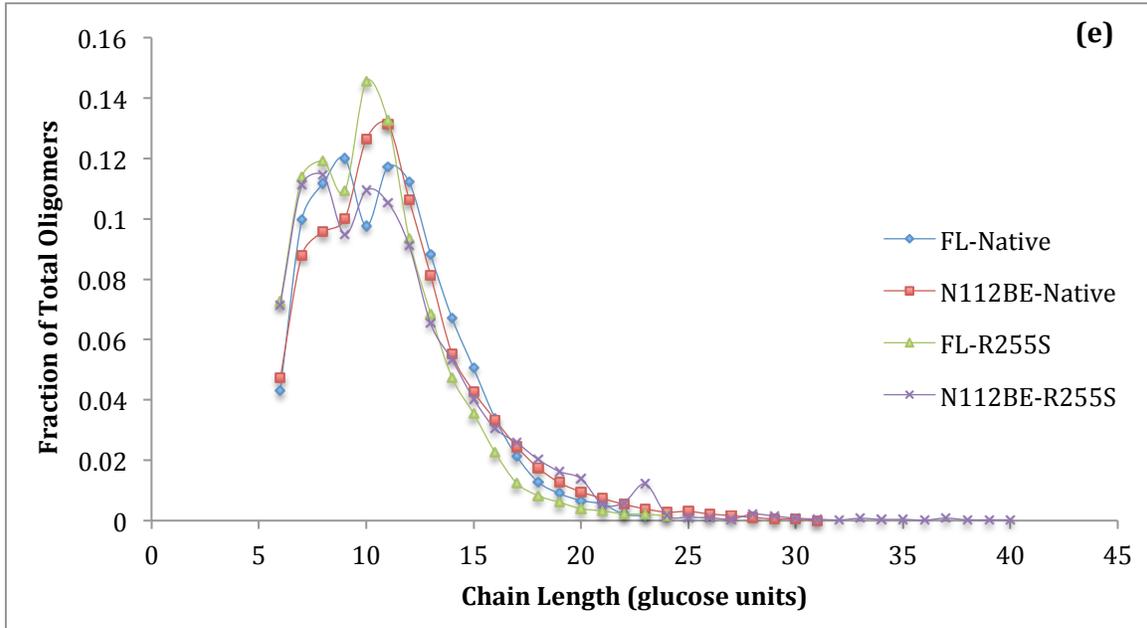


Figure 2.8 (cont'd)

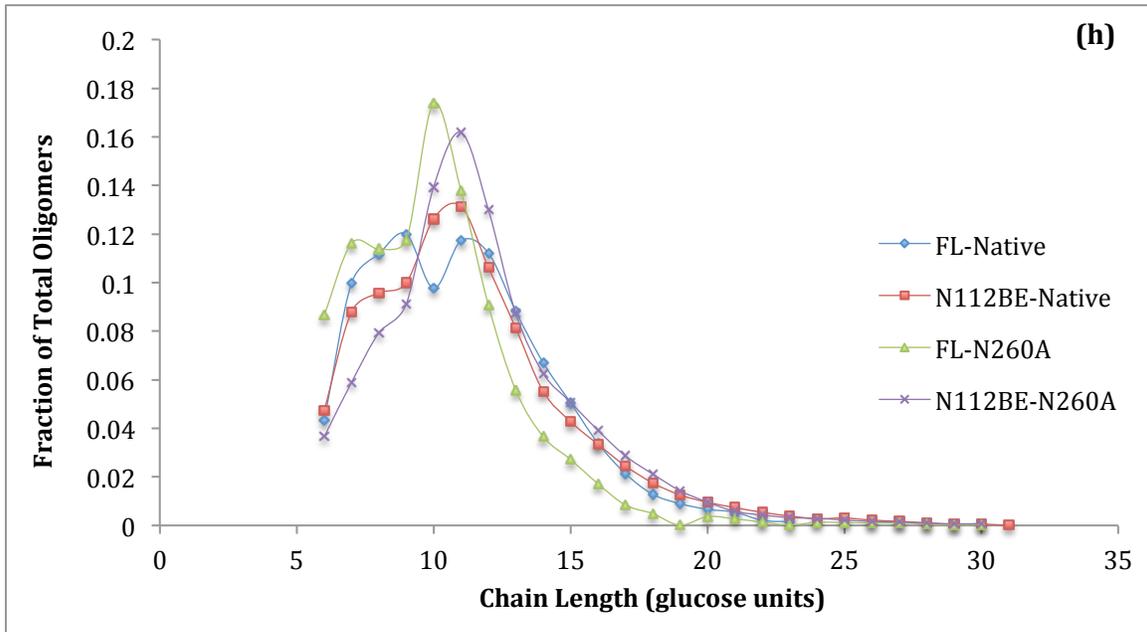
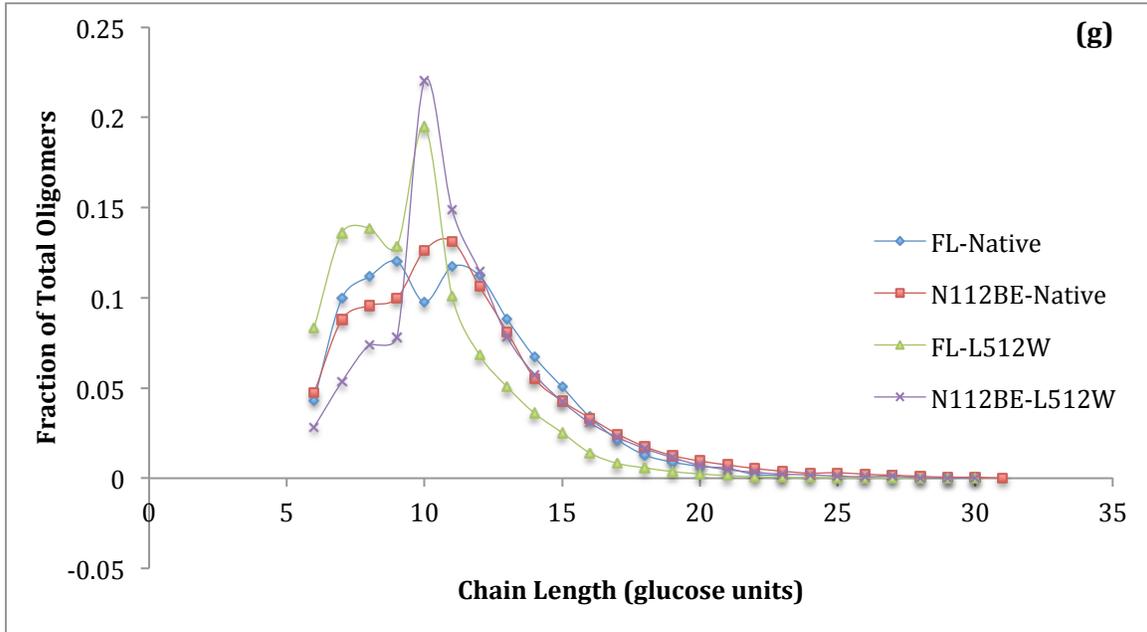
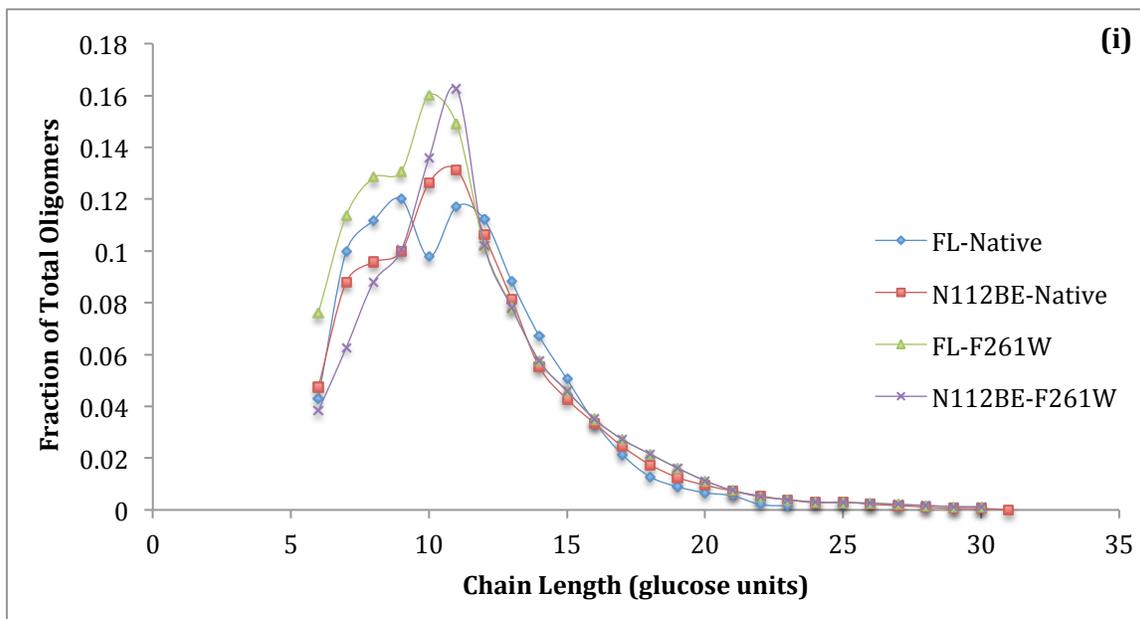


Figure 2.8 (cont'd)



The chain length distribution assay results are shown in **Figure 2.8** where the blue lines represent the full-length native protein, the red lines, the truncated native, and the green and purple lines, the full length and truncated mutant proteins, respectively. The x-axis shows the length of the oligomer chains measured in number of glucose units, and the y-axis is the individual fractions of total oligomers, making the sum of all fractions for each sample equal to 1. The highest degree of polymerization of transferred oligomers was 40 glucose units for the mutant N112BE-R255S, followed by chains of 37 glucose units transferred by N112BE-D585A, while the other mutants transferred chains of no longer than 31 glucose units; for each mutant however, the sum of all transferred

chains with DP > 22 was negligible (lower than 0.01 or 1%). We show 4 graphs in each diagram, the full-length and truncated versions of both native and mutant proteins. Graphs **a**, **b**, **c**, **d** and **e** represent E590R, D585A, D585K, H587F, and R255S, respectively, all belonging to binding site I. Graph **f** represents the mutant K189A and belong to binding site IV, while **g** to L512W and belongs to site VI. The last two graphs, **h** and **i**, represent N260A and F261W and belong to binding site VII.

The FL-native protein shows two peaks, the first with 11.1% and 11.9% of oligomers at 8 and 9 glucose units, respectively, and the second, 11.7% oligomers at DP = 11; while N112-native shows a shoulder at DP = 8 (9.5%) and DP = 9 (9.9%), then peaks higher than the FL at DP = 10 (12.6%) and DP = 11 (13%). In binding site I, mutant E590R (graph a) exhibits similar patterns to the native proteins. The FL version shows two peaks, one with 12.0% and 12.1% oligomers for DP = 7 and 8, respectively, and another with 15% oligomers at 10 glucose units. N112BE-E590R shows a shoulder with 8.7% and 9% of oligomers at DP = 8 and DP = 9, respectively, and peaks at DP = 10 (12%) – 11 (12.4%). Graph b showing D585K deviates greatly from the native proteins, as the full-length exhibits 3 peaks for DP = 9 (10.2%), 12 (14.5%) and 14 (12.6%), while the truncated has one wide peak, containing DP = 7 (16.6%), 8 (17.9%) and 9 (16.6%). D585A (graph c) on the other hand, is closer to the native proteins in branching activity, and it is quite noticeable how the FL and N112BE versions of this mutant are similar together, as they both have two peaks each. FL-D585A mostly transfers chains of DP = 8 (10.8%) and DP = 11 (14.1%), while N112BE-

D585A preferably works on oligomers of DP = 9 (8.5%) and DP = 12 (12.9%). H585F (graph d) is another similar mutant to the native proteins; the FL protein shows two peaks at DP = 8 (11.6%) and DP = 10 (14%), while the truncated version shows a shoulder at DP = 7 (9.2%) and DP = 8 (9.9%) and a peak at DP = 11 (12.9%). R255S (graph e) is the last mutant in binding site I tested for branching activity; this deviates greatly from the native proteins. The FL and N112BE proteins both peak at DP = 8 and DP = 10 with varying percentage of oligomers as follows: 11.9% and 14.5% for the FL and 11.4% and 10.9% for the truncated.

Only one mutant in binding site IV, K189A (graph f), was assayed for branching activity. FL-K189A is very different than the corresponding native protein as it shows two peaks at DP = 8 (9.5%) and DP = 10 (14.7%), while N112BE-K189A peaks twice instead of once as the native, one peak at DP = 7 (8.3%) and a more intense one at DP = 11 (14.6%). A single mutant belonging to binding site VI, L512W (graph g), was tested; this showed a much higher activity than the native proteins. FL-L512W shows a wide peak at DP = 7 (13.5%) and DP = 8 (13.8%), and another sharp peak at DP = 12 (19.5%). N112BE-L512W has a shoulder at DP = 8 (7.3%) and a much sharper peak at DP = 10 (22%). The last two mutants, part of binding site VII, resemble the native proteins in branching activity and are very active. FL-N260A (graph h) shows one peak at DP = 7 – 8 (11.6% - 11.3 %) and another peak at DP = 10 (17.3%), while N112BE-N260A shows a shoulder at DP = 8 (7.9%) and a peak at DP = 11 (16.2%). F261W also shows two peaks in the FL version of the mutant, one at

DP = 8 (12.8%) and another at DP = 10 (16%), while the truncated protein presents a shoulder at DP = 8 (8.8%) and a peak at DP = 11 (16.3%).

There is clearly no single pattern shared for all the mutants, although most mutants tested are more active than their corresponding native counterparts. The last three proteins, L512W, F261W and N260A are greatly more active than the native proteins. This is attributed in part to the fact that the N-terminal domain affects the chain transfer pattern considerably. In addition, the last three mutations greatly change the interactions with the surrounding sugars and therefore affect the activity of the entire protein. Most mutants show peaks at DP = 8 through DP = 11, with one exception, FL-D585K, which shows a pretty strong peak at DP = 14. This is partly due to the location of this residue in binding site I; it must be directly interacting with the branch to be transferred where the lysine causes a repulsion of the sugar to be held versus an attraction by the glutamate.

## 2.6 Conclusion and Hypothesis

### 2.6.1 Structure-function correlation

The truncated protein surely presents a different model than the wild type full-length. N112BE's activity is affected more dramatically by mutations since it is missing the extra N-terminal domain of the FL-EcBE protein. The highest activity seen in the truncated protein is 1.49 U/mg for L512W but does not go higher than 0.88 U/mg for Q545A in the full-length. When a mutant is inactive in both cases, the percent activity can go down to 2-3%. In addition, the same mutants do not have the same impact on both proteins and sometimes even contradicting effects. The mutation specifically modifies how a certain binding site holds the substrate during the branching reaction; we obviously saw diverse effects in all binding sites, but as a general feature, site II had the least effect, even when mutated 3 times (H596A+W595L+K546A). This site does not seem to be essential for the protein's activity for either form of the enzyme.

Binding site I is more crucial for the activity of the truncated protein than it is for the native full-length while site III plays the key role for FL-EcBE. In site I, only one mutation, D585K, strongly affects the activity in the full-length protein (16%), while this mutation has a smaller effect in the truncated protein (32%). On the other hand, two mutations in site 1 E590A and R255A, detrimentally distress only the truncated EcBE. In binding site III, W478A and F477A reduce FL-EcBE activity to 5% and 9%, respectively, while N112EcBE retains half of its activity. The structure of the *Mycobacterium tuberculosis* branching enzyme (MycoBE)<sup>23</sup>

is the most comparable structure to EcBE and shows the extra N-terminal domain, which is missing in the structures of EcBE. It is located away from the overall protein packs adjacently to the amino-terminus (on the left of **Figure 2.1.b** closest to binding site IV); we expect the same location in EcBE. Truncation of 112 amino acids reduced the activity to about 60%, so the N-terminus does participate in the reaction probably by interacting with long natural substrates. This external 112 amino acids could be correlating with binding site III, which resides within the catalytic domain and close to the active site. In the truncated protein, binding site I becomes more critical for the protein's activity than site III, since it is the closest of the sites to the active center.

Binding site IV does not impact either protein form tremendously, but sites V and VI have large effects in both forms, with greater consequences for the truncated protein. Substrates have to exclusively depend on other sites in the absence of the N-terminal 112 residues; so knocking down a site could render the enzyme completely inactive as we see in sites V and VI. Tryptophans exist in every site but their mutations in these two sites, and especially in site VI, are more prominent as W628A and W628R are 9 and 5% active in N112BE, while another hydrophobic residue plays a key role in the activity of FL-EcBE, L512W (5%). Also in site VII, W262A inactivates both protein versions, 8% for N112BE and 3% for FL-EcBE. Sites V, VI and VII seem to be imperative for an active protein.

Double mutant experiments teach us that the sites close in location work together to achieve the branching activity. For example, a protein with combined

mutations from both sites IV and VII, W262A + K189D, decreased the activity of N112EcBE to 3%. Similarly, sites III and VI associate together in holding the substrate as explained by two double mutants; one combining single mutations of inactive proteins, W478A + W628A, and another joining two mutations of highly active enzymes, F477A + L512W, giving rise to another highly active protein.

The chain length specificity assay shows that all mutants tested play a significant role in the branching activity of both protein versions, however, binding sites VI and VII were the most detrimental. These two increase the activity greatly, and the truncated doubles the branching of DP = 10 (22%). The effect exerted by the mutations only shows the importance of each binding site in the transferring of the right size of branches, and the right amount of the specific size of sugars.

### **2.6.2 Hypothesis on enzymatic activity**

Natural substrates are very long and might need to wrap around the protein before entering the active site. The external binding sites would then need to work together to accomplish the branching reaction. Binding sites that are located in close proximity, such as III + VI and IV + VII, cannot be simultaneously mutated or will result in an almost completely dead protein. The missing N-terminal in the truncated protein gears the enzyme's mode of binding from site III, which had the highest effect in the full-length protein, to site I, and leads to other sites such as site VI standing out more. Sites I, VI and, and VII are crucial for the activity of each enzyme as each contains a mutation that kills the protein.

Mutations in sites II, IV and V, though they reduce the activity greatly do not lead to a completely dead protein, and therefore are more likely to involve binding enzyme to its polymer substrate than directing oligonucleotides into the active site. Sites I, VI and VII are the sites that are most likely to involve directing transferred chains and acceptor chains into the active site.

Site III overall and only one mutation in site VI present an ambiguous result as it is crucial for the full-length protein but has less impact on the truncated protein. Mutations W478A, F477A and L512A show 52, 237 and 98% activities in the truncated version, while they show 5, 9, and 5% activities in the full-length. These are both sites in the bottom of the protein, looking at the exit channel of the active site. The exit channel is more sensitive to mutation in wt full-length versus N112BE. The full-length version has the far N-terminal domain, which might be assisted by this site somehow to accomplish activity. In the MycoBE, the far N-terminal is located away from the protein's N-terminal domain we see in EcBE structure, however, the same location does not have to be true for EcBE, in which site III could be the closest to the protein. We can only explain this when the full-length crystal structure of EcBE becomes available.

In the following chapter, we present the crystal structure of EcBE bound to a maltooctaose, showing other binding sites, some being the same. There, we will compare the protein with the one explained in this chapter and suggest other mutations for future work.

## REFERENCES

## REFERENCES

1. Baba, T., Kimura, K., Mizuno, K., Etoh, H., Ishida, Y., Shida, O., and Arai, Y. (1991) Sequence conservation of the catalytic regions of amylolytic enzymes in maize branching enzyme-I *Biochem. Biophys. Res. Commun.* **181**, 87-94.
2. Nakajima, R., Imanaka, T., and Aiba, S. (1986) Comparison of amino acid sequences of eleven different  $\alpha$ -amylases, *Appl. Microbiol. Biotechnol.* **23**, 355--360.
3. Borovsky, D., Smith, E. E., and Whelan, W. J. (1975) Purification and Properties of Potato 1,4- $\alpha$ -D-Glucan: 1,4- $\alpha$ -D-Glucan 6- $\alpha$ -(1,4- $\alpha$ -Glucano)-Transferase, *Eur. J. Biochem.* **59**, 615-625.
4. Boyer, C. D., and Preiss, J. (1978) Multiple Forms of (1,4)- $\alpha$ -D-Glucan, (1,4)- $\alpha$ -D-Glucan-6-Glycosyl Transferase From Developing *Zea mays* L. Kernels, *Carbohydrate Research* **61**, 321-334.
5. Brzozowski, A. M., and Davies, G. J. (1997) Structure of the *Aspergillus oryzae* Alpha-Amylase Complexed with the Inhibitor Acarbose at 2.0 Å Resolution, *Biochemistry* **36**, 10837-10845.
6. Jespersen, H. M., MacGregor, E. A., Henrissat, B., Sierks, M. R., and Svensson, B. (1993) Starch- and Glycogen-Debranching and Branching Enzymes: Prediction of Structural Features of the Catalytic ( $\alpha/\beta$ )<sub>8</sub>-Barrel Domain and Evolutionary Relationship to Other Amylolytic Enzymes, *Journal of Protein Chemistry* **12**, 791-805.
7. Jespersen, H. M., MacGregor, E. A., Sierks, M. R., and Svensson, B. (1991) Comparison of the domain-level organization of starch hydrolases and related enzymes, *Biochemistry J.* **280**, 51-55.
8. Svensson, B. (1994) Protein engineering in the  $\alpha$ -amylase family- catalytic mechanism, substrate specificity, and stability, *Plant Molecular Biology* **25**, 141-157.
9. Uitdehaag, J. C. M., van Alebeek, G. J., Van Der Veen, B., Dijkhuizen, L., and Dijkstra, B. W. (2000) Structures of Maltohexaose and Maltoheptaose Bound at the Donor Sites of Cyclodextrin Glycosyltransferase Give Insight into the Mechanisms of Transglycosylation Activity and Cyclodextrin Size Specificity, *Biochemistry* **39**, 7772-7780.
10. Feng, L. (2009) X-Ray Crystallographic Studies of Branching Enzyme/Polysaccharide Complex and RNA Polymerase III Transcription Factor TFIIIB Complex, *Dissertation*, p. 10.

11. Abad, M. C., Binderup, K., Rios-Steiner, J., Arni, R. K., Preiss, J., and Geiger, J. H. (2002) The X-ray crystallographic structure of *Escherichia coli* Branching Enzyme, *Journal of Biological Chemistry* 277, 42164-42170.
12. Binderup, K., Mikkelsen, R., and Preiss, J. (2000) Limited proteolysis of Branching Enzyme from *Escherichia coli*, *Archives of biochemistry and biophysics* 377, 366-371.
13. Kim, M. I., Kim, H. S., Jung, J., and Rhee, S. (2008) Crystal structures and mutagenesis of sucrose hydrolase from *Xanthomonas axonopodis* pv. glycines: insight into the exclusively hydrolytic amylosucrase fold, *Journal of molecular biology* 380, 636-647.
14. Binderup, K., Mikkelsen, R., and Preiss, J. (2002) Truncation of the amino terminus of Branching Enzyme changes its chain transfer pattern, *Archives of biochemistry and biophysics* 397, 279-285.
15. Matsumoto, A., Nakajima, T., and Matsuda, K. (1990) A Kinetic Study of the Interaction between Glycogen and Neurospora crassa Branching Enzyme, *Journal of Biochemistry* 107, 123-126.
16. Matsumoto, A., Nakajima, T., and Matsuda, K. (1990) Mode of Action of Glycogen Branching Enzyme from Neurospora crassa, *Journal of Biochemistry* 107, 118-122.
17. Hizukuri, S., Takeda, Y., Yasuda, M., and Suzuki, A. (1981) Multi-branched nature of amylose and the action of debranching enzymes. , *Carbohydrate Research* 94, 205-213.
18. Takeda, Y., Guan, H.-P., and Preiss, J. (1993) Branching of amylose by the Branching Isoenzymes of maize endosperm, *Carbohydr. Res.* 240, 253-263.
19. Krisman, C. R. (1962) A Method From the Instituto Facultad de Ciencias of for the Calorimetric Estimation Glycogen with Iodine, *Anal. Biochem.* 4, 17-23.
20. Guan, H. P., and Preiss, J. (1993) Differentiation of the Properties of the Branching Isozymes from Maize (*Zea mays*), *Plant Physiology* 102, 1269-1273.
21. Binderup, K., and Preiss, J. (1998) Glutamate-459 Is Important for *Escherichia coli* Branching Enzyme Activity, *Biochemistry* 37, 9033-9037.
22. Kuriki, T., Stewart, D. C., and Preiss, J. (1997) Construction of Chimeric Enzymes out of Maize Endosperm Branching Enzymes I and II: Activity and Properties, *Journal of Biological Chemistry* 272, 28999-29004.

23. Pal, K., Kumar, S., Sharma, S., Garg, S. K., Alam, M. S., Xu, H. E., Agrawal, P., and Swaminathan, K. (2010) Crystal structure of full-length *Mycobacterium tuberculosis* H37Rv Glycogen Branching Enzyme: insights of N-terminal beta-sandwich in substrate specificity and enzymatic activity, *The Journal of biological chemistry* 285, 20897-20903.

## CHAPTER 3: X-RAY CRYSTALLOGRAPHIC STUDIES OF *E. COLI* BE IN COMPLEX WITH OCTAOSE

### 3.1 Research Objective

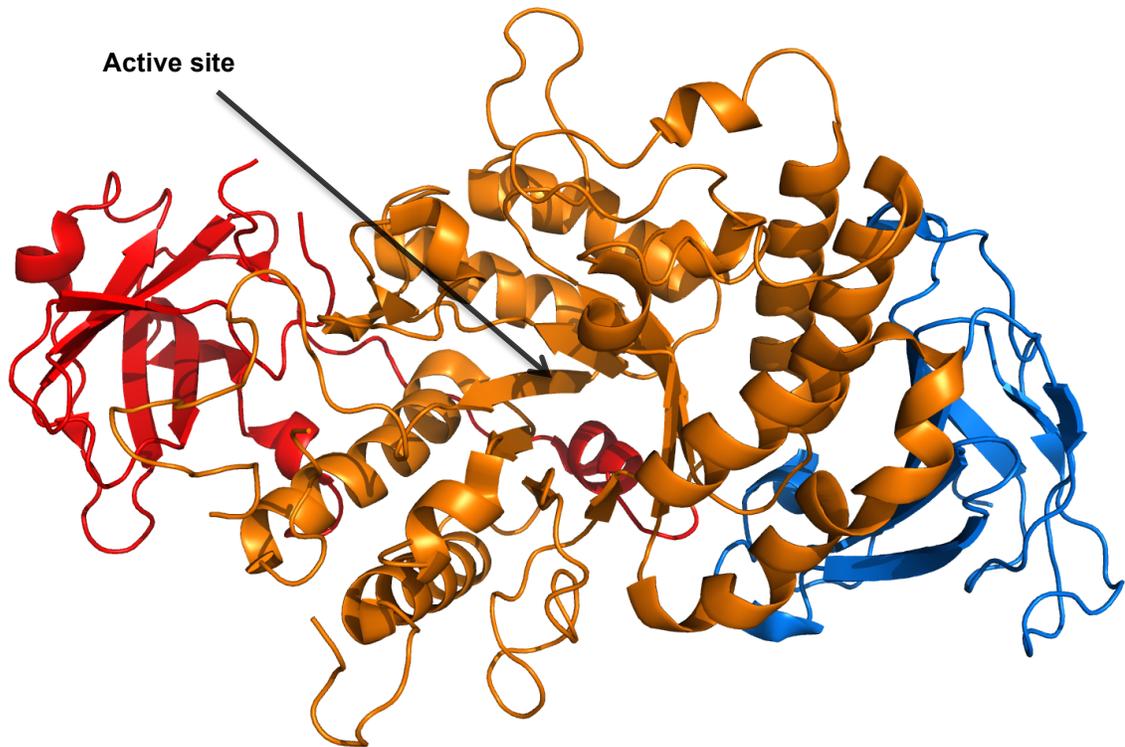
Branching enzyme plays a dual simultaneous role of cleaving  $\alpha$ -1,4-bonds and forming  $\alpha$ -1,6-branches in oligosaccharides. Amino acid sequence alignment and site-directed mutagenesis studies showed that branching enzyme (BE) is a member of the amylolytic family of enzymes.<sup>1-3</sup> Although the central catalytic domain is highly conserved among this family, the amino- and carboxy- termini are very different. In an effort to understand the role of each of the terminal domains, Binderup *et al.* undertook proteolysis experiments specifically on *E. coli* branching enzyme (*EcBE*).<sup>4</sup> Limited proteolysis of the wild type *EcBE* by proteinase K gave a robust polypeptide product of 70-KDa that only decomposed further in the presence of excess digestive enzyme. They showed that this truncated *EcBE* with 112-amino-acid deletion at the N-terminus resulted in a protein that retained 60% of enzyme activity and was 3 to 4 fold less catalytically efficient than the full-length protein.<sup>4</sup> Attempts at *EcBE* crystallization date much older than the year 2002; however, all trials failed until Abad and coworkers tried crystallizing the truncated version (**Figure 3.1**).<sup>5</sup>

The first crystal structure of any branching enzyme was that of *Escherichia coli* BE determined in 2002.<sup>5</sup> The structure confirms the predicted similarity to the  $\alpha$ -amylase family of enzymes despite the low sequence identity.<sup>4, 5</sup> The overall

structure is elliptical in shape. The COOH-terminal domain is arranged in seven  $\beta$ -strands. The active site is formed into a groove-like domain with eight alpha helices making the boundaries of the groove and eight beta-sheets creating the bottom of the groove (**Figure 3.1**). Seven  $\beta$ -strands cross the NH<sub>2</sub>-terminal domain, in agreement with previous sequence alignment studies.<sup>6-9</sup> Four conserved regions among the members of the  $\alpha$ -amylase family are also conserved in *EcBE*; these regions contain the catalytic residues Y300, D335, H340, R403, D405, E458, H525, and D526 according to *EcBE* numbering revealed based on biochemical data and on the crystal structures of unbound and in complex with substrates of cyclodextrin glucanotransferase (CGT) and  $\alpha$ -amylase.<sup>10, 11</sup>

*BE* binds and catalyzes natural linear polysaccharides, which are not rigid structures but flexible and long. The substrates then require support for enzymatic activity to take place; therefore, *BE* should have several binding sites to provide this groundwork. Feng and coworkers soaked crystals of *EcBE*, reproduced after Abad's crystals, in a number of oligosaccharides ranging from maltose to maltoheptaose in length and also in different cyclodextrins.<sup>12</sup> These studies demonstrated the existence of seven binding sites in all four asymmetric units combined, one of the binding sites being exclusive to cyclodextrins. The restrictive binding of cyclic oligosaccharides in one site becomes more realistic when we consider the structure of the natural linear substrates of *EcBE* forming helical turns every six glucose units.<sup>12</sup> Although hypotheses can be made on the mode of binding of *EcBE* to its substrate after the unfolding of seven binding

sites, the mystery remains in the way the enzyme interacts with the polysaccharide when binding occurs at the active site. With a seven-unit oligomer, binding was only visible on the surface of the enzyme, which got the Geiger laboratory very curious to see binding in the catalytic domain in order to better explain the enzymatic mechanism and compare the results with other members of the same enzyme family for which structures of holo-enzymes with substrate binding in the central domain have been solved. We grew crystals of EcBE and soaked them in maltooctaose (M8) and in dodecaose (M12) carbohydrates. While the data on the enzyme in complex with M12 was very weak and will not be displayed here, the present structure of *E. coli* branching enzyme in complex with M8 confirms the existence of multiple external binding sites, namely ten, five of which are novel in comparison to EcBE – M7 and the other five are common. We also present possible explanation for the denied binding in EcBE – M8 at two sites that previously affirmed significance to the protein's activity.



**Figure 3.1: Crystal structure of *EcBE*.**<sup>5</sup> The active site is displayed in orange, the N-terminal in red and the C-terminal in blue.

## 3.2 The Three Dimensional Structure of EcBE Bound to Maltotetraose

### 3.2.1 Structure determination

Although the crystal structure of EcBE was previously reported by the Geiger group and belonged to space group  $P2_1$ ,<sup>5</sup> we grew our 112 residue-amino-terminal truncated protein starting from a different vector as explained in Materials and Methods – chapter 6, and therefore obtained a different form of the crystals. **Table 3.1** outlines the x-ray diffraction data collected from a single crystal at the Advanced Photon Source (Chicago, IL), LS-CAT 21-ID-D beamline. The diffraction images were processed using HKL2000.<sup>13</sup> The data was further refined using CCP4 (**Table 3.1**).<sup>14, 15</sup> The structure was solved by molecular replacement using the EcBE structure (PDB ID: 1M7X) as a search model and data was refined to 3.0 Å, and a completeness of 96.3%, with an  $R_{\text{work}}$  of 0.2051 and  $R_{\text{free}}$  of 0.3188. The crystal structure belongs to space group  $P 3_2 2 1$  and contains 4 monomers in each asymmetric unit as its precedents, numbered A, B, C, and D similarly to holo and apo-EcBE.<sup>5, 12</sup> The present form of the protein (chain A) consists of 20009 non-hydrogen atoms, 902 oligosaccharide atoms and 579 water molecules. The enzyme binds to 14 different glucose chains, each consisting of a distinct number of glucose units ranging from one to eight, forming together 8 disparate binding sites (**Figure 3.2**). The overall structure of the protein is formed, as previously shown, of an NH<sub>2</sub>-terminal domain (residues 117 – 240), a central catalytic domain (residues 241 – 612), and a COOH-terminal (residues 613 – 728).

<i>Data collection</i>	
Wavelength (Å)	0.978
Unit-cell parameters (Å)	a = 146.489, b = 146.489, c = 294.920
Unit-cell parameters (°)	$\alpha = 90, \beta = 90, \gamma = 120$
Space group (Å)	P3 <sub>2</sub> 21
Resolution range (Å)	50.00-2.80 (2.85-2.80)
Completeness (%)	88.2 (40.0)
$R_{\text{merge}} (I)^a$	0.14 (0.90)
$\langle I \rangle / \langle \sigma_I \rangle$	22.9 (0.94)
<i>Refinement</i>	
Resolution range (Å)	38.85 – 3.00
$R$ -factor <sup>b</sup>	0.2051 (0.33)
$R_{\text{free}}^c$	0.3188 (0.40)
Completeness (%)	96.33
Root mean square deviation bonds	0.0121
Root mean square deviation angles	1.6537
Protein non-hydrogen (NH) atoms	18,880
Oligosaccharides NH atoms	664
Water molecules	473

<sup>a</sup>  $R_{\text{merge}} = \sum \sum_i |I_i - \langle I \rangle| / \sum |I|$ , where  $I_i$  is an individual intensity measurement and  $\langle I \rangle$  is the average intensity for this reflection, with summation over all data.

<sup>b</sup>  $R$ -factor:  $\sum ||F_o| - |F_c|| / \sum |F_o|$ .

<sup>c</sup>  $R_{\text{free}}$  :  $R$ -factor of 10% reflections removed before refinement.

**Table 3.1: Data collection and refinement statistics**

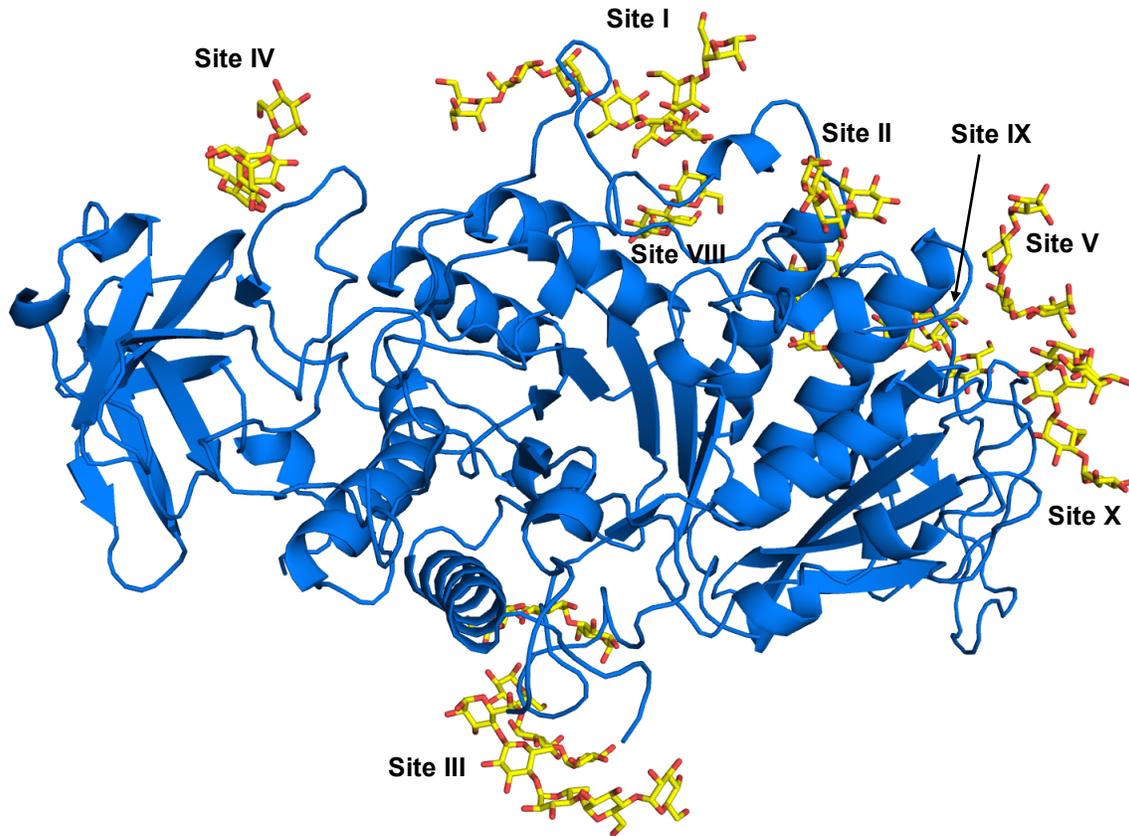
### 3.2.2 Comparing the molecules in the asymmetric unit

The crystal structure of *EcBE* reveals four molecules per asymmetric unit. Each monomer elongates from residue 117 to 728, with the last one and two amino acids disordered in each of molecules B (Mol. B) and D (Mol. D), respectively. The two disordered regions seen in *EcBE* – M7, residues 361 – 373 and 414 – 429, are not well ordered here either and do not show the following residues: 364 – 369 and 415 – 417 (Mol. A), 361 – 369 and 422 – 425 (Mol. B), 365 – 371 and 419 – 425 (Mol. C), and 361-362, 365-370, and 417-419 (Mol. D). One other region consisting of residues 211 to 215, also previously disordered in the M7 complex, is now ordered in all molecules except in D, density for residues 212 and 213 is very weak.

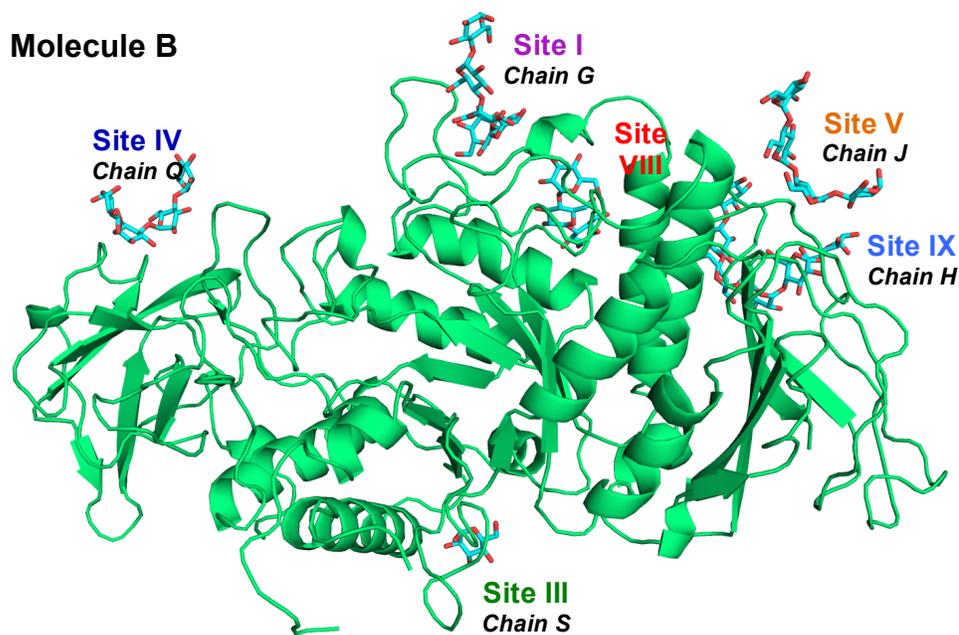
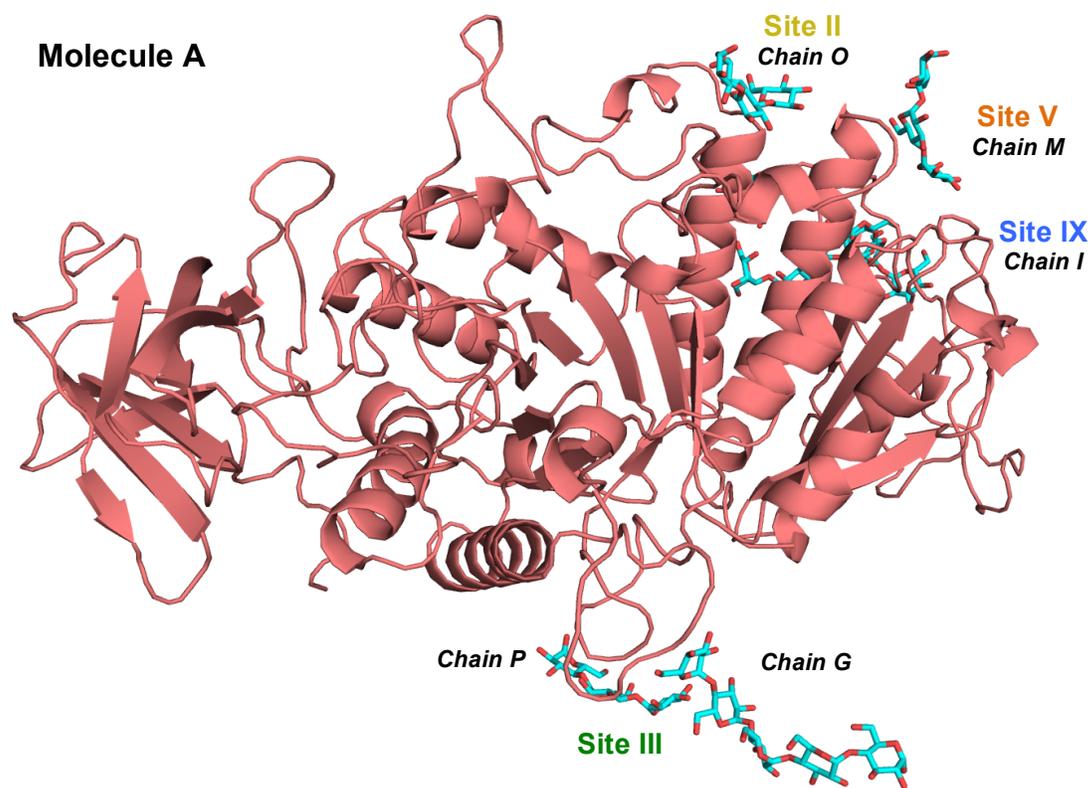
Binding sites are uniquely filled in each molecule (**Figure 3.2**). **Table 3.2** classifies the sugar chains according to the binding sites they are bound to, shows the number of glucose units seen per site in each molecule, and lists the residues involved in the interactions in each binding site. Site I is filled in molecules B and C with five and eight glucose units respectively, and the oligosaccharides overlap quite well and adopt the same orientation. The chain that fills this site in Mol. B, chain G, is actually sandwiched between molecules A, B and C in the protein's asymmetric unit (**Figure 3.5**). When the four molecules A through D are overlaid, chain G bound to Mol. B overlaps with chain E – Mol. C in site I and with superimposing reducing ends (**Figure 3.3**). On the other hand, the same chain G bound to A and C make part of binding site III but adopt different conformations in these molecules (**Figure 3.3**).

Chain P, which binds to molecules A and B, present a similar case. The interactions this chain makes and location it adopts when A and B are overlaid are very different and represent two different binding sites, III and VIII. Since all the other chains besides G and P bind to individual molecules in the asymmetric unit, they make distinct binding sites bound to distinct oligosaccharides (**Figure 3.3**). Binding site IV is only filled in Mol. B and only 3 out of 4 visible glucose units make contact with the protein. Also, binding site X is only occupied in Mol. C and while density for five Glc units is visible; only four units interact with the protein with the non-reducing end sugar not binding. The rest of the sites described here all feature overlapping chains bound to more than a single molecule. Site II occupied in molecules A and C by chains O (3 Glc units) and N (2 units) respectively; the shorter chain reducing-end overlaps with the non-reducing end of the longer one. Binding site III is overwhelmed with sugars, in addition to chain G that binds to molecules A and C, and to chain P binding to Mol. A, chain L of Mol. C, formed of 3 glucose units, also binds in this site, and chain S of Mol. B overlaps the center glucose of chain L. Site V is occupied in 2 molecules, A and B, with chains M (3 units) and J (4 units); the non-reducing end of longer chain overlaps with the reducing end of the other similar to site II. Binding site IX, occupied in Mol. A, B and D by chains I, H and K respectively, resembles sites II and V in that binding oligosaccharides do not all take a single direction. Chain I showing 7 glucose units, runs in opposite direction to the other two and its Glc6 and Glc8 do not make any interaction with the protein's residues. Chains H's and K's reducing-end sugars start at Glc6 and Glc4 and overlap through Glc2 and

Glc1 of chain I units respectively. The opposite direction of sugars bound within each site must imply some differences in roles and will be discussed later.



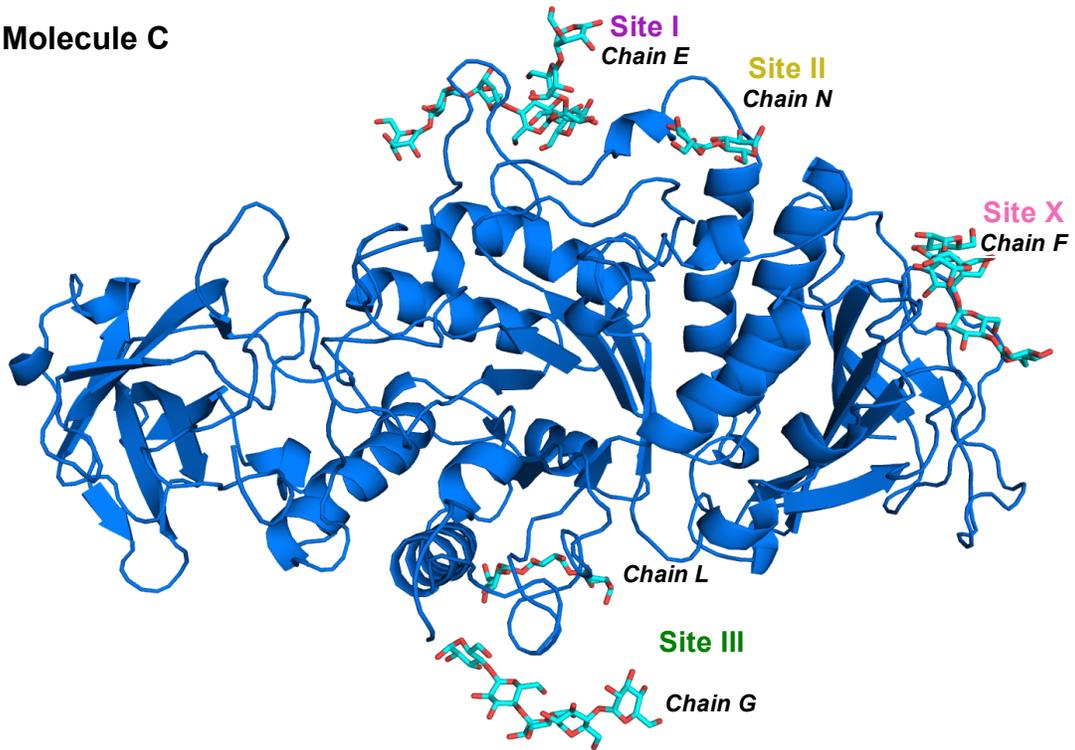
**Figure 3.2: Composite representation of EcBE – M8 complex showing the relative positions of the nine binding sites.** All the ligands were overlaid onto Mol. C (blue cartoon). The oligosaccharides bound to EcBE are colored by type, C in yellow and O in red.



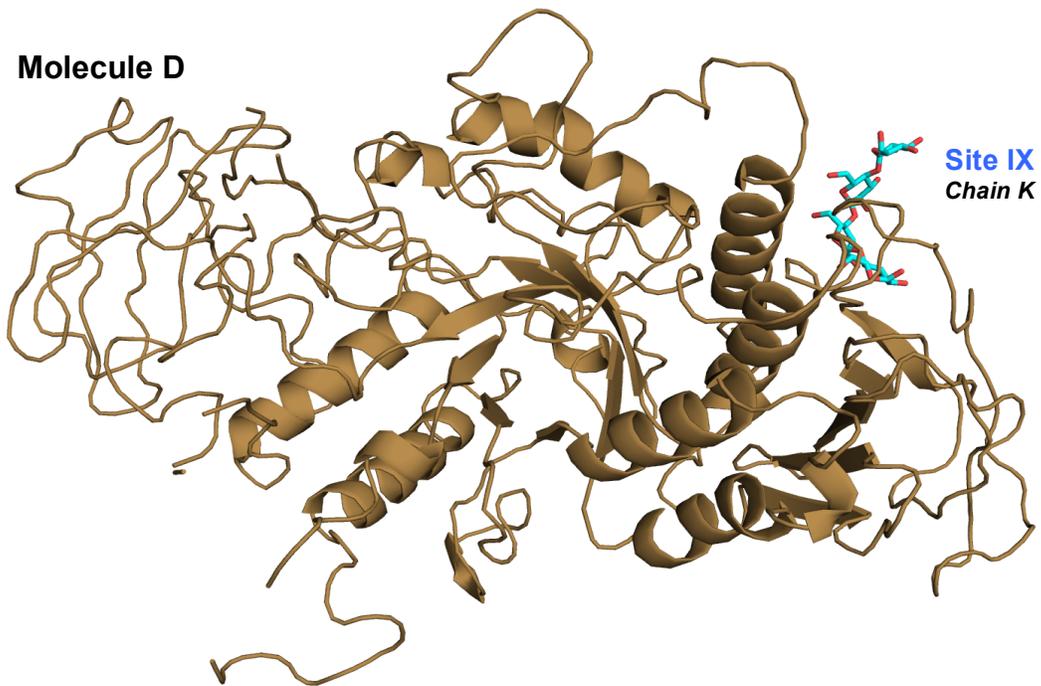
**Figure 3.3: Binding sites in each of the four molecules per asymmetric unit.** Oligosaccharides are colored by type, with C in cyan and O in red, and the cartoon of protein molecules is colored by chain, A in deep salmon, B in bright green, C in blue and D in sand. Oligosaccharide chains bound are mentioned for each binding site.

Figure 3.3 (cont'd)

**Molecule C**



**Molecule D**



Binding Site	Mol.	Glc chain	Glc units no.	Residues involved in interactions
I	C B	E G	8 5	R255, H256, T257, D258*, N259*, N260, F261*, W262*, L263*, S264*, E267*, Q271*, R576, S583, L584 (Mol. B only), D585, W586, H587, E590 <i>* Interaction found only in Mol. C – chain E</i>
II	A C	O N	3 2	L536, D537, A543, K546, F547, G575, R576, L588, W595, H596
III	A A B C C	G P S G L	5 3 1 5 3	M472, G473 D471, M472 <i>All residues with ** interact in chain S</i> E437, R440, D471 F239**, E437, R440, R444, E448, S467**, R468, P469**, Q470, D471, G476**, F477**, W478**, N518**
IV	B	Q	4	W159, L187, K189, L201, Q211, E215 and N198 of symmetrical molecule C
V	A B	M J	3 4	D542, W544, Q545, P659, P661, S689, P717
VIII	B	P	3	R254, Y275, W278, W586, L589, Q600
IX	A B D	I H K	8 5 4	(R601, S683, N693) <sup>+</sup> , D605, E677, I678, L679 <sup>++</sup> , N680, H685, Y686, N690, G692, G694, G695 <i><sup>+</sup> Interaction found only in Mol. B – chain H</i> <i><sup>++</sup> Interaction found only in Mol. A – chain I</i>
X	C	F	5	D496, P497, V498, Y499, Q501, E635, T658, V660, R662

**Table 3.2: Summary of binding sites, molecule occupied, glucose chain name and length, and the residues involved in the interactions.** Residues listed in the last column interact with all molecules occupied in the corresponding site unless otherwise noted, except for site III, where each chain interacts with different residues, as all chains do not overlap. In this site, residues interacting are listed next to each chain.

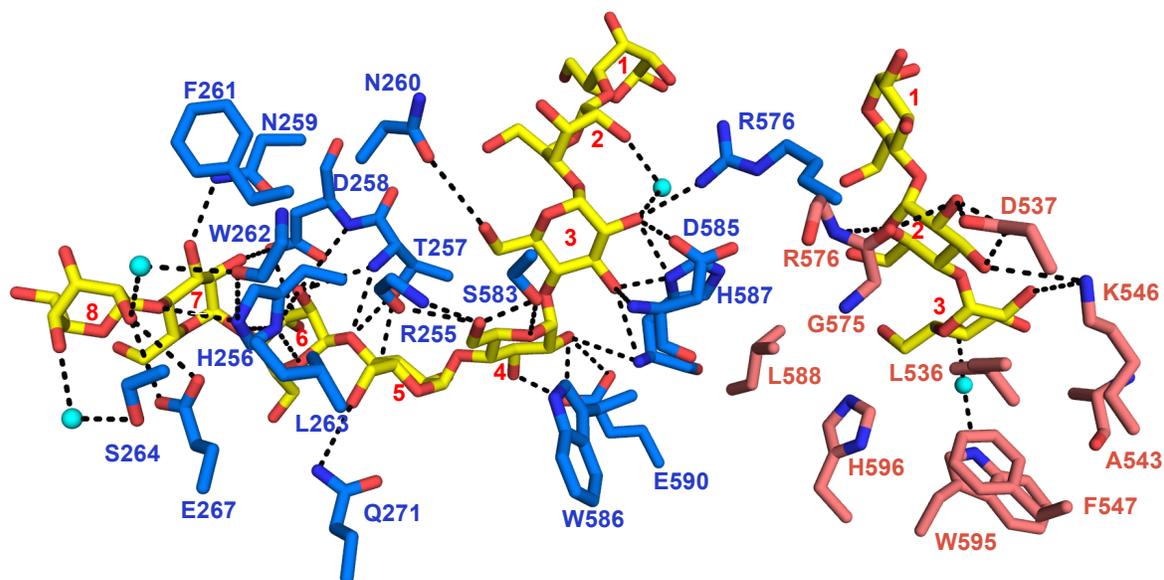
### 3.2.3 Detailed interactions in each binding site

#### 3.2.3.1 Binding sites I and II

Site I, occupied in both molecules B and C, binds one of the two maltooctaose showing all eight glucose units in the EcBE – M8 complex. **Figure 3.4** shows the detailed interactions between the oligosaccharide and the residues in blue (Mol. C). The first glucose on the reducing end numbered 1 (Glc1) does not make any direct interaction with the protein. O2 of Glc2 makes a hydrogen bond with a water molecule that also connects to O2 of Glc3 by hydrogen bonding. Arg576, Asp585, and His587 side chains also interact with O2 of Glc3 through hydrogen binding. The main chain nitrogens of D585 and His587 make H-bonds with O3 of Glc3, and O3 of Glc3 and O2 of Glc4, respectively. Glc3 also makes another hydrogen bond by means of its O6 to the side chain oxygen of Asn260. The fourth glucose unit makes 3 sets of hydrogen bonds: both side chain oxygens of Glu590 make an H-bond each to O2 and one of the oxygens binds to O3, Ser583 hydrogen-binds to O5 and O6 and the main chain nitrogen and carbonyl of Arg255 bind to O6. Glucose 5 is also held by several hydrogen bonds: one between O6 and Gln271 side chain nitrogen, two others made by Thr257 side chain to each of O3 and O4, and one between O4 and the main chain nitrogen of Thr257. In Glc6, O2 binds to the main chain nitrogens of Asp258 and Thr257 and O3 to both oxygens side chain of Asp258, which also makes a hydrogen bond to O2 of Glc7. The net of hydrogen bonds displayed in **Figure 3.4** also includes hydrogen bonds interactions between the side chain N3 and NH1 of His256 to

Glc6 (O2, O4 and O5) and Glc7 (O4), respectively. Oxygen 3 of Glc7 also binds to the side chain nitrogen of Asn259. In addition, Glc8 makes 2 hydrogen bonds to the side chain of Glu267. Two other residues, Trp262 and Ser264, connect between O2 and O4 of Glc8 and another pair of residues, His256 and Glu267, through water molecules 269 and 268, respectively. Despite this wealth of hydrogen bonds seen in this site, the hydrophobic interactions are very shy restricting to only two residues, Leu263 and Phe261, interacting with Glc6 and Glc7. Moreover, the side chain of Glu231 belonging to B molecule in the symmetrical unit approaches this site and makes 2 hydrogen bonds with O2 and O3 of Glc5. Finally, one additional interaction, not present in site I Mol. C but seen in Mol. B, O3 of Glc3 makes a hydrogen bond with the main chain carbonyl of Leu584 (**Table 3.2**).

The 3 glucose units seen in the overall structure in **Figure 3.2**, site II, bind to Mol. A and form fewer interactions compared to site I. Glc1 on the reducing end makes no interaction **Figure 3.4**, while 3 oxygens in Glc2 interact: O2 binds to the main chain carbonyl of Gly575 and to both side chain oxygens of Asp537, O3 to the side chains of Asp537 and Lys546. The latter residue also binds to O2 of Glc3, which makes another interaction to Trp595 through water molecule 537. Hydrophobic interactions are also made in this site: three residues, Ala543, Leu536, and Phe547, surround the non-reducing end Glc3 and one Leucine, L588, is located close to Glc2.



**Figure 3.4: Binding sites I and II detailed interactions.** The protein residues are colored by atom type, with carbons in blue for Mol. C and deep salmon for Mol. B (also labeled according to colors), oxygens in red and nitrogen in dark blue. The glucose units' backbone is colored in yellow and oxygens in red and. The longest oligosaccharide present in each site (chain E of Mol. C and chain O of Mol. B) is depicted here and its units are numbered in red starting at the reducing end sugar. Cyan spheres represent water molecules, and black dotted lines, H-bonds.

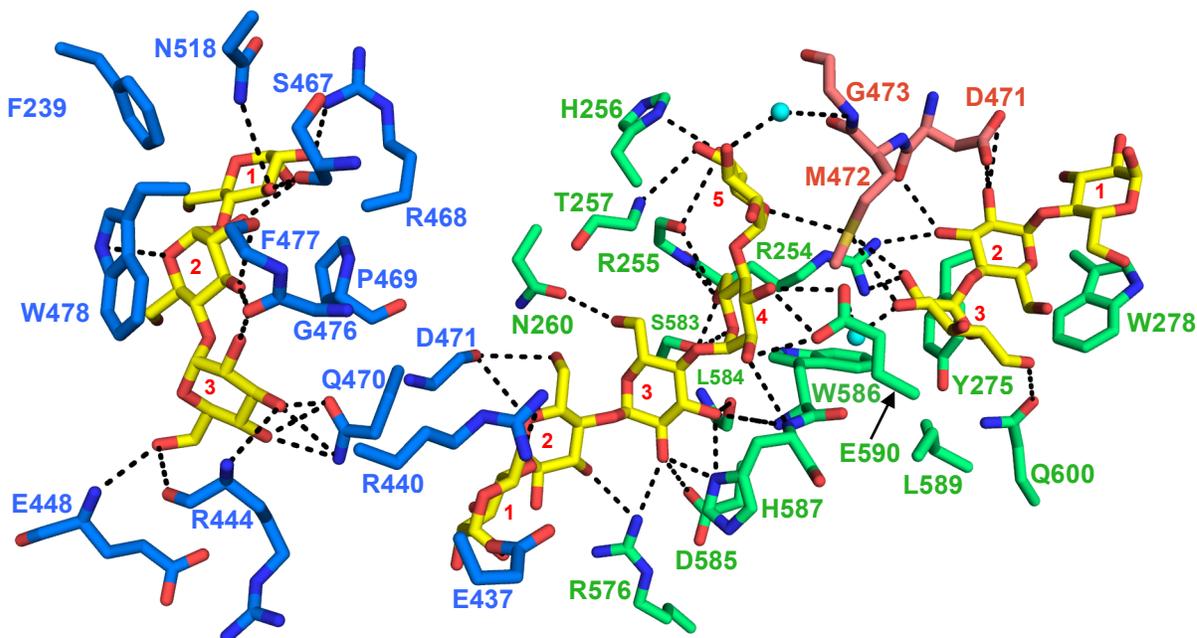
### 3.2.3.2 Binding sites III and VIII

Three distinct chains occupy binding site III in quite different locations but all wrap around one loop as shown in the overall structure (**Figure 3.2**). Chain L, shown in Mol. C **Figure 3.3**, is formed of three glucose units that make several hydrogen bonds with the protein residues. In Glc1 on the reducing end, O2 binds to the side chain of Arg468 and O3 to the side chain nitrogen of Asn518 and the main chain carbonyl of S467 (**Figure 3.5**). Glc2 shows more interactions than Glc1, as O2 hydrogen binds to main chain carbonyl of S467 and Phe477, and O3 to those of Phe477 and Gly476, while O5 binds to the side chain nitrogen of W478, which also makes sugar-amino acid stacking interaction with Glc2, and so does Pro469. In addition, Phe239 makes hydrophobic interactions with Glc1 and Glc2. As for GLc3, we see another set of hydrogen bonds when O2 binds to the main chain carbonyl of G476, O3 to the nitrogen and carbonyl main chain of Arg444 and to the amide oxygen and nitrogen side chain of Gln470; the latter 2 atoms also bind to O4. Oxygen 6 also makes 2 H-bonds, with main chains carbonyl and nitrogen of Arg444 and Glu448, respectively.

Another chain, G, is sandwiched between molecules A, B, and C, and binds to all three molecules (**Figure 3.5**). While the interactions it makes with Mol. A and Mol. C are part of site III, its interactions in Mol. B overlap with chain E of Mol. C and are therefore part of site I (which are detailed in section 3.2.3.1). Glucoses 1 and 2 on the reducing end interact with Mol. C; interactions that are absent in either chain in site I leaving those two glucose units unbound there. Oxygen 1 of Glc1 binds to the side chain of Glu437 and O6 makes 2 H-bonds with the side

chain of Arg440. In Glc2, O5 and O6 bind to the main chain carbonyl of Asp471. Chain G, Glc2 through Glc5, then makes several interactions with Mol. B overlapping with site I. Glc 5 also interacts with Mol. A, making a hydrogen bond with Met472 side chain, and another one with Gly473 main chain carbonyl through a water molecule.

Moreover, molecule A interacts with chain P, the third chain on the right in **Figure 3.5** as part of binding site III. Both carboxylate oxygens of Asp471 bind to O2 of Glc2 and one of them to O3 of the same glucose. Met472 side chain also binds to O2 of Glc3. The rest of the interactions with chain P listed hereafter are made by residues of Mol. B and represent the only interactions seen in binding site VIII, which is occupied by chain P only in molecule B. Trp478 stacks onto Glc1 locking the reducing end sugar in place. Tyr272 ring approaches Glc2 without making hydrogen bonding. The side chain terminal nitrogens of Arg254 bind to O3 of Glc2 and O2 of Glc3, making together 2 H-bonds. Gln600 side chain carbonyl binds to O6 of Glc3 terminating the hydrophilic interactions in this site. Two other residues, Trp586 and Leu589 make hydrophobic contacts with Glc3.



**Figure 3.5: Detailed interactions in binding sites III and VIII.** Oligosaccharide chains, L (Mol. C), G (Mol. A, B, and C), and P (Mol. A and B), are shown from left to right, binding to molecules described in parentheses above. Molecules A, B and C are not overlaid here but shown as they appear in one asymmetric unit. When all four molecules of the asymmetric unit are overlaid, oligosaccharide P (most right) bound to Mol. B (site VIII) is located differently than in Mol. A. All the rest of the binding in this figure is part of site III. Coloring is as in figure 3.4 with the additions of Mol. B carbons shown in green and sulfur atoms in dark yellow.

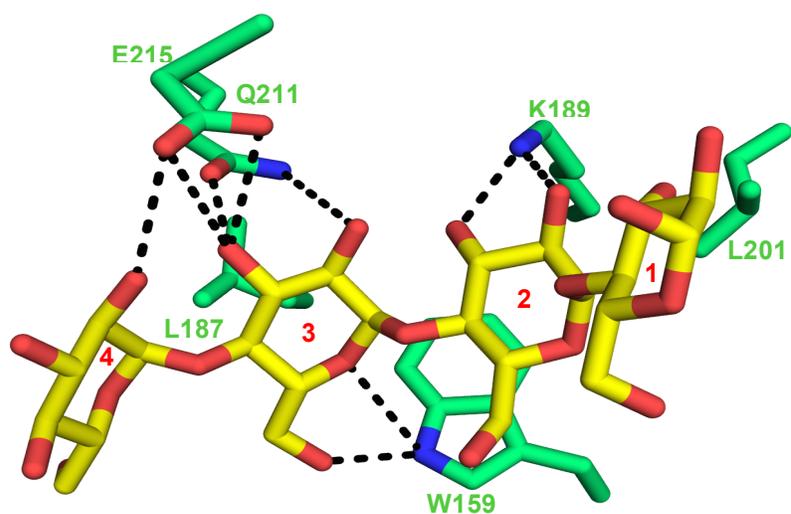
### 3.2.3.3 Binding sites IV and IX

Binding site IV is occupied only in Mol. B with chain Q formed of four glucose units, all interacting with the protein's residues (**Figure 3.6**). Glc1, on the reducing end, makes hydrophobic interactions with Leu201 in addition to a hydrogen bond made with the side chain of a residue in the symmetrical molecule C, Asn198, through a water molecule. A sugar-protein stacking

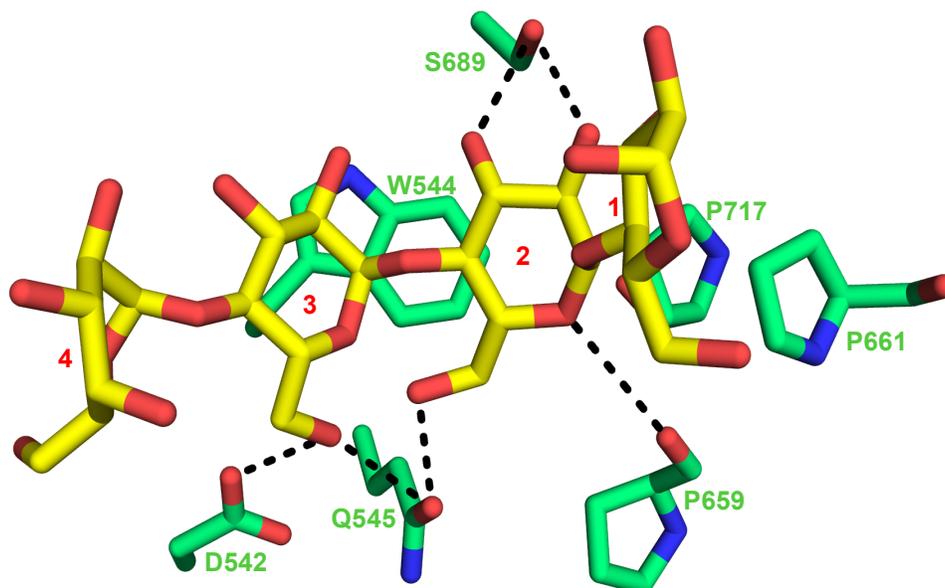
interaction is seen between Glc2 and Trp159, which also makes 2 H-bonds to O5 and O6 of Glc3. Lys189 side chain also makes 2 H-bonds with O2 and O3 of Glc2. Glu215 side chain interacts with O2 and O3 of Glc3 and with O2 of Glc4, which also hydrogen-binds to the side chain amide nitrogen and carbonyl of Gln211, respectively.

#### **3.2.3.4 Binding site V**

Chain J, represented in **Figure 3.7**, binding to site V overlaps with chain M of molecule A in opposite directions. The reducing end sugar of M superimposes with the non-reducing end of J and extends to GLc2 of chain J. Proline 661 interacts with Glc1, and Pro717 with Glc2. The side chain of Ser689 makes hydrogen bonds with O2 and O3 of Glc2. The main chain carbonyl of Pro659 also binds to O5 of Glc2, while O6 binds to the side chain oxygens of Gln545 and Asp542. A tryptophan, W544, stacks against Glc2 and Glc3 in this site holding the sugar tightly.



**Figure 3.6: Detailed interactions in binding sites IV.** Oligosaccharide is chain Q binding to Mol. B (green). Coloring is same as in figure 3.5.



**Figure 3.7: Detailed interactions in binding site V.** Sugar chain J of molecule B is shown; chain M of molecule A runs in opposite direction with Glc1 starting at Glc4 of J, and ends at Glc2 of J. Coloring is as in figure 3.6. There are four visible residues but Glc4 does not interact with the protein.

### 3.2.3.5 Binding site IX

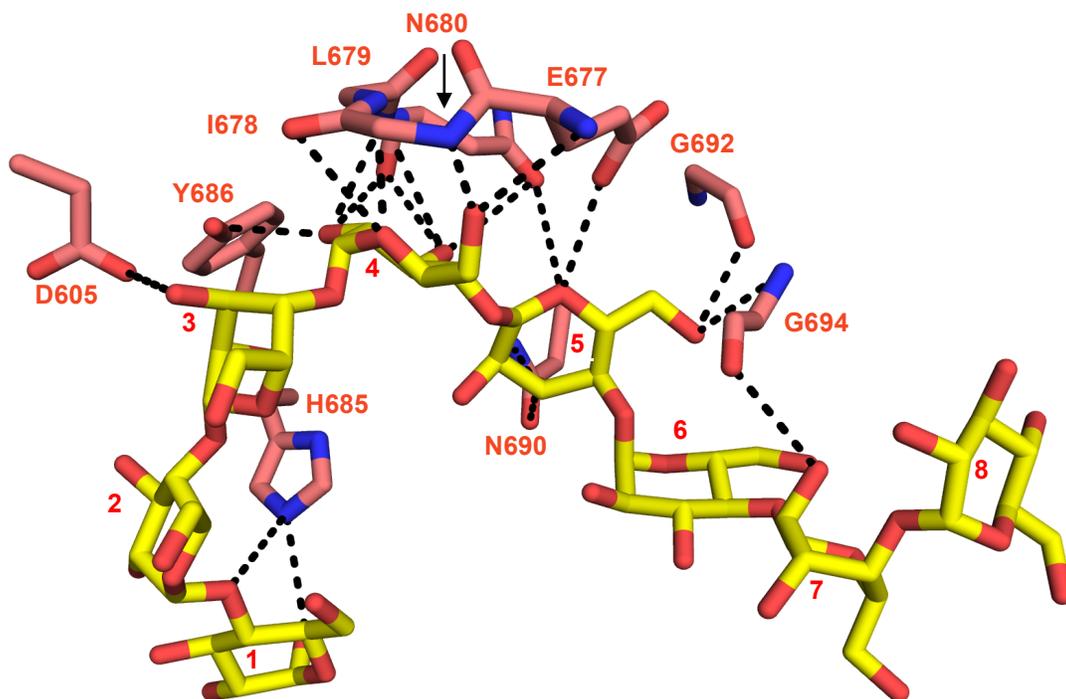
Another maltooctaose, chain I, spans this site in molecule A; however despite the visible full density for M8, Glc2, Glc6 and Glc8 do not interact with the protein's residues. The oligosaccharide overlaps with chains H of Mol. B and K of Mol. D, both of which bind in the same direction, opposite to chain I, as follows: chain H – Glc1 on the reducing end overlaps with Glc6 of I and its Glc5 on the non-reducing end, with Glc2 of I, while in the case of K, formed of 4 glucoses total, Glc1 overlaps with Glc4 of I and vice versa. The interactions described hereafter are those existent between Mol. A and chain I. The only type of interactions made in this site is hydrogen bonding (**Figure 3.8**). In Glc1, on the reducing end, O1 and O4 bind to the side chain of His685. In Glc3, a single H-bond is seen between O3 and D605. The following 2 glucoses are overwhelmed with interactions. O2 of Glc4 makes 3 H-bonds: one with the side chain of Trp686, another one with the main chain carbonyl of Asn680, which also binds to O3 of Glc4 and O5 of Glc5, and a third one with Asn680 main chain nitrogen, which also binds to O3 of Glc4. The side chain oxygen of Asn260 also binds to O3 of Glc4. Oxygen 5 of Glc4 interacts with the main chain carbonyl and nitrogen of Ile678 and Leu679, respectively, while O6 binds to both main chain nitrogens of Ile678 and Glu677. In Glc5, O5 binds to the side chain oxygen of Glu677, O3 to both oxygen and nitrogen of Asn690 side chain, and O6 to the main chains carbonyl and nitrogen of Gly692 and Gly694, respectively. Glc7 shows a single H-bond between O2 and the main chain carbonyl of G694.

This site presents a few other interactions in chain H of Mol. B not seen with the oligosaccharide occupying Mol. A. Arg601 side chain binds to O6 of Glc5, which overlaps with Glc2 in chain I. This represents the only interaction between the protein and a glucose unit not bound in chain I, while the following interactions listed here show more residues interacting in this site in the same location as previously explained. For instance, Ser683 side chain binds to O6 of Glc2, which overlaps with Glc5 in chain I and O2 of Glc2 interacts with the main chain nitrogen and oxygen of Asn693; Glc2 overlaps with Glc5 of chain I.

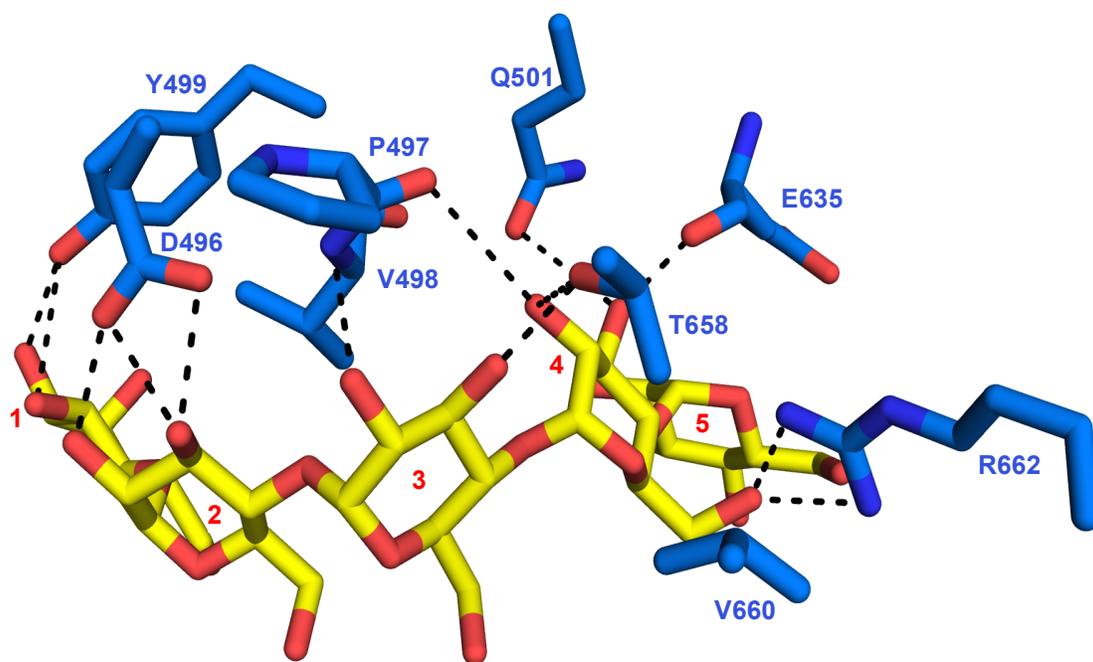
#### **3.2.3.6 Binding site X**

This site is occupied only in Mol. C with a 5 glucose-unit carbohydrate, of which the first 4 show interactions to the protein by means of hydrogen bonding mostly (**Figure 3.9**). O2 and O3 of Glc1 bind to the side chain of Tyr499. The side chain of Asp496 makes 2 H-bonds with O3 and a single one with O2 of Glc2. Similar to the 2 precedent glucoses, in Glc3, O2 and O3 are the only oxygens interacting; they bind to the main chain nitrogen of Val498 and the side chain of Thr658, respectively. O2 of Glc4 binds to the side chain of Thr658 and the main chain carbonyl of Pro497, and O3, to the side chain carbonyl of Gln501 and the main chain carbonyl of Glu635. Also, O6 of Glc4 makes 2 H-bonds to the side chain of R662 and Val660 interacts hydrophobically with Glc4. This concludes all binding sites interactions in the four molecules per asymmetric unit of EcBE – M8 complex. In the following paragraphs, we will compare the

common binding sites to previously discovered sites in structure EcBE – M7 and attempt explanation for lack of binding in sites VI and VII in the present structure.



**Figure 3.8: Detailed interactions in binding site IX.** Oligosaccharide represented is chain I of Mol. A, which overlaps with chains H of Mol. B and K of Mol. D. The latter 2 chains bind in opposite directions to the one depicted in this figure, as follows: chain H – Glc1 on the reducing end overlaps with Glc6 of I and its Glc5 on the non-reducing end, with Glc2 of I, while in the case of K, formed of 4 glucoses total, Glc1 overlaps with Glc4 of I and vice versa. Coloring is as in figure 3.5.



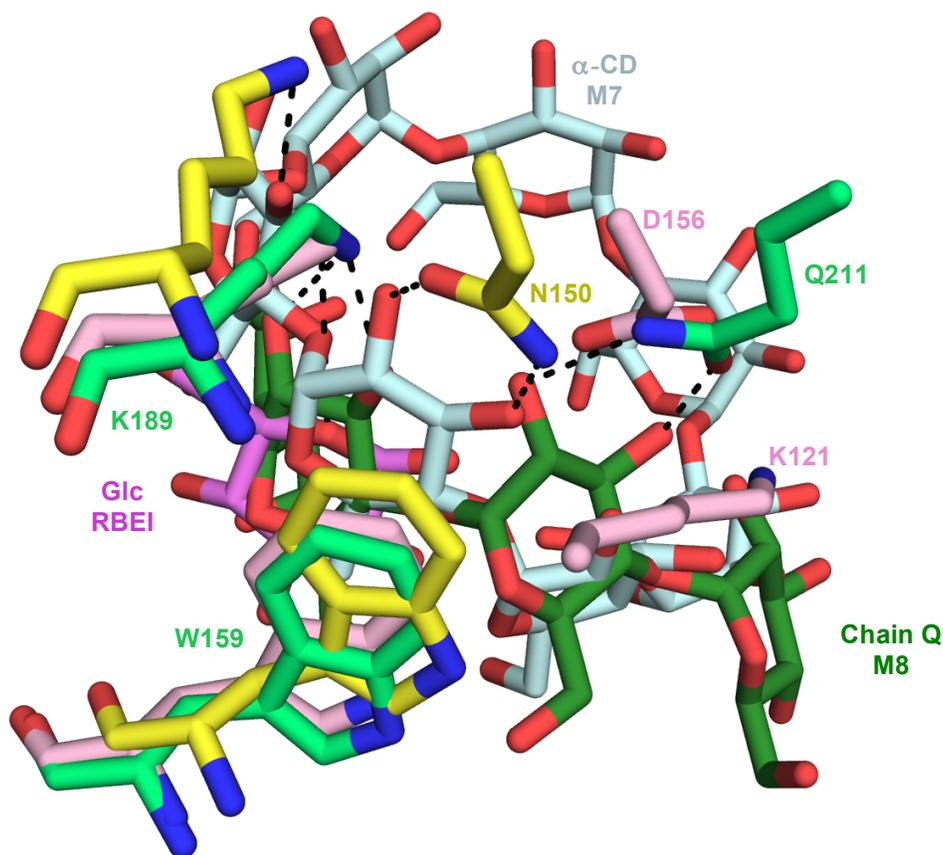
**Figure 3.9: Detailed interactions in binding site X.** Oligosaccharide formed of 5 glucose units occupy this site only in Mol. C. Coloring is similar to figure 3.5.

### 3.3 Structural Comparison

#### 3.3.1 CBM48 domain

According to its amino acid sequence, EcBE is classified as a member of the GH13 family according to the carbohydrate active enzymes (CAZy) and contains, in the N-terminal region, a carbohydrate-binding module, known as CBM48.<sup>16</sup> The crystal structure of RBEI, determined in 2011, verified the existence of this module in BE.<sup>17</sup> The x-ray crystal structure of 5'-AMP activated protein kinase (AMPK)  $\beta$ -subunit in complex with  $\beta$ -cyclodextrin (PDB 1Z0M) was determined, this structure shows a carbohydrate binding pocket containing all observed aspects of CBM domains and was identified as CBM48.<sup>18</sup> Despite large differences between the two structures, CBM48 of AMPK and of EcBE (residues 125-222) superimpose fairly well. Particularly, two residues, W159 and K189 (EcBE numbering) conserved in all branching enzymes, interact with the sugars bound in this site to AMPK and rice BEI in complex with maltopentaose structure (PDB 3VU2),<sup>19</sup> as they do in binding site IV in EcBE – maltoheptaose and maltooctaose complexes (**Figure 3.10**). Q211 of EcBE makes hydrogen bonding in this site to O2 and O3 of a single glucose unit, N150 does the same in AMPK to the neighboring glucose unit, while in RBEI, sequence alignment with EcBE shows that K121 occupies the position of Q211<sup>20</sup> and could potentially interact in the same way with the sugar if bound. D156 side chain adopts almost the same position as Q211 of EcBE and is also likely to make hydrogen bonds at this position. This indicates that many CBMs bind oligosaccharides in this site in the

same manner, although interactions here are not seen in all carbohydrate binding enzyme complexes, such as RBEI bound to maltododecaose (chapter 4), possibly pointing to less affinity of this site compared to other sites in BE.



**Figure 3.10: CBM48 overlay EcBE, RBEI and AMPK.** All residues and sugars are colored by atom type, with O in red, N in dark blue, and proteins' and oligosaccharides' carbons listed in order hereafter: bright and dark green (EcBE – M8), yellow and light blue (AMPK), and light pink and violet (RBEI – M5). The dotted black lines represent hydrogen bonds made by residues to their bound carbohydrates. W159 and K189 of EcBE and their counterparts in the other proteins make identical interactions with the corresponding sugars. Q211, though not conserved, similar residues in the other enzymes show comparable mode of binding.

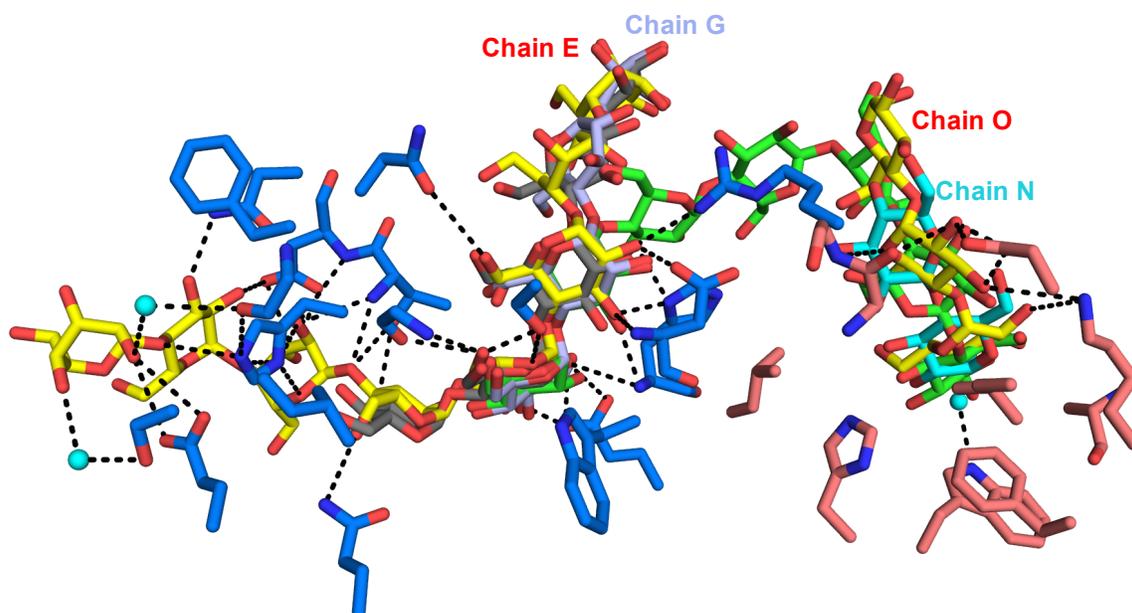
### 3.3.2 Comparing EcBE – M8 to EcBE – M7

#### 3.3.2.1 Binding sites I and II compared

The maltoheptaose bound EcBE structure is filled in molecule B with an M7 molecule bridging sites I and II, which bind the non-reducing and reducing ends, respectively. In this complex, the first 2 glucose units on each side of the oligosaccharide bind to the protein residues, while the middle 3 glucoses do not, except for couple interactions with Arg576 (green chain in **Figure 3.11**). Another sugar chain, colored in grey in **Figure 3.11**, occupies site I in Mol. D. This sugar adopts a different trajectory: the reducing end, although not making direct interactions with the protein, points away from the active site and from binding site II. In both of the sugars bound to sites I and II in this complex, we see site I binding the non-reducing end while site II binds the reducing end of the sugar.

In the EcBE – M8 complex, the total number of sugars bound in both sites, I and II, reaches four considering all molecules combined. In Mol. C alone, an M8 molecule binds entirely to site I, and overlaps with the sugar bound to Mol. D of the EcBE maltoheptaose complex. Both sugars have their reducing ends pointing in the same direction. In addition, Mol. B shows density for a Maltotetraose in this site (light blue in **Figure 3.11**), this sugar superimpose with the reducing end of the two sugars discussed here. Therefore, all sugars binding in this site have their reducing ends pointing away from site I interacting-residues and away from the protein's active site. This confirms that binding site I is exclusive for binding the non-reducing end of the substrate in EcBE.

In Mol. C of EcBE – M8 complex, another short sugar, a maltose, is visualized in site II. This sugar, colored in cyan in **Figure 3.11**, binds its reducing end in the same position as M7 (in green). The fact that the same molecule binding the M8 in site I also binds this sugar, shows that binding a longer sugar than M7 did not cause the oligosaccharide to span both sites, but instead to prefer binding the longer sugar in site I with the reducing end pointing away from the active site. This indicates that site I has stronger affinity to sugars than site II. As a matter of fact, residues involved in the interactions in site I are conserved among the members of the Enterobacteria family, while site II residues are far less conserved.<sup>20</sup> In addition, these facts indicate that site I works independently of site II; in other words, site II does not help support a sugar binding in site I. Moreover, the sugar colored in yellow in site II binds in Mol. A – EcBE / M8 complex and directs its reducing end away from the active site and in opposite direction to the other sugars discussed that bind in this site. This clearly indicates that binding site II is not restricted for binding the reducing end of a sugar. Mutations in this site did not affect the protein's activity tremendously as seen in chapter 2, even a triple mutant such as H596A+W595L+K546A kept activity of the truncated protein at 95% compared to the native truncated. Besides W595A, a mutation that lowered the protein's activity to 23% only in the truncated version but kept the full-length enzyme as active as the native full length, there is no mutation that affected the activity more than about 50%. All this evidence presented here lead to believe that binding site II only helps support the polymer on the surface of the protein without communicating directly with the active site.



**Figure 3.11: Overlay of all oligosaccharides bound in sites I and II in EcBE – M8 and EcBE – M7 complexes.** Residues and oligosaccharides shown here are same as in figure 3.4 with these sugars additions: in site I, chain G of Mol. B is colored in light blue, another sugar colored in grey binds to Mol. D of the EcBE – M7 complex; and in site II, chain N of Mol. C is colored in cyan. An M7 molecule spans sites I and II in EcBE – M7.

### 3.3.2.2 Binding sites III compared

In the EcBE – M7 complex, the electron density shows a maltose (green in **Figure 3.12.a**) that overlaps with chain L of the current structure in site III. The two glucose-unit sugar directs its reducing end towards that of the non-reducing end in the maltooctose structure, which also binds chain S, a single glucose sugar, in Mol. B at the same position. Chain S directs its reducing end similarly to the oligosaccharide that binds to the maltoheptaose bound complex at this position, in opposite direction to chain L. Chain S makes similar interactions to those that chain L makes, especially the sugar-protein stacking hydrophobic

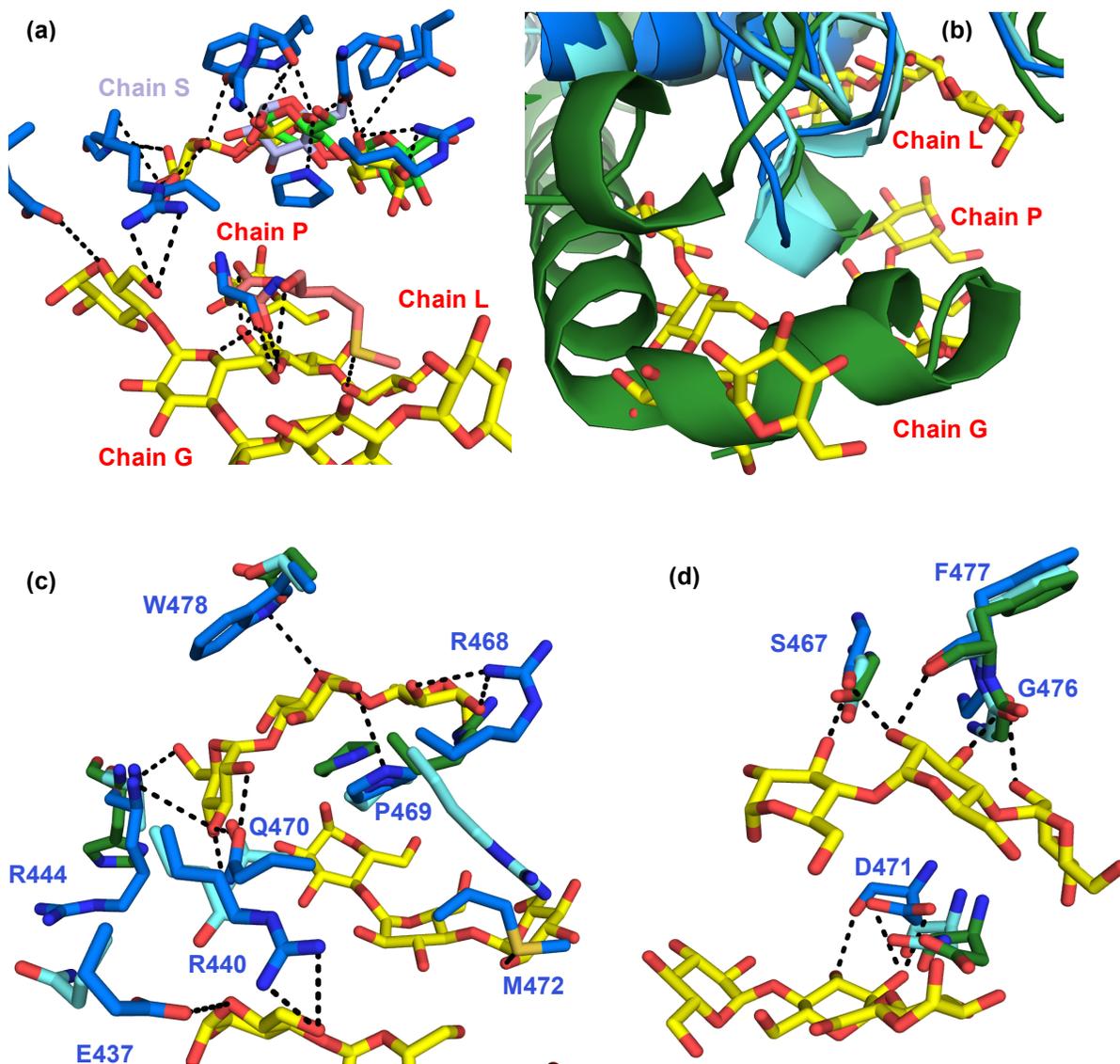
interaction with W478 in this site in addition to several other hydrogen bonds, despite the contrary direction. The sugar bound to the EcBE – M7 complex also does not miss the stacking interaction with W478. In addition, the slightly longer chain L of EcBE – M8 interacts with Glu437, Arg440, Arg444, Asp471, and Met472 (**Figure 3.5**), residues that did not show binding to the carbohydrate in this site before, even in the cyclodextrin bound EcBE structure.<sup>12</sup> Two other sugars, chain G bound to Mol. A and C and chain P bound to Mol. A in this site, occupy positions that were not previously filled in the EcBE – M7 complex (**Figure 3.5**). Chain L occupies an almost mid-position between chains G and P (as if a longer helix would wrap around and fill in the gaps): its non-reducing end is at 11.5 Å away from chain G's reducing-end, and its reducing end is at 7.7 Å from chain P's non-reducing end (**Figure 3.5**). Both of these distances could accommodate 3 and 2 glucose-units respectively. If an oligosaccharide longer than M8 is soaked in EcBE, it could span the same region now occupied by the 2 reducing-end residues of chain G, chains L and P, combined. It is possible that the loop structure formed of the binding residues in this site provides support to the natural polymer that wraps around it connecting the 3 bound chains during the branching reaction and as the sugar exits the protein.

To gain a better insight on the structure of other BEs in this site, we compared the structures of a glycogen BE, *Mycobacterium tuberculosis* BE (MycoBE),<sup>21</sup> and a starch BE, RBEI (data of chapter 4), to EcBE around this site. The structural features of MycoBE overlap well with those of EcBE and could therefore bind oligosaccharides in the same location, while RBEI clashes with all

of chain G and parts of L and P (**Figure 3.12.b**). We also overlaid the residues interacting in EcBE that have counterparts in the other 2 structures (**Figures 3.12.c, 3.12.d**). Pro469, Gly476 and Phe477 are conserved among all the members of GH13 family enzymes regardless of the species.<sup>20</sup> Several other amino acids occupying similar locations to their counterparts in EcBE could make nearly the same interactions as in the EcBE – M8 complex. For instance, a Thr473 in MycoBE and a Cys408 in RBEI replace Ser467, whose main chain interacts with the sugar in EcBE; the 3 residues main chains overlap allowing for analogous interactions. Same is the case of R444 replaced by Ala450 in MycoBE and His385 in RBEI. Three residues in EcBE, E437, R440, and Q470, completely not conserved in RBEI, have their side or main chains overlapping with Q443, Q446 and T476 in MycoBE, which are therefore capable of forming similar contacts with the substrate. R468, though conserved in the 2 other structures, its counterparts adopt different orientations, especially R474 of MycoBE, which can actually interact with chain P similarly to M472. A special case, D471 of EcBE, is conserved in the other 2 proteins but does not keep the exact same location. D471 makes 3 H-bonds with chain P, two with its side chain and one its main; its counterparts are close enough to bind in this location as well. Trp478 seems to be key for oligosaccharide binding in this site of EcBE since all three bound chains to this site, L, G and P, make hydrophobic stacking interaction with this residue, yet it is not conserved among the  $\alpha$ -amylase family nor is it among other BEs; it is replaced by Asp419 in RBEI and Ser484 in MycoBE. This is confirmed by the activity assay data previously presented that showed W478A to only

decrease the activity by about 50% in the truncated version of the protein. W478, a seemingly critical residue by examining the structure, is essential for holding the substrate but not for the branching activity. This result also means that supporting the polymer is not mostly dependent on W478 as is the case for some other sites with a tryptophan essential residue, although all three chains overlapping in the maltoheptaose and maltooctaose complexes with EcBE stack against it; likewise, this result aligns with the belief that this site is the putative substrate exit site for GH13 family enzymes. Carbohydrate binding enzymes, other than BEs, that bind oligosaccharides in their active site and/or external sites, not only lack binding in site III but also some disqualify for it due to structural features filling this area.<sup>11, 19, 22-34</sup> This indicates that BE, which resembles the structures of other proteins in the same family, has a different mode of binding, definitely linked to the difference in the reaction it performs.

Site VI is another site located at the exit of the glucan. Although we do not see binding in this site in the present structure, which could be due to the different crystal packing or to the longer oligosaccharide soaked compared to the EcBE – M7 complex, activity assay data proved that binding site VI is critical for the protein's activity as a single mutation W628A resulted in a dead protein (4% activity). Sites III and VI seem then to play different roles where one supports the sugar and the second directly communicates possibly an acceptor chain with the active site.

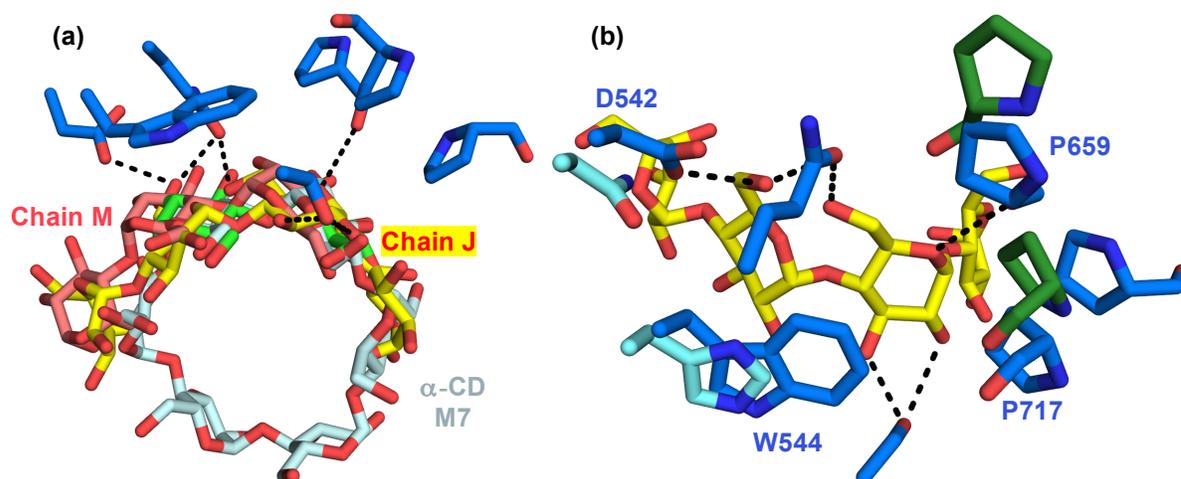


**Figure 3.12: Comparing EcBE, MycoBE and RBEI around binding site III.** The structures are colored in blue, cyan and green, respectively. **a.** Overlay of all oligosaccharides bound in site III. Coloring is as in in figure 3.5; yellow chains (labeled in red) and light purple chain belong to the EcBE – M8 complex. **b.** Overlay of loops from the 3 structures. While RBEI clashes with chains G, and parts of P and L (EcBE – M8 complex chains), MycoBE overlaps well with EcBE and could make similar interactions with its substrate. **c.** Overlay of MycoBE and RBEI residues that are able to make analogous interactions with all three chains. **d.** Same as b, showing other interactions in order not to crowd the figure.

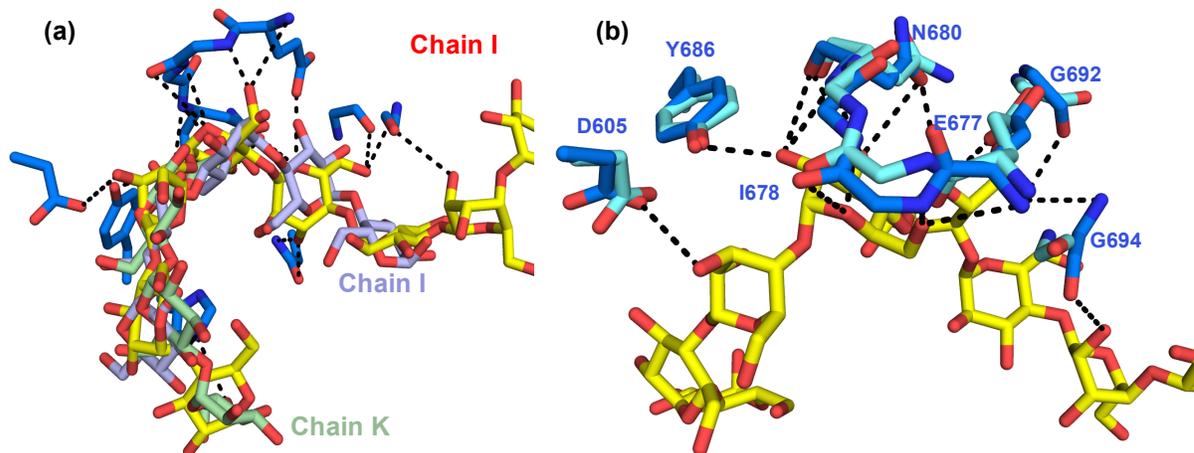
### 3.3.2.3 Binding sites V

In site V, two chains, J (Mol. B, yellow) and M (Mol. A, deep salmon), overlap together and with a maltose (bright green) and an  $\alpha$ -cyclodextrin (light cyan) from the EcBE – M7 complex (**Figure 3.13.a**). All four binding sugars direct their non-reducing end away from the active site, except for chain J. In addition, the four chains make all the same interactions with the protein's residues. An important hydrophobic sugar-residue stacking interaction between the ligand and W544 is seen in all the molecules that bind in this site in both complexes (**Figure 3.13.a**). Mutation of this residue to an Ala showed a reduced activity to about 25 % in the truncated protein, however the mutation to an Ala or a Lys increased the activity in the full-length version. This site seems also to hold the substrate polymer on the protein's surface instead of directing the ligand towards the active site.

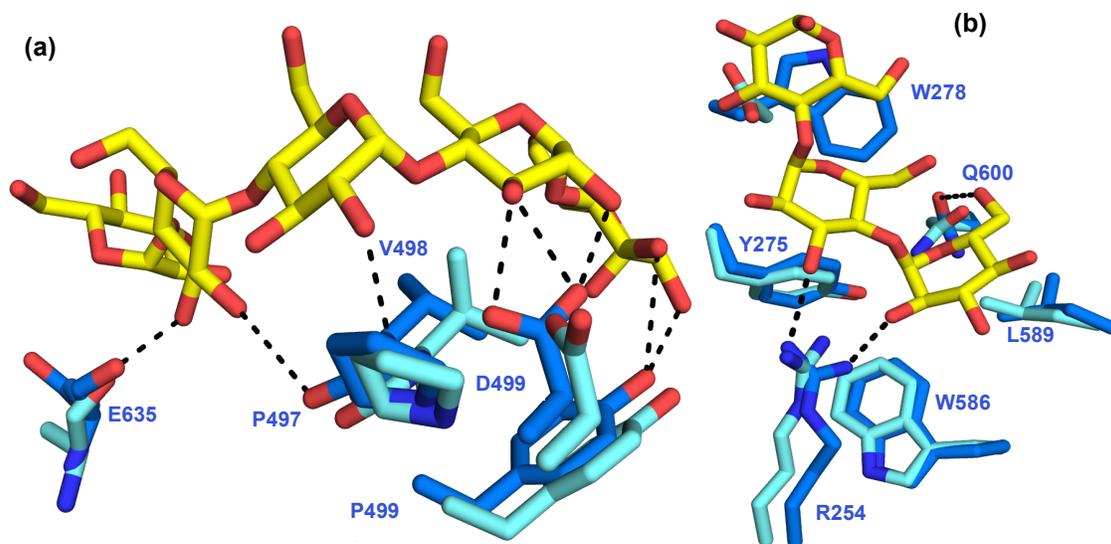
Examining site V in the EcBE – M8 complex confirms that this site is essential for *E. coli* BE external binding of the sugar rather than for all BEs because we find in RBEI and in MycoBE couple residues in this site that are either conserved or adopting the same orientation. In MycoBE, N548 and D550 replace D542 and W544, and in RBEI, P616 and P684 replace P659 and P717, respectively (**Figure 3.13.b**). In addition, no binding is seen in this site in any of the carbohydrate-binding-enzymes structures bound to substrates currently available.



**Figure 3.13: Binding site V compared.** **a.** Overlay of oligosaccharides from the M8 (yellow and deep salmon) and the M7 (light cyan and bright green) complexes with EcBE (residues in blue). **b.** Overlay of MycoBE (cyan) and RBEI (dark green) residues that can make analogous interactions to those of EcBE (blue).



**Figure 3.14: Binding sites IX compared.** **a.** Overlay of oligosaccharides in the EcBE – M8 complex: chain I (Mol. A) is colored in yellow, H (Mol. B) in light blue, and K (Mol. D) in pale green. **b.** Most residues of EcBE (blue) have superimposable counterparts in MycoBE (cyan); those that do not are not shown for either structure. Oligosaccharide bound is shown in yellow and black dots represent H-bonds.



**Figure 3.15: Binding sites occupied in one molecule of EcBE – M8.** Residues of EcBE that have superimposable counterparts in MycoBE are shown. Coloring is as figure 3.14-b. **a.** Binding site X. **b.** Binding site VIII: W478 is the only residue that is shown but not conserved.

### 3.3.2.4 Binding site VIII, IX and X

Binding site IX is occupied in 3 molecules of the EcBE – M8 complex. Chain I (M8, Mol. A) directs its reducing end to the opposite side of chain H (M5, Mol. B) and chain K (M4, Mol. D) (**Figure 3.14.a**). The majority of the interactions made in this site are concentrated in the center of the longest oligosaccharide; so that the other two chains do not miss most contacts with the protein. In fact, chain H, three glucose-units shorter than chain I, interacts with three additional residues, R601, S683, N693, which the longest chain does not come close enough to in molecule A (**Table 3.2**).

Binding site X, filled only in molecule C, along with sites V and IX have the characteristic of bridging the central TIM barrel domain with the COOH-terminal domain. All three sites interact with residues from both domains (**Table 3.2**). This probably explains why they are not conserved in any other GH13 family enzyme. Overlaying several structures shows that structural features in these sites either clash or are far enough to lack binding with the sugars bound in the present structure.<sup>11, 19, 22-34</sup>

Site VIII, which is on the opposite side of the protein in comparison to sites V and X, belongs to the central domain. Nevertheless, superimposing with the rest of the GH13 enzymes shows clashing with the sugar. Overlaying interacting residues from sites VIII, IX and X with the MycoBE and RBEI residues shows a different scenario than the overlay with the other GH13 enzymes. RBEI exhibits one loop in each case that clashes with the sugar and another that could potentially bind however it shows only a single residue conserved with that of EcBE but in a slightly different location. MycoBE site is surprisingly very well conserved in all three of these sites, with most of the residues overlapping with the EcBE ones (**Figure 3.14.b, 3.15.a and 3.15.b**). The residues that are not conserved in MycoBE are omitted from the figures, except for one, W278, which makes a hydrophobic stacking interaction with the sugar but is replaced by an Asp in MycoBE. Still, the overall interactions in sites VIII, IX and X seem to be conserved with EcBE, hence the similar substrate, glycogen.

### 3.4 Hypothesis and Conclusion

Although the GH 13 family enzymes exhibit similar external binding sites to those seen in EcBE,<sup>31, 35</sup> none superimpose exactly with any of the eleven sites observed in the maltoheptaose and maltooctaose complexes. Even those enzymes that bind oligosaccharides on the protein's surface do not present more than three disparate binding sites, however, in contrary to EcBE, most proteins will bind a shorter glucan than M8 in the active site. In addition, the active site in all four molecules is not encumbered by many loops like in the other structures, nor is it by the symmetry molecules. Clearly, a maltooctaose is not long enough to bind in the conspicuous groove, expected to bind the substrate. It is also possible that the active site has a lower affinity for oligosaccharides than the peripheral sites. BE that only reacts with long polymers such as glycogen and starch transfers chains of at least five glucose units but has a stronger affinity for chains of DP 10-12.<sup>36</sup> Furthermore, despite the discovery of four new binding sites and the more extended interactions seen in sites I and III in comparison with the maltoheptaose – EcBE complex, none of the newly interacting residues is closer to the active site more than site I.

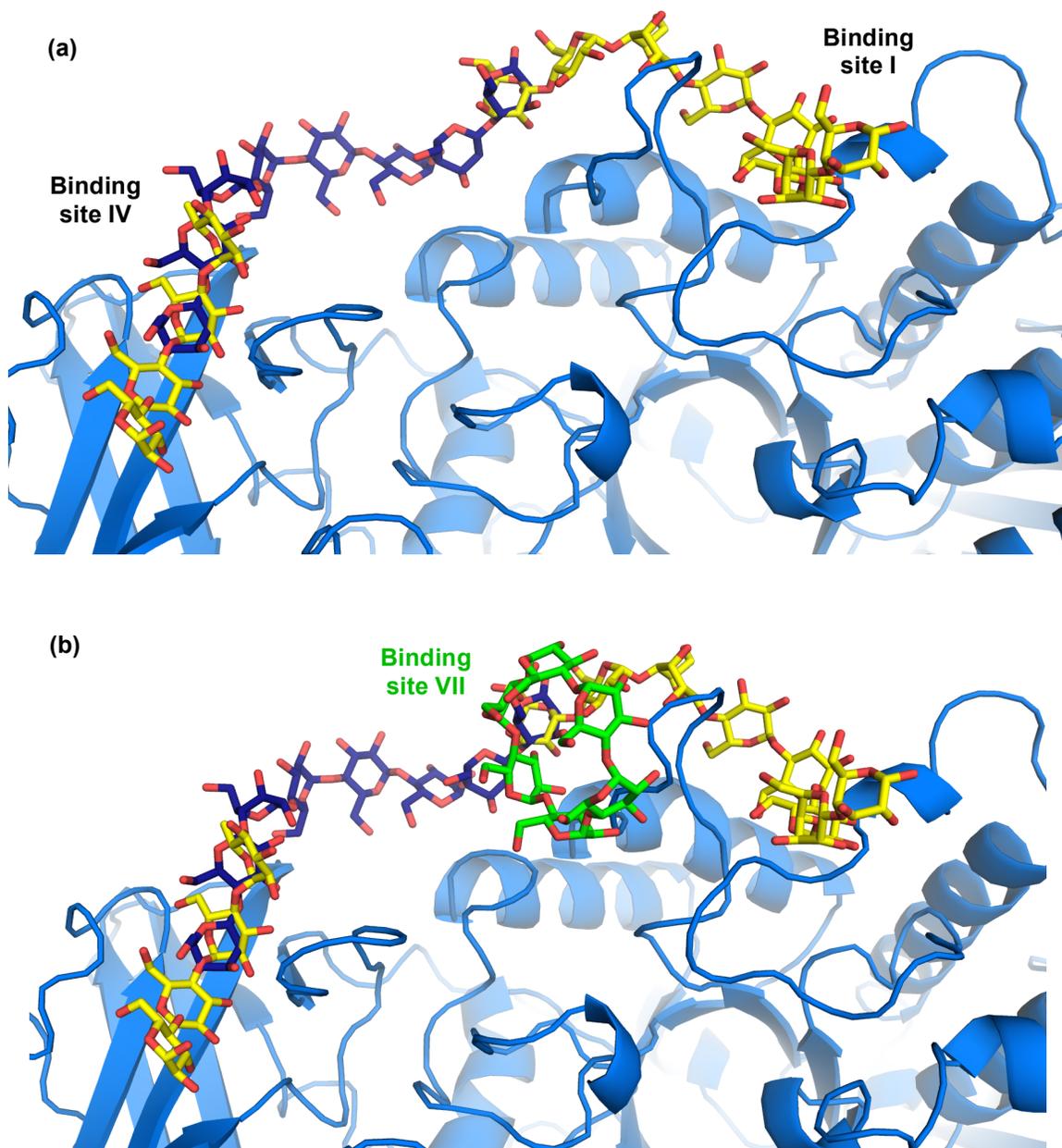
Furthermore, comparing the binding sites discloses of an important feature of EcBE. All the sites are rather spaced away by at least 8 Å from each other; such are sites V and II. The exceptions for this finding is the distance from site VIII to site I, where a single residue, W586, makes a side-on hydrophobic interaction with site VIII while making a hydrogen bond with a glucose in site I, and the 3.3 Å distance between residues binding in sites V and X (CA of S689 and N690). The

relatively long distance between binding sites proves that BE cannot transfer chains shorter than six units and prefers chains of  $DP \geq 8$  or more units.<sup>4</sup> Close examination of the binding sites that approach each other further shows that the oligosaccharides bound run parallel to each other. Modeling an M8 extending from site V lays the oligosaccharide around the surface of the protein but cannot meet any end of sites IX or X bound sugars. In addition, site VIII sugar adopts a perpendicular orientation to those of sites V, IX and X. Although a single residue binds to both sites I and VIII, it appears that the sugar in site VIII directs away from the active site as that of binding site I prohibiting connection between the 2 oligosaccharides. Also, the sugars in sites V – IX, and V – X bind parallel rather than linearly even though they are at a distance of less than a maltose apart. The binding sites then provide ground for random sugars binding rather than simultaneous binding to one or more external site or the active site, indicating that concurrent binding only happens if the substrate is a long polymer. In addition, EcBE transfers rather longer polymers as mentioned above. It is known that BE acts mainly by transglycosylation and not by hydrolysis,<sup>37, 38</sup> so it cleaves a chain from the donor glucan then transfers it into the acceptor chain at position  $\alpha$ -1,6 therefore binding both chains simultaneously; this makes the superior number of glucan binding sites over any other enzymes of the GH13 family not very surprising. These facts lead to believe that the different external binding sites play distinguished roles in the protein's branching reaction. However, we could also conclude that some of the external binding sites (II, IV and V) constrain the protein onto its substrate, while others (I, VI and VII) are essential

for the branching reaction as they surely demolished the activity when mutated. Sites IX and X appear to have similar roles as site V because of location and parallel binding, biochemical activity and chain transfer assays on the newly discovered binding sites residues will reveal more information on the possible roles of each and combined sites.

Sites I and IV are located in an appropriate position to bind the non-reducing end of the sugar. In fact, site I has proven to bind only the non-reducing end in both protein complexes with M7 and M8. In contrary to sites V, IX and X, sites I and IV bound sugars appear to align continuously. Therefore, we modeled in an M8 between the two sites extending from the non-reducing end unit in site I towards site IV and found out that 5 glucose units are needed connect the 2 sugars unbroken by clashes with the protein as well as interacting with several residues along the way (**Figure 3.16.a**). We also know from assay data that site I is critical for the enzyme's activity (R255A – 2%, E590A – 12%), while site IV, though also necessary, has a lower impact on the protein's activity (W159A – 30%, Q211A – 63%, Q211R – 12%). This knowledge recommends that the non-reducing end of the long polymer binds to site I, which sends the sugar directly to the active site, while site IV holds the extension emanating from site I. It is very likely that the sugars bound to site I so far observed point away from the active site because they were not long enough to reach to the active site (18 Å between site I and the active site, which fit at least 4 sugars) and site I interacting residues have a higher affinity towards oligosaccharides than the active site ones.

On the other hand, binding site VII, which does not bind a sugar in the EcBE – M8 complex for similar reasons to those explained about site VI (crystal packing and longer sugar), binds only a cyclodextrin in the protein's complex with M7 but does not bind a linear saccharide. Comparing sites I and VII clarifies that site I shares with site VII binding to the same protein residues. In fact, the angle between the plane going transversally through the glucose units of the cyclodextrin and that laying with the first glucose on the non-reducing end of binding site I, is an acute angle, indicating that these two sites do not communicate together, so the sugar can not turn from site I towards site VII (**Figure 3.16.a**). In addition, both of these sites turned out critical for the protein's activity, and a binding site that can tremendously reduce the activity must be interacting with the active site directly and probably working independently of other sites. These combined facts may indicate that site I directs the reducing end towards the active site, where it is cleaved and sent towards site VII to be bound to the protein while the acceptor glucan enters the active site for branching. Additional combined activity assay mutations are needed to conclude more definitive separate roles of binding sites.



**Figure 3.16: Binding sites I, IV, and VII relation explained.** **a.** Modeling a sugar between binding sites I and IV shows that five glucose units fill in the space between the 2 sugars, extending from the non-reducing end in site I towards site IV. Modeled sugar is in dark blue, bound sugars are in yellow and protein cartoon in blue. All oxygens are colored in red. **b.** Binding site VII sugar is shown in green. The angle between the plan of the cyclodextrin and that of the closest glucose from site I is an acute angle and the sugars do not actually connect.

## REFERENCES

## REFERENCES

1. Baba, T., Kimura, K., Mizuno, K., Etoh, H., Ishida, Y., Shida, O., and Arai, Y. (1991) Sequence conservation of the catalytic regions of amylolytic enzymes in maize branching enzyme-I *Biochem. Biophys. Res. Commun.* **181**, 87-94.
2. Mizuno, K., Kimura, K., Arai, Y., Kawasaki, T., Shimada, H., and Baba, T. (1992) Starch Branching Enzymes from Immature Rice Seeds, *Journal of Biochemistry* **112**, 643-651.
3. Romeo, T., Kumar, A., and Preiss, J. (1988) Analysis of the *Escherichia coli* glycogen gene cluster suggests that catabolic enzymes are encoded among the biosynthetic genes, *Gene* **70**, 363-376.
4. Binderup, K., Mikkelsen, R., and Preiss, J. (2000) Limited proteolysis of Branching Enzyme from *Escherichia coli*, *Archives of biochemistry and biophysics* **377**, 366-371.
5. Abad, M. C., Binderup, K., Rios-Steiner, J., Arni, R. K., Preiss, J., and Geiger, J. H. (2002) The X-ray crystallographic structure of *Escherichia coli* Branching Enzyme, *Journal of Biological Chemistry* **277**, 42164-42170.
6. Bao, Y., Kishnani, P., Wu, J. Y., and Chen, Y. T. (1996) Hepatic and Neuromuscular Forms of Glycogen Storage Disease Type IV Caused by Mutations in the Same Glycogen-branching Enzyme Gene, *J. Clin. Invest.* **97**, 941-948.
7. Chen, Y. T., and Burchell, A. (1995) Glycogen Storage Diseases, *The Metabolic and Molecular Basis of Inherited Diseases. C.R. Scriver, A.L. Beaudet, W.S. Sly, and D. Valle, eds. (New York, McGraw-Hill, Inc., Health Professions Division)* **1**, 935-965.
8. DiMauro, S., and Tsujino, S. (1994) Non-Lysosomal Glycogenoses, *Myology (Engel, A. G. and Franzini-Armstrong, C., eds)* **2**, 1554-1576.
9. Schroder, J. M., May, R., Shin, Y. S., Sigmund, M., and Nase-Hiippmeier, S. (1993) Juvenile hereditary polyglucosan body disease with complete branching enzyme deficiency (type IV glycogenosis), *Acta Neuropathol.* **85**, 419-430.
10. Borovsky, D., Smith, E. E., and Whelan, W. J. (1975) Purification and Properties of Potato 1,4-alpha-D-Glucan: 1,4-alpha-D-Glucan 6-alpha-(1,4-alpha-Glucano)-Transferase, *Eur. J. Biochem.* **59**, 615-625.

11. Uitdehaag, J. C. M., van Alebeek, G. J., Van Der Veen, B., Dijkhuizen, L., and Dijkstra, B. W. (2000) Structures of Maltohexaose and Maltoheptaose Bound at the Donor Sites of Cyclodextrin Glycosyltransferase Give Insight into the Mechanisms of Transglycosylation Activity and Cyclodextrin Size Specificity, *Biochemistry* 39, 7772-7780.
12. Feng, L. (2009) X-Ray Crystallographic Studies of Branching Enzyme/Polysaccharide Complex and RNA Polymerase III Transcription Factor TFIIIB Complex, *Dissertation*, p. 10.
13. Otwinowski, Z., and Minor, W. (1997) Processing of X-Ray Diffraction Data Collected in Oscillation Mode, *Methods in Enzymology* 276, 307-326.
14. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures, *Acta Crystallographica D* 67, 355-367.
15. Murshudov, G. N., Vagin, A. A., and J., D. E. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method, *Acta Crystallographica D, EU Validation contract: BIO2CT-92-0524* 53, 240-255.
16. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics, *Nucleic acids research* 37, D233-238.
17. Noguchi, J., Chaen, K., Vu, N. T., Akasaka, T., Shimada, H., Nakashima, T., Nishi, A., Satoh, H., Omori, T., Kakuta, Y., and Kimura, M. (2011) Crystal structure of the Branching Enzyme I (BEI) from *Oryza sativa* L with implications for catalysis and substrate binding, *Glycobiology* 21, 1108-1116.
18. Polekhina, G., Gupta, A., van Denderen, B. J., Feil, S. C., Kemp, B. E., Stapleton, D., and Parker, M. W. (2005) Structural basis for glycogen recognition by AMP-activated protein kinase, *Structure* 13, 1453-1462.
19. Chaen, K., Noguchi, J., Omori, T., Kakuta, Y., and Kimura, M. (2012) Crystal structure of the rice branching enzyme I (BEI) in complex with maltopentaose, *Biochemical and biophysical research communications* 424, 508-511.
20. Feng, L., Fawaz, R., Hovde, S., Gilbert, L., Chiou, J., and Geiger, J. H. (2015) Crystal Structures of Escherichia coli Branching Enzyme in Complex with Linear Oligosaccharides, *Biochemistry* 54, 6207-6218.

21. Pal, K., Kumar, S., Sharma, S., Garg, S. K., Alam, M. S., Xu, H. E., Agrawal, P., and Swaminathan, K. (2010) Crystal structure of full-length *Mycobacterium tuberculosis* H37Rv Glycogen Branching Enzyme: insights of N-terminal beta-sandwich in substrate specificity and enzymatic activity, *The Journal of biological chemistry* 285, 20897-20903.
22. Dumbrepatil, A. B., Choi, J. H., Park, J. T., Kim, M. J., Kim, T. J., Woo, E. J., and Park, K. H. (2010) Structural features of the Nostoc punctiforme debranching enzyme reveal the basis of its mechanism and substrate specificity, *Proteins* 78, 348-356.
23. Kanai, R., Haga, K., Akiba, T., Yamane, K., and Harata, K. (2004) Biochemical and Crystallographic Analyses of Maltohexaose-Producing Amylase from Alkalophilic *Bacillus* sp. 707, *Biochemistry* 43, 14047-14056.
24. Katsuya, Y., Mezaki, Y., Kubota, M., and Matsuura, Y. (1998) Three-dimensional Structure of *Pseudomonas* Isoamylase at 2.2 Å Resolution, *Journal of molecular biology* 281, 885-897.
25. Knegt, R. M. A., Strokopytov, B., Penninga, D., Faber, O. G., Rozeboom, H. J., Kalk, K. H., Dijkhuizen, L., and Dijkstra, B. W. (1995) Crystallographic Studies of the Interaction of Cyclodextrin Glycosyltransferase from *Bacillus circulans* Strain 251 with Natural Substrates and Products, *Journal of Biological Chemistry* 270, 29256-29264.
26. Malle, D., Iwamoto, H., Katsuya, Y., Utsumi, S., and Mikami, B. (TBA) Crystal structure of pullulanase type I from *Bacillus subtilis* str. 168 in complex with maltose and alpha-cyclodextrin, *TBA*.
27. Mikami, B., Adachi, M., Kage, T., Sarikaya, E., Nanmori, T., Shinke, R., and Utsumi, S. (1999) Structure of Raw Starch-Digesting *Bacillus cereus* beta-Amylase Complexed with Maltose, *Biochemistry* 38, 7050-7061.
28. Mikami, B., Iwamoto, H., Malle, D., Yoon, H. J., Demirkan-Sarikaya, E., Mezaki, Y., and Katsuya, Y. (2006) Crystal structure of pullulanase: evidence for parallel binding of oligosaccharides in the active site, *Journal of molecular biology* 359, 690-707.
29. Payan, F., and Qian, M. (2003) Crystal Structure of the Pig Pancreatic 'Alpha-Amylase Complexed with Malto-Oligosaccharides, *J Protein Chem* 22, 275-284.
30. Qian, M., Haser, R., Buisson, G., Duee, E., and Payan, F. (1994) The Active Center of a Mammalian  $\alpha$ -Amylase. Structure of the Complex of a Pancreatic alpha-Amylase with a Carbohydrate Inhibitor Refined to 2.2-Å Resolution, *Biochemistry* 33, 6284-6294.

31. Strokopytov, B., Penninga, D., Rozeboom, H. J., Kalk, K. H., Dijkhuizen, L., and Dijkstra, B. W. (1995) X-ray Structure of Cyclodextrin Glycosyltransferase Complexed with Acarbose. Implications for the Catalytic Mechanism of Glycosidases, *Biochemistry* 34, 2234-2240.
32. Tan, T. C., Mijts, B. N., Swaminathan, K., Patel, B. K., and Divne, C. (2008) Crystal structure of the polyextremophilic alpha-amylase AmyB from *Halothermothrix orenii*: details of a productive enzyme-substrate complex and an N-domain with a role in binding raw starch, *Journal of molecular biology* 378, 852-870.
33. Uitdehaag, J. C. M., Mosi, R., Kalk, K. H., van der Veen, B., Dijkhuizen, L., Withers, S. G., and Dijkstra, B. W. (1999) X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the a-amylase family, *Nature Structural Biology* 6, 432-436.
34. Vester-Christensen, M. B., Abou Hachem, M., Svensson, B., and Henriksen, A. (2010) Crystal structure of an essential enzyme in seed starch degradation: barley limit dextrinase in complex with cyclodextrins, *Journal of molecular biology* 403, 739-750.
35. Robert, X., Haser, R., Gottschalk, T. E., Ratajczak, F., Driguez, H., Svensson, B., and Aghajari, N. (2003) The Structure of Barley  $\alpha$ -Amylase Isozyme 1 Reveals a Novel Role of Domain C in Substrate Recognition and Binding, *Structure* 11, 973-984.
36. Binderup, K., Mikkelsen, R., and Preiss, J. (2002) Truncation of the amino terminus of Branching Enzyme changes its chain transfer pattern, *Archives of biochemistry and biophysics* 397, 279-285.
37. Mikkelsen, R., Binderup, K., and Preiss, J. (2001) Tyrosine residue 300 is important for activity and stability of branching enzyme from *Escherichia coli*, *Archives of biochemistry and biophysics* 385, 372-377.
38. Takata, H., Kuriki, T., Okada, S., Takesada, Y., Iizuka, M., Minamiura, N., and Imanaka, T. (1992) Action of Neopullulanase: Neopullulanase catalyzes both hydrolysis and transglycosylation at alpha-1,4 and alpha-1,6-glycosidic linkages., *Journal of Biological Chemistry* 267, 18447-18452.

## CHAPTER 4: X-RAY CRYSTALLOGRAPHIC STUDIES OF RICE BEI IN COMPLEX WITH LINEAR OLIGOSACCHARIDES

### 4.1 Research Objective

Branching enzyme (BE) belongs to the  $\alpha$ -amylase family of enzymes and shares several structural and functional features with these members, mainly the TIM barrel domain.<sup>1, 2</sup> Site directed mutagenesis studies of the active center in neopullulanase<sup>3</sup> suggested that the amylolytic enzymes share a common catalytic mechanism.<sup>4</sup> The active site residues involved in catalysis are confined to four conserved primary sequence regions existing in all GH13 enzymes<sup>5</sup> yet the reactions catalyzed vary between hydrolysis and/or transglycosylation of  $\alpha$ -1,4 and  $\alpha$ -1,6 glycosidic linkages.<sup>4, 6</sup> BE catalyzes the cleavage of  $\alpha$ -1,4-glycosidic linkages followed by the transfer of the freshly dissected branch to form  $\alpha$ -1,6-branches. Consequently, the enzyme plays a critical role in the biosynthesis of starch in plants and glycogen in animals and bacteria. In addition, branching enzymes (BEs) from different sources produce very diverse molecules of starch and glycogen, varying tremendously in molecular weights, sizes of chains, and numbers of branches.<sup>7-21</sup>

To date, several starch branching enzymes (SBEs) have been identified from multiple sources including maize,<sup>5, 22-30</sup> rice,<sup>31-36</sup> wheat,<sup>37-40</sup> barley,<sup>41-43</sup> potato,<sup>44, 45</sup> cassava,<sup>46</sup> and pea and soybean.<sup>47-49</sup> In rice endosperm, there are three isoforms of the branching enzymes, BEI, BEIIa, and BEIIb, the later two being products of the same gene<sup>31, 33, 34</sup> that is differentiated after transportation into

the amyloplast<sup>33, 50</sup> and post-translational modification of the synthesized precursor gene.<sup>33</sup> It is known that BEI and BEIIb are typically specifically expressed in endosperm while BEIIa is omnipresent in every tissue,<sup>32, 36</sup> suggesting a different enzymatic role. A comprehensive biochemical study on the properties of rice BE isozymes<sup>36</sup> supports a previous observation stating that the branching of the amylopectin amorphous lamellae is catalyzed by BEI while the branches residing on the border between the amorphous and the crystalline lamellae are almost exclusively catalyzed by BEIIb.<sup>51</sup> In this thorough study, Nakamura, Y. *et al.* subjected the 8-amino-1,3,6-pyrenetrisulfonic acid (APTS)-labeled products of RBEI, RBEIIa and RBEIIb enzymatic reactions towards amylopectin and amylose debranched by bacterial debranching enzymes to capillary electrophoresis, unprecedented experience then and superior over most previous investigations, which analyzed the reaction products by gel filtration or HPAEC-PAD. They demonstrated that, in regards to amylopectin, BEIIa transfers a wide range of relatively short outer chains (DP6 – 15), and uses donor substrates of intermediate (DP17 – 26) and long (DP ≥ 40) chains, while BEIIb transfers chains of DP6 and DP7 and attacks almost exclusively intermediate outer chains as donor substrates. On the other hand, BEI forms branches of DP6 – 15 from donor chains of DP ≥ 12; BEI must be active towards inner chains since it also produces branches of DP26 – 39, in opposite to the other two isozymes.<sup>36</sup> This study also indicates that all BE isoforms can react with both amylose and amylopectin, and their distinct chain length preferences are not changed by glucan substrates.

They also differ in substrate preference; in vitro experiments showed that maize BEI mainly acts on amylose to create a smaller number of branches than do BEII that preferentially branches amylopectin.<sup>24, 27, 30</sup> Studies on the chain length transferred in maize isozymes of the protein revealed more diversity among them; BEI essentially transfers chains with degree of polymerization (DP) larger than 10, while BEII predominantly transfers shorter chains with DP 3-9 but mostly DP6 and DP7<sup>24</sup>. Although there are an abundance of biochemical studies on the rice BE isozymes, the detailed mode of binding to substrate and catalytic mechanism remain obscure.

The crystal structure of RBEI was recently determined.<sup>50</sup> The sequence of the mature RBEI (755 amino acids) consists of a carbohydrate-binding module 48 (CBM48) (residues 59-160), the central catalytic domain (residues 161-587), and an  $\alpha$ -amylase C-terminal domain (residues 588-702).<sup>50</sup> The overall structure forms a groove-like shape between two globular domains, where the polysaccharide is expected to bind.<sup>50</sup> Two residues residing in the active center, Asp344 and Glu399, are believed to play critical roles in the catalysis reaction as a nucleophile and a general acid/base, respectively.<sup>50, 52</sup> The crystal structure of rice BEI in complex with maltopentaose identified three binding sites, 1, 2, and 3.<sup>53</sup> In site 1, the oligosaccharide is sandwiched between the N-terminal, CBM48, and  $\alpha$ -amylase domains; in site 2, it is stabilized by two strands in the CBM48 domain; and in site 3, the sugar is located between the N-terminal and catalytic domains. This confirms CBM48's role in carbohydrate binding in BEI. Specific interactions between CBMs and oligosaccharides have been previously studied

in detail.<sup>54</sup> *Aspergillus niger* gluco-amylase (A. amy) starch binding domain, a member of the CBM20 family, is the most well studied of all CBMs.<sup>55</sup> Comparing the structures of CBM20 (A. amy) and CBM48 (BEI) reveals that the first and second sites in A. amy resemble sites 2 and 1 in BEI, respectively. Site 2 showed a weak electron density in the BEI-maltopentaose complex, while site 1 had an obvious clear density, indicating the higher affinity site 1 has for the oligosaccharide than 2, similar to the higher binding affinity seen in the second site over the first site in A. amy.<sup>53, 55</sup> Binding longer oligosaccharides to BEI might present different binding sites or explain the three sites discovered more precisely. We present here the crystal structure of BEI in complex with a twelve glucose-unit saccharide.

## 4.2 Why a Twelve Unit Oligomer?

The crystal structure of rice BEI was previously determined in its apo form<sup>50</sup> and later the protein was co-crystallized with maltopentaose.<sup>53</sup> Although binding was seen in three distinct sites in BEI bound to maltopentaose (BEI-M5), no sign of binding was detected at or near the active site, or in the substrate-binding groove, despite excessive carbohydrate concentration in the co-crystallization. In fact, several prolonged loops encumber the central domain limiting substrate access, in comparison with *E. coli* branching enzyme bound to maltooctaose (EcBE-M8). Therefore, binding to the external binding sites might change those loop conformations and lead to active center binding.<sup>53</sup> Since BEI typically reacts with long oligosaccharides, and preferentially transfers glucan chains over 8 residues in length, it was anticipated that longer oligosaccharides might identify new binding sites.

In addition, BEI transfers chains up to 40 units in length but most preferably those with degree of polymerization DP 6-15.<sup>36</sup> This fact gives great potential for active center substrate binding when using oligosaccharides of  $DP \geq 6$ . We obtained maltooctaose and maltododecaose, generously provided by our collaborators Dr. Kwan Hwa Park and Dr. Sung Hoon Park maltooctaose (M8) and maltododecaose (M12).

As shown for the potato-tuber-SBEI (**Table 4.1**), SBEI isoforms have significantly increased affinity for longer malto-oligosaccharides.<sup>56</sup> Since rice BEI shares 81% sequence identity with the potato corresponding protein,<sup>33</sup> and both enzymes exhibit similar patterns of action,<sup>57, 58</sup> we expect rice BEI to have

comparable affinities for oligosaccharides. Although M8 and M12 are not reported in this table but the DP = 10 and DP = 13 samples represent a mixture of oligosaccharides lengths with averages of 10 and 13 units, respectively.

Compound	Dissociation Constant (mM)
Maltose	> 50
Maltotriose	11.7
Maltotetraose	1.1
Maltopentaose	0.75
Maltohexaose	0.25
Maltoheptaose	0.16
DP 10	0.14
DP13	0.13

**Table 4.1: Dissociation constants for multiple  $\alpha$ -glucans and potato tuber BEI.** <sup>56</sup>

### 4.3 The Three Dimensional Structure of Rice BE I Bound to Dodecaose

#### 4.3.1 Crystallization: attempts and successes

In rice endosperm, rice branching enzyme I is synthesized as an 820-amino acid predecessor, and then transported to the amyloplast where it loses the N-terminal signaling sequence formed of the n-terminal 65 amino acids, resulting in the mature RBEI (1-755 amino acids). Noguchi *et al.* subjected the protein to limited trypsin digestion that resulted in the truncation of 53 amino acids at the C-terminus; the resulting fragment was called BEI core domain (BEI $\Delta$ C).<sup>50</sup> Crystal structures of both rice mature BEI and BEI $\Delta$ C were obtained. We obtained the full-length rice BEI gene from the National Institute of Agrobiological Sciences in Japan.<sup>59</sup> The complete DNA sequence can be found on the Rice Genome Resource Center's website reference above. We cloned the gene into a modified pet28a-Sumo vector as described in Chapter 6. The protein was expressed in the BL21 codon plus strain of *E. coli*, purified to about 90% homogeneity as estimated from SDS-gel by Ni-agarose affinity chromatography confirmed protein identity by Western blot against anti-His antibody, and was further purified by gel filtration chromatography. Attempts at crystallizing the entire protein using a multitude of commercially available conditions from Hampton Research (Crystal Screen 1 & 2, PEG/Ion 1 & 2, Salt Rx 1&2 and Index) at 4°C and 25°C employing the hanging drop method failed to produce good quality crystals but instead resulted in needle clusters of crystals, which were unsuccessfully improved by screening around their conditions.

Alongside ameliorating crystals of full-length RBEI, we looped out the leader sequence (65 residues) at the N-terminus of the protein and experimented with trypsin digestion after its purification; however, all trials of digestion resulted in several bands, while the desired one was never predominant. Therefore, we decided to insert in the COOH-terminal domain, just after residue 702, the TEV protease cutting sequence, and cleaved the purified protein with the highly specific TEV cysteine-protease commonly used for controlled cleavage. The product truncated at the C-terminus is then obtained after further purification and is referred to as RBEI core domain ( $\text{BEI}_{\Delta\text{C}}$ ). We attempted crystallization of both mature and core domain RBEI and succeeded with both; however, the mature protein crystals had very weak diffraction, which became worse upon soaking with M12. On the other hand, crystals of the core domain showed good diffraction that was sustained when soaked with dodecaose (M12). We present here the crystal structure of rice  $\text{BEI}_{\Delta\text{C}}$  in complex with M12 refined to a resolution of 2.35Å.

### 4.3.2 Structure determination

The crystal structure of RBEI, soaked with M12 was determined by molecular replacement using the rice BEI<sub>ΔC</sub> structure previously determined (PDB ID: 3AMK) as a search model and data was refined to 2.35 Å. The data was complete to 95.24%, and refined to an  $R_{\text{work}}$  of 15% and  $R_{\text{free}}$  of 22%. **Table 4.2** shows the x-ray diffraction data collected from a single crystal at the Advanced Photon Source (Chicago, IL), LS-CAT 21-ID-G beamline. The diffraction images were processed using HKL2000.<sup>60</sup> The data was further refined using CCP4 (**Table 4.2**).<sup>61, 62</sup> The structure shows one molecule in the asymmetric unit and belongs to space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>. The present form of the protein encompasses a total of 702 residues; but shows full density for residues 9-694 and some weak density for residues 6-8, which were left out of the refinement. One loop region, Gln469-Gly473, had weak density in previous BEI<sub>ΔC</sub> structures,<sup>50, 53</sup> but is completely visible in the current structure, including residues H467 and D468, therefore allowing us to model in a sugar molecule extending from the current M12 to the active site (*Detailed binding sites*, section 4.3.3.2). Only one loop in to the C-terminal domain encompassing residues 662-672, has weak to no density as was previously seen in the apo structure.<sup>50, 53</sup>

<i>Data collection</i>	
Wavelength (Å)	0.976
Unit-cell parameters	a = 47.670, b = 80.107, c = 182.716
Space group (Å)	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Resolution range (Å)	40-2.3 (2.34-2.30)
Completeness (%)	89 (57.9)
$R_{\text{merge}} (I)^a$	0.048 (0.383)
$\langle I \rangle / \langle \sigma_I \rangle$	18.88 (2.3)
<i>Refinement</i>	
Resolution range (Å)	39.97 - 2.35
$R$ -factor <sup>b</sup>	0.1503 (0.24)
$R_{\text{free}}^c$	0.2234 (0.34)
Completeness (%)	95.24 (72)
Root mean square deviation angles	1.5919
Protein atoms	5451
Oligosaccharides atoms	189
Water molecules	557

<sup>a</sup> $R_{\text{merge}} = \sum \sum_i |I_i - \langle I \rangle| / \sum |I|$ , where  $I_i$  is an individual intensity measurement and  $\langle I \rangle$  is the average intensity for this reflection, with summation over all data.

<sup>b</sup> $R$ -factor:  $\sum ||F_o| - |F_c|| / \sum |F_o|$ .

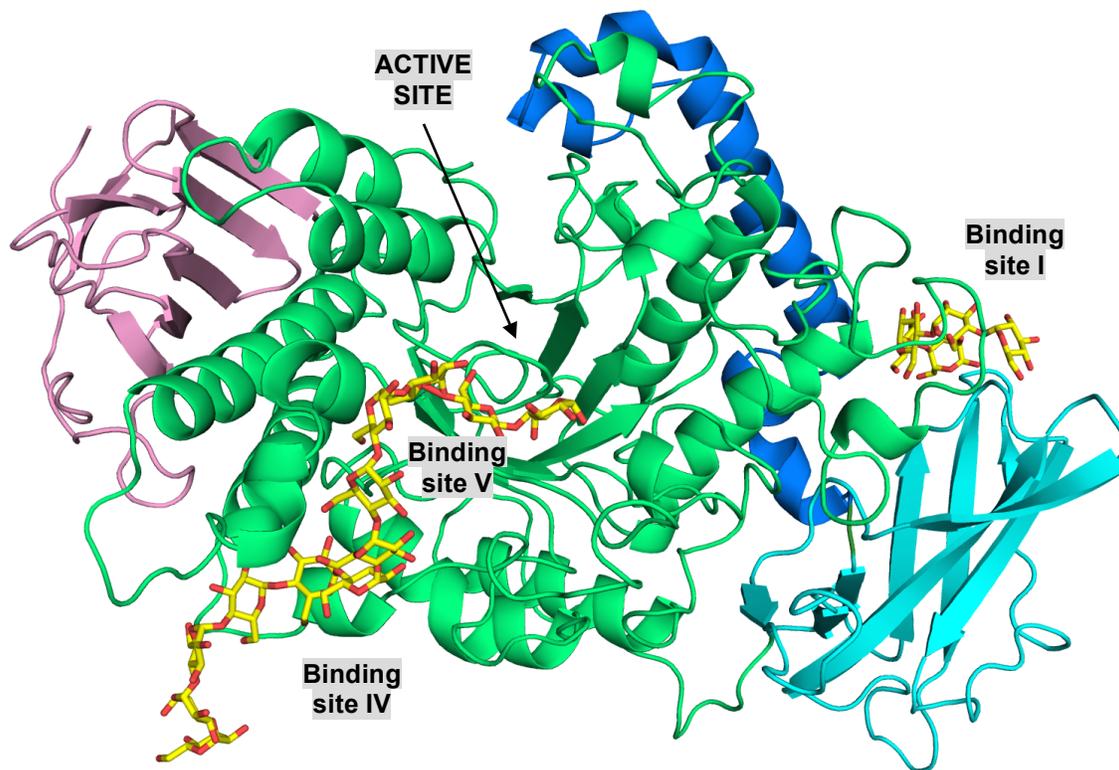
<sup>c</sup> $R_{\text{free}}$  :  $R$ -factor of 10% reflections removed before refinement.

**Table 4.2: Data collection and refinement statistics**

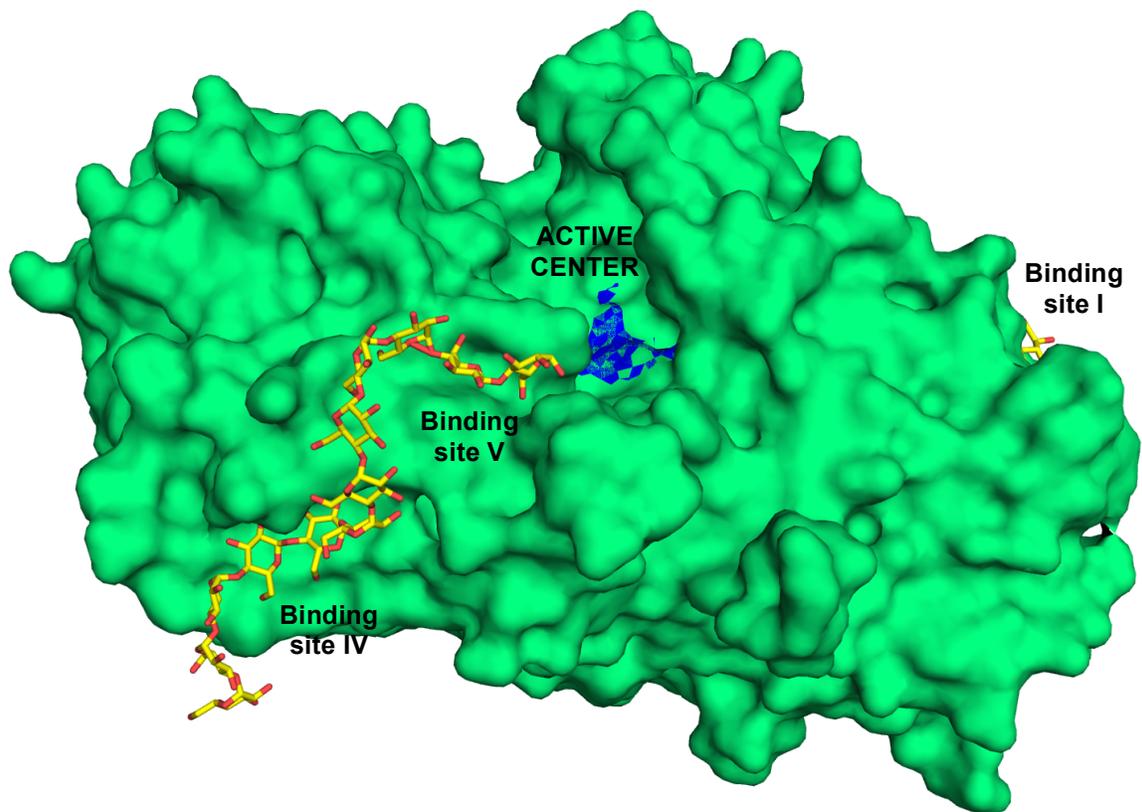
### 4.3.3 Overall structure

The structure of RBEI soaked in dodecaose showed two molecules of M12 bound to one enzyme molecule. One carbohydrate shows density for five glucose units, binding in site I; the other, shows density for all twelve glucose units, and reaches the opening of the groove-like active site domain and binds to sites IV and V. The overall structure is divided into four domains (**Figure 4.1**):

- N-terminal domain (in blue, residues 9-58): formed of three consecutive  $\alpha$ -helices, the second one constituting most of this domain, spans residues 18-45. These elements of secondary structure are connected by short loops, and make direct contact with the helices of the catalytic center on one side.
- CBM48 domain (in cyan, residues 59-160): consists of nine antiparallel  $\beta$ -strands connected by relatively long loops of 10-12 residues. This  $\beta$ -sandwich like fold is commonly observed in several starch-binding proteins.<sup>55, 63</sup>
- Catalytic (central) domain (in green, residues 161-587): described previously and common to all GH13 family enzymes, consisting of 8 pairs of alternating  $\alpha$ -helices  $\beta$ -strands keeping the sheets towards the inside of an almost cone-shaped active center, where the substrate binds.
- C-terminal domain (in pink, residues 588-702): forms another  $\beta$ -sandwich domain. Just like the N-terminal domain, it comes in close contact with the catalytic domain on the opposite side of the N-terminus'.



**Figure 4.1: Overall structure of RBEI in complex with M12.** The N-terminal domain is shown in blue, CBM48, cyan, center catalytic domain, green, and the C-terminal domain in pink. Carbohydrates are represented in sticks and colored by atom type, carbons in yellow and oxygens in red. The dodecaose, M12, binds exclusively into the catalytic domain and hangs over the catalytic groove without reaching inside (sites IV and V), while the five glucose units visible for the second site (site I), are sandwiched between three domains, the N-terminal, the carbohydrate binding module, and the catalytic domain.



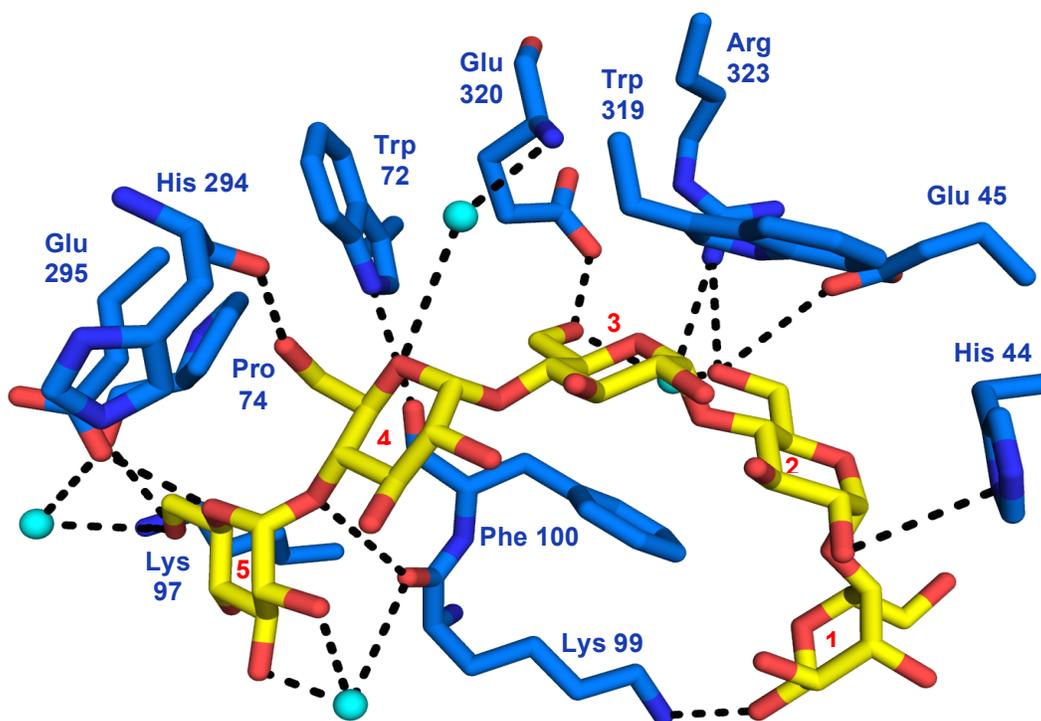
**Figure 4.2: Surface depiction of RBEI in complex with M12.** A prominent groove almost separates the molecule into two globular domains. At the center of this groove resides the active center, whose residues involved in catalysis are shown in blue. These catalytic residues, Y235, D270, H275, R342, D344, E399, H467 and D468 according to RBEI sequence numbering,<sup>50</sup> were predicted based on biochemical and structural data of  $\alpha$ -amylase and CGT.<sup>64, 65</sup>

#### 4.3.3.1 Binding site I

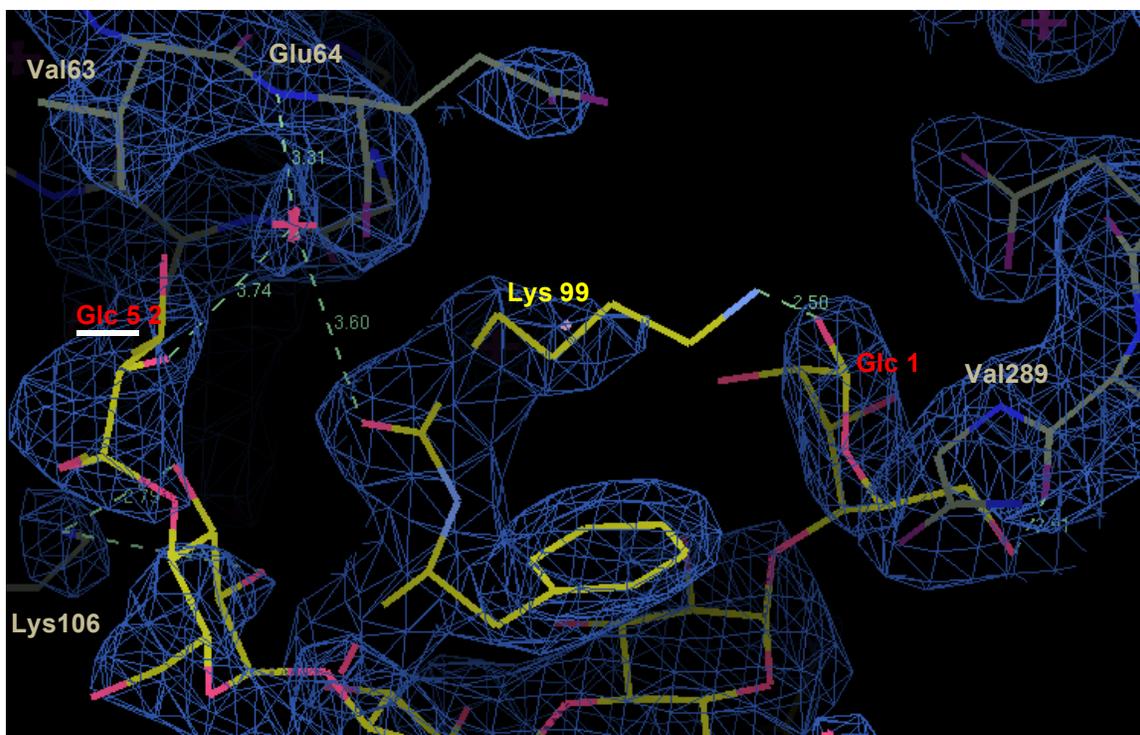
Five units of glucose are visible in binding site I (**Figure 4.3**). The saccharides, numbered 1 through 5 starting from the reducing end, make several hydrogen-bond interactions with the protein. Oxygen 1 of glucose (Glc) 1 binds to the side chain of Lys99, which also extends towards the non-reducing end of the glucan and makes a hydrogen bond to the main chain carbonyl to O4 of Glc4. Glc2 O2 makes an H-bond to His44, and its O6 makes hydrogen bonds to both Glu45 and Arg323 side chains. Another hydrogen bond is also made between O6 of Glc3 and the side chain of Glu320. Glc4 also forms a hydrogen bond between O6 of the sugar and the main chain carbonyl of His294. The first glucose on the non-reducing end, numbered 5 here, uses two consecutive oxygens, O5 and O6, to make hydrogen bonds to the Glu295 and Lys97 side chains, and to the Pro74 main chain oxygen carbonyl, respectively. A few hydrophobic interactions are also seen in this site, such as Pro74 interacting with Glc5 and Trp72 with Glc4. More interestingly, Trp319 and Phe 100 stack against the carbohydrate on opposite sides as if to sandwich glucoses 2 and 3. We expect these 2 residues to be critical for the protein's activity based on data on *E. coli* branching enzyme that showed all tryptophan stacking interactions with sugars significantly affected the activity (Chapter 2).

A handful of water molecules showing strong positive density are located near the glucose units, adding to the interactions between the protein and the oligosaccharide. This is exhibited in instances like water 371 connecting Glc3 to Arg323, water 195 hydrogen bonding to both Glc4 and Glu320, and water 444

mediating an interaction between Glc5 and Glu295. A unique interaction that water molecule 490 makes in this site connects the carbohydrate (Glc5) to Lys99 and to symmetry related molecule (**Figure 4.4**). It is worth mentioning that glucoses 1, 4 and 5 all interact with the symmetry related molecule with both hydrogen bonding (Glc1 to Val 289 and Glc4 to Lys106) and hydrophobic interactions (Glc5 to Val63) (**Figure 4.4**).



**Figure 4.3: Binding site I: detailed interactions between the sugar and RBEI.** The protein atoms are colored by type: C in marine, O in red and N in dark blue. M12 atoms are also colored by type with C in yellow and O in red. Glucose units are numbered in red. Hydrogen bonds are shown by dotted black lines. Water molecules interacting in this site are represented by spheres and colored cyan.



**Figure 4.4: Oligosaccharide binding to RBEI and its symmetry related molecule.** RBEI atoms are colored by type: C in yellow and rest as in figure 4.3. M12 atoms and glucose units' numbers are also colored as in figure 4.3. The symmetry related molecule's carbons are shown in beige as well as their numbering, O and N are colored just a nudge darker than the rest of the molecule. Hydrogen bonds are shown by dotted green lines. Water molecule 490, represented in a pink star, has an obvious positive density and binds to Lys99 of RBEI, Glc5 and Glu64 of symmetry molecule. The rest of the interactions with the symmetry mate is also shown: Glc1 hydrogen binds to Val289, Glc4 to Lys106, and Glc5 makes hydrophobic contact with Val63. The figure shows the  $1.0 \sigma$   $2F_o - F_c$  electron density map of RBEI around M5.

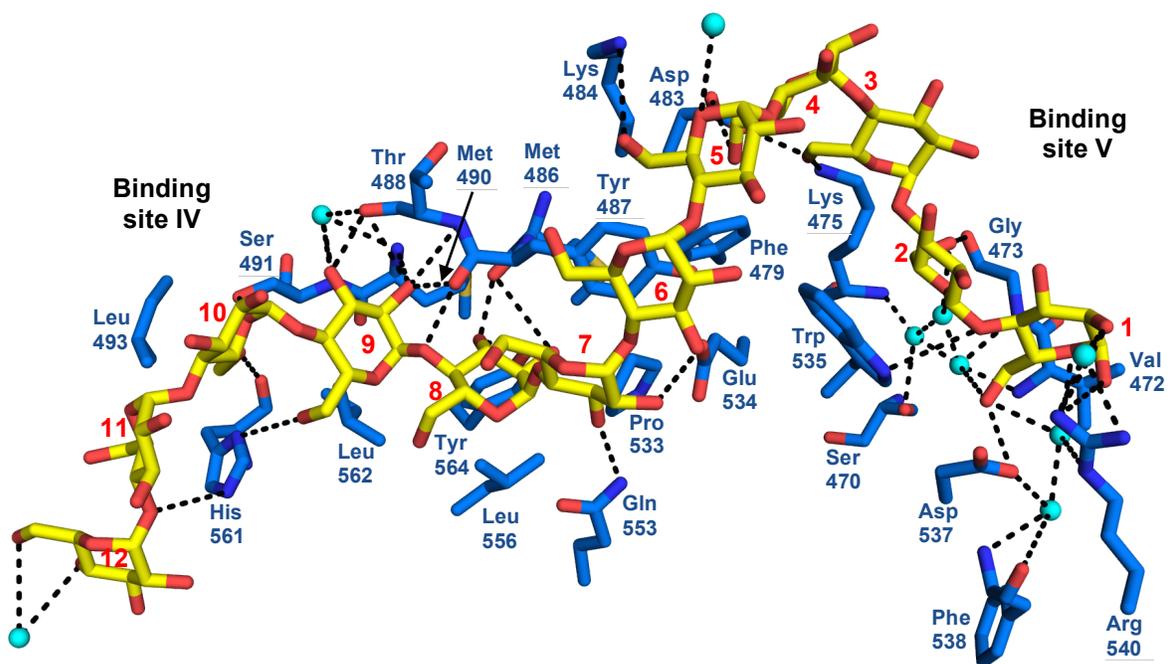
#### 4.3.3.2 Binding sites IV and V

Density for the second oligosaccharide molecule binding to RBEI is clearly visible in total and shows multiple interactions with the protein and with an adjacent symmetry mate as well. Every glucose unit in this carbohydrate makes interactions with the residues of RBEI except for the last glucose on the non-reducing end (Glc12). Binding sites are numbered IV and V starting away from the active. We split M12 into two binding sites of six glucoses each first because six consecutive glucoses form a full turn of amylose helix, and second because we see the number of interactions decline at Glc6 compared to the other glucoses. Several hydrogen bonds are seen in this site. Glc12 is the last sugar in M12, located the furthest away from the active site. It shows only two hydrogen bonds to water molecule number 492. At close proximity to this sugar resides a residue that extends alongside the carbohydrate, His561, whose main chain makes a hydrogen bond with O6 of Glc10 and whose side chain makes hydrogen bonds with O4 of Glc 11 and O6 of Glc9. Oxygens O5 and O6 of Glc10 also bind to the side chain of Ser491. O2 of Glc 9 makes hydrogen bonds with several main chain moieties, the carbonyl of Tyr487, nitrogen of Met490, and carbonyl and nitrogen of Thr488; the latter residue's main chain carbonyl also interacts with O3 of Glc9. A water molecule (number 28) interacts with O3 of Glc9 and three residues' main chains, Thr488, Met490 and Ser491. In the same binding site, O4 of Glc8 also makes hydrogen bonds to the main chain carbonyl of Tyr487, and both O2 and O3 of Glc8 bind to each of the side chains of Tyr564 and the main chain carbonyl of Met486. In Glc7, O2 and O3 bind to the side

chains of Glu 534 and Gln553, respectively. Additionally, hydrophobic interactions are seen in binding site IV. Leucines 493, 562 and 556 interact with Glc11 and 10, Glc 9 and Glc 8, respectively. A proline residue, 553, is also in close proximity to Glc 8 and 7.

Binding site V shows yet more interactions especially for the glucose units closer to the active site. Multiple hydrogen bonds can be visualized here, such as between O6 of Glc5 and the side chain of Lys484, oxygens 6 of both Glc4 and Glc3 and the side chain of Asp483, O6 of Glc3 and the Lys475 side chain, and O6 of Glc2 with the main chain carbonyl of Gly473. Near Glc5, water molecule 506 connects the O5 of this glucose to a symmetry mate; this bond is however one of many present between M12 and the neighboring molecule. Going back to the first two glucoses by the active center, there is a triad of water molecules with a clearly positive density closest to Glc2, where both O5 and O6 bind to water 554, which itself binds to waters 521 and 369. Those two waters, in turn, make two other hydrogen bonds: water 521 binds to the main chain carbonyl of Ser470 and the nitrogen of Lys475, and water 369 binds to the main chain nitrogen of Val472 and to O6 of Glc1, which also binds to the side chain of Asp537. Another net of hydrogen bonds is seen here: water molecule 556 binds to O5 and O6 of Glc1 connecting them to the side chain nitrogen NE of Arg540 and to H<sub>2</sub>O-141. The latter water molecule, in turn, binds to the side chain of Asp537 and both nitrogen and carbonyl main chain of Phe538. The interactions of Glc1 also extend to both terminal side chain nitrogens of Arg540 by means of O1 and O2. Stacking hydrophobic ring interactions with the sugar are also featured in this

site, such as those seen between Trp535 and glucose units 2 and 3, and Tyr487 and Glc6. One other hydrophobic residue, Phe479, aligns parallel to the sugar as to stack against it but is not closely facing it to provide some support. A last interaction is noted between Val 472 and Glc1.



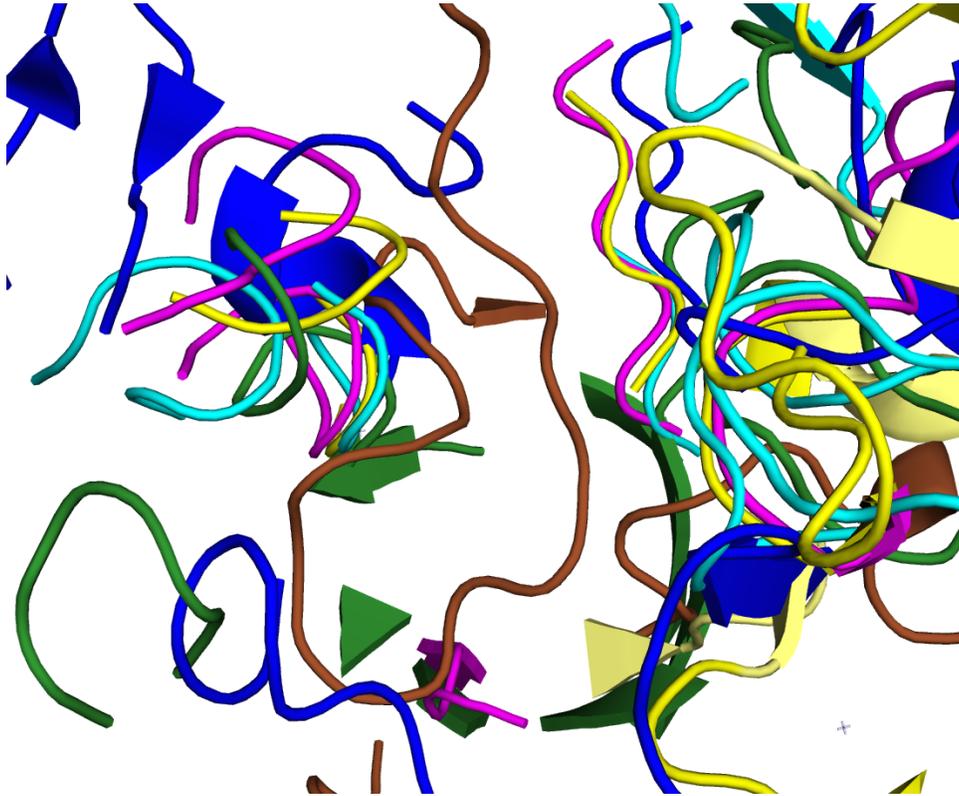
**Figure 4.5: Binding sites IV and V: detailed interactions between the oligosaccharide and RBEI.** Numbers for glucose units, M12 and the protein atoms are colored as in figure 4.3. Hydrogen bonds and water molecules are also represented as in figure 4.3.

### 4.3.4 Structural comparison

#### 4.3.4.1 To GH13 family enzymes

While the members of the GH13 enzymes share the TIM barrel domain structure, home of the protein's activity, they differ in their specific mode of binding to the substrate. Comparing the current structure of RBEI to the reported  $\alpha$ -amylase family structures bound to glucans in the active site reveals that the interactions in these proteins at the opening of the active site on the side where the substrate normally enters tremendously exceed those that RBEI makes.<sup>66-78</sup> Branching enzyme differs mainly in the number and length of loop structures just outside of the active site (**Figure 4.6**). To visualize this difference, RBEI was overlaid with pig pancreatic  $\alpha$ -amylase (PDB 1UA3) and *Halothermothrix orenii* alpha-amylase (PDB 3BC9). Both of these structures contain loops making direct contact with hydrogen bonds and hydrophobic interactions with their substrates, loop 138-164 in the former, and loop 380-392 in the latter, for which counterparts in RBEI are much shorter, 360-372 and 399-403, respectively. Two other examples, *Bacillus cereus*  $\beta$ -amylase (brown in Fig. 4.6) and maltohexaose producing amylase (yellow in Fig. 4.6) surround their substrate with four loops in each case, for which equivalents in RBEI are either much shorter or absent. *Escherichia coli* branching enzyme is also similarly quite open around the active site, like RBEI. Though the branching enzyme is different in this region, the active site residues are conserved. This implies the unique role of both branching ( $\alpha$ -1,4

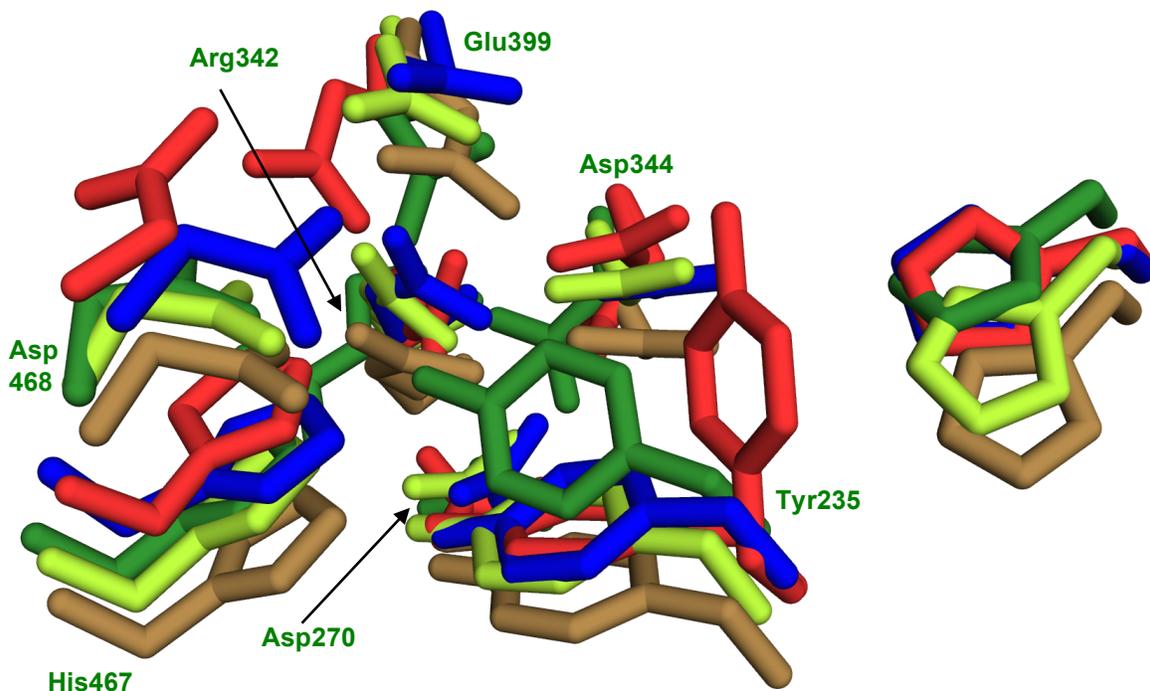
bonds) and transglycosylation (forming  $\alpha$ -1,6 bonds) that BE plays among the GH13 family of enzymes.



**Figure 4.6: Loops around the active site in GH13 family enzymes.** RBEI is represented in green cartoon. For clarity, only some of the GH13 family structures bound to glucans in the active site are shown here. Those proteins with longer or non-existing loops counterparts in RBEI are depicted for each of pig pancreatic  $\alpha$ -amylase (PDB 1UA3, blue),<sup>73</sup> *Halothermothrix orenii* alpha-amylase (PDB 3BCD, magenta),<sup>76</sup> maltohexaose-producing amylase (PDB 1WPC, yellow),<sup>67</sup> *Thermoactinomyces vulgaris* R-47  $\alpha$ -amylase II (PDB 1JIB, cyan),<sup>78</sup> and *Bacillus cereus*  $\beta$ -amylase (PDB 1B9Z, brown).<sup>71</sup>

#### 4.3.4.2 Active site residues

Eight residues involved in catalysis are conserved among the members of the GH13 family of enzymes despite the structural differences around the active site.<sup>2, 79, 80</sup> In **Figure 4.7**, these residues from several enzymes,  $\alpha$ -amylase (PDB 1UA3),<sup>73</sup> isoamylase (PDB 1BF2),<sup>81</sup> pullulanase (PDB 2E8Z),<sup>82</sup> and cyclodextrin glucanotransferase (PDB 4JCM),<sup>83</sup> were overlaid. The numbering in the figure follows the RBEI sequence. Asp270, located at the exit side of the groove-like active site domain, does not make direct interactions with the substrate in the bound structures, indicating a conformational role, such as holding the residues interacting with the glucan. Asp344 and Glu399 are essential for catalysis in BE.<sup>84</sup> In addition, when His275, Asp344, Glu399, and His467 were individually replaced by Ala, the protein became inactive towards amylose.<sup>52</sup> Based on these facts, Kumar *et al.* suggested that Asp344 and Glu399 play the respective roles of a nucleophile and a general acid/base catalyst, while the other two residues help stabilize the transition state during the reaction.<sup>85</sup> This can be verified structurally since D344 and E399 are oriented differently in RBEI than their equivalents in the other proteins, while H275 and H467 conserve their orientation. Similarly, Tyr235, Arg342, and Asp468, which adopt analogous orientation to their counterparts, probably play structural roles in the protein binding to the substrates.



**Figure 4.7: Overlay of active site residues in GH13 enzymes.** Proteins are colored as follows: RBEI in green (numbering is according to RBEI sequence), pig pancreatic  $\alpha$ -amylase (PDB 1UA3, blue), <sup>73</sup> *Pseudomonas* isoamylase (PDB 1BF2, lime), <sup>81</sup> *Bacillus subtilis* pullulanase type I (PDB 2E8Z, sand), <sup>82</sup> and *Alkalophilic Bacillus Clarkii* cyclodextrin glucanotransferase (PDB 4JCM, red). <sup>83</sup>

#### 4.3.4.3 Carbohydrate binding module

The target glycosidic bonds in polymeric saccharides are usually not exposed to the active site of the carbohydrate active enzymes. Carbohydrate Binding Modules (CBMs) play the essential role of promoting the association of these target enzymes to their insoluble substrates. <sup>86</sup> CBMs were previously classified based on amino acid sequence similarity. <sup>87</sup> In order to conform to the glycoside hydrolase families' classification, the CBM groups are now organized by families and numbered in Arabic numerals as well. The CBM families have in some cases similar binding specificity, help identify functional residues or predict polypeptide

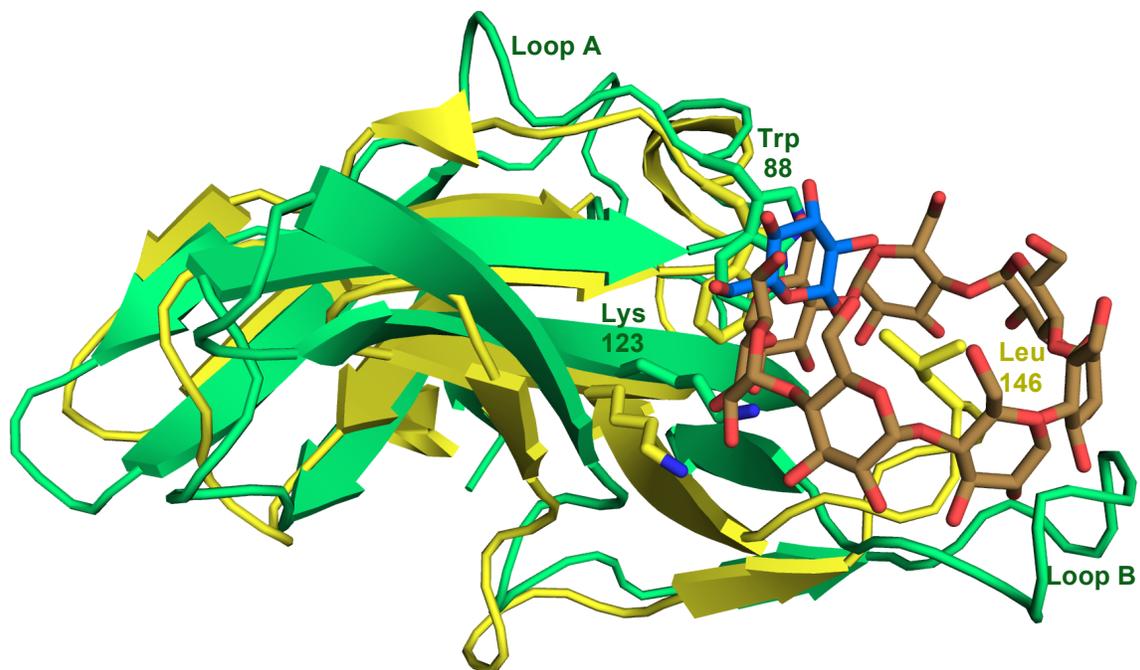
folds (<http://www.cazy.org/Carbohydrate-Binding-Modules.html>). Currently, there are 71 CBM families that differ in their ligand specificity. Characterized CBM families contain proteins that bind to insoluble storage saccharides like starch and glycogen. In the last almost 30 years, three-dimensional studies on structures of CBM members provided a wealth of invaluable information on the biological functions of these proteins; in addition, the elucidation of CBMs in complex to oligosaccharides gave beneficial insight into substrate recognition by corresponding proteins. CBMs are also classified into 3 major types: A, B, and C. Type B, which CBMs of BE belongs to, is known as a chain binder because it binds oligosaccharides of at least 4 glucose units. Type B CBMs is also characterized by binding to several subsites, implying a topography that is capable of interacting with individual glucan chains rather than crystalline surface, counting therefore key orientations of specific aromatic residues and on direct hydrogen bonds.<sup>63</sup> Several CBM20 family enzymes have been analyzed excessively;<sup>55, 71</sup> two carbohydrate binding sites were identified; they bind raw starch and direct linear starch towards the active site.<sup>88</sup> The carbohydrate active enzymes (CAZy) classifies BE a member of the GH13 family of enzymes according to its amino acids sequence,<sup>89</sup> and its carbohydrate binding module as CBM48; this was later verified by crystal structure analysis.<sup>50</sup> The CBM48 domain is formed of 100 residues and makes a part of the entire N-terminal domain, which was shown to be responsible for binding raw starch and leading it to the active site.<sup>76, 90</sup> The x-ray crystal structure of AMP-activated protein kinase (AMPK)  $\beta$ -subunit in complex with  $\beta$ -cyclodextrin (PDB 1Z0M) was determined,

showing a carbohydrate binding pocket resembling CBM domains and was recognized as CBM48 based on its sequence.<sup>91</sup> We overlaid this structure with the current RBEI CBM48, residues 59-160, and noticed an almost perfect superposition between the two structures, with only minor differences (**Figure 4.8**). CBM48 forms a  $\beta$ -sandwich of nine alternating  $\beta$ -sheets overall connected by loops, except in between the second and third  $\beta$ -strands in AMPK where a short  $\alpha$ -helix crosses the connecting loop. In RBEI-M12, a loop connecting  $\beta$ 5 and  $\beta$ 6 (loop A in **Figure 4.8**) encompasses residues 108-119 while its counterpart in AMPK is formed of only four residues (118-121). Another structural difference is in the N-terminus of this domain between the two proteins: AMPK  $\beta$ 1 is parallel to  $\beta$ 2 of RBEI,  $\beta$ 2 of AMPK to  $\beta$ 3 of RBEI, and so on, which causes the last  $\beta$ -sheet of AMPK,  $\beta$ 9, to adopt an antiparallel conformation relative to  $\beta$ 1 of RBEI.

Even though binding of  $\beta$ -cyclodextrin (CD) to AMPK happens at a position where a CD could not bind in RBEI because of a loop more extended in RBEI (145-156, loop B in **Figure 4.8**) versus AMPK (143-149), some residues involved in the binding are conserved between the two structures. For instance the Leu146 side chain centers the cyclodextrin structure in AMPK, its counterpart Phe151 could be making a similar or stronger hydrophobic stacking interaction with a sugar provided the carbohydrate binds with a slight variation in position to avoid a clash. Three other residues are also analogous or partially conserved between the two structures, Gln124, Trp133, and Asn150 in AMPK, correspond to Asn121, Asp135, and Asp156 in RBEI, respectively (the residues mentioned

so far are not shown in **Figure 4.8** for clarity). Finally and most importantly, two residues, W88 and K123 (RBEI numbering), are conserved among all BEs as well as in AMPK CBM48. In fact, a single glucose is seen in RBEI-M5 (PDB 3VU2) in this site (called site 2 there) in the most convenient position for stacking against the conserved Trp in this domain.<sup>53</sup> This Glc also makes interactions with D135 and D156, the equivalent of those that interact with the cyclodextrin in the AMPK complex. In order to shed some light on the role of each of W88 and K123 in RBEI, we mutated these residues in the mature protein (missing the leader sequence at the N-terminus necessary for transport) individually into Ala and compared their activity to the native mature BEI. One would expect that residues conserved among all the branching enzymes to have a detrimental effect on the activity and this was seen for the W88A mutation, which rendered the protein completely dead (0% activity). The activity of the K123A was partially lost with 58% compared to the wild type. Clearly the sugar-protein stacking interactions are essential for the protein's branching activity similarly to what we saw in EcBE-M7, where in every binding site, the loss of a Trp interaction with the sugar resulted in a significant decrease in activity, with the exception of site II, which turned out not to be critical for the protein's activity. Perhaps other surrounding residues in K123A can compensate for the hydrogen bonds that K123 makes with the substrate. In the presence of a dodecaose, binding in the CBM48 domain of RBEI was not seen, while the structure of RBEI-M5 shows a single glucose at this position. In fact, the crystal packing in our RBEI-M12 obstructs binding completely at this site. Soaking the crystals repeatedly resulted

in their dissolution until we reached a minimal soaking time enough for observing oligosaccharide. It is very likely that longer times soaking (over 5 hours) destabilized the crystal lattice resulting in crystal dissolution. This makes us believe that soaking is not the ideal method to visualize what happens in natural binding but it was the only way to produce well-diffracting crystals of the complex; however, dodecaose is not only highly insoluble but is also commercially unavailable.

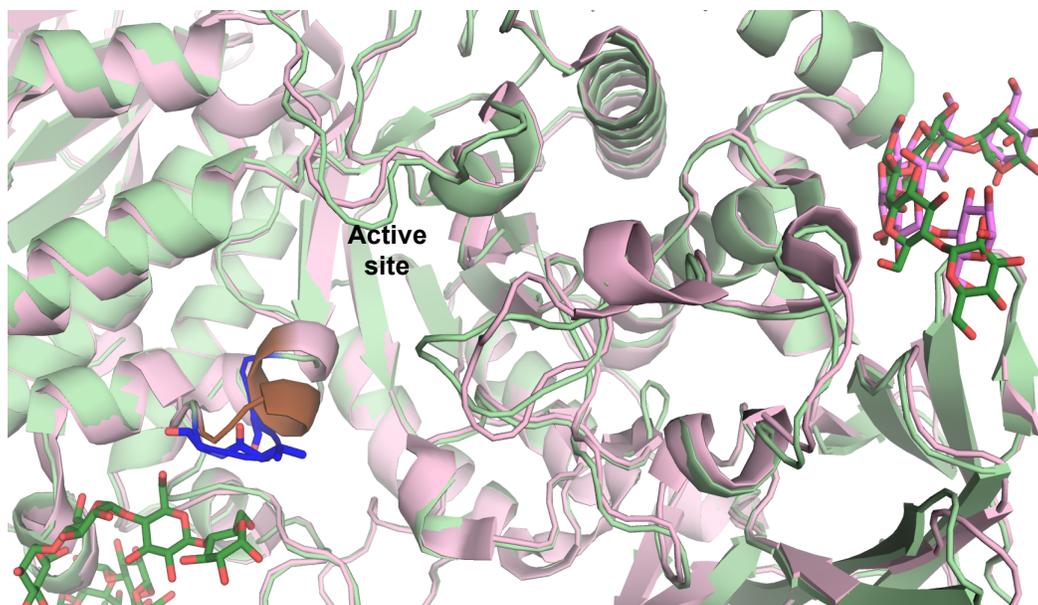


**Figure 4.8: Overlay of CBM48 domains of RBEI-M12 and AMPK-beta-CD.**<sup>91</sup> RBEI cartoon and conserved residues W88 and K123 are shown in green with those of AMPK in bright yellow. Beta-CD is colored brown. All stick representations are colored by atom type, with O red, and N blue. One glucose unit with C colored in blue is the sole glucose seen at this position in the RBEI-M5 structure. This single glucose stacks against W88 and hydrogen bonds to K123, the two conserved residues in all BEs.

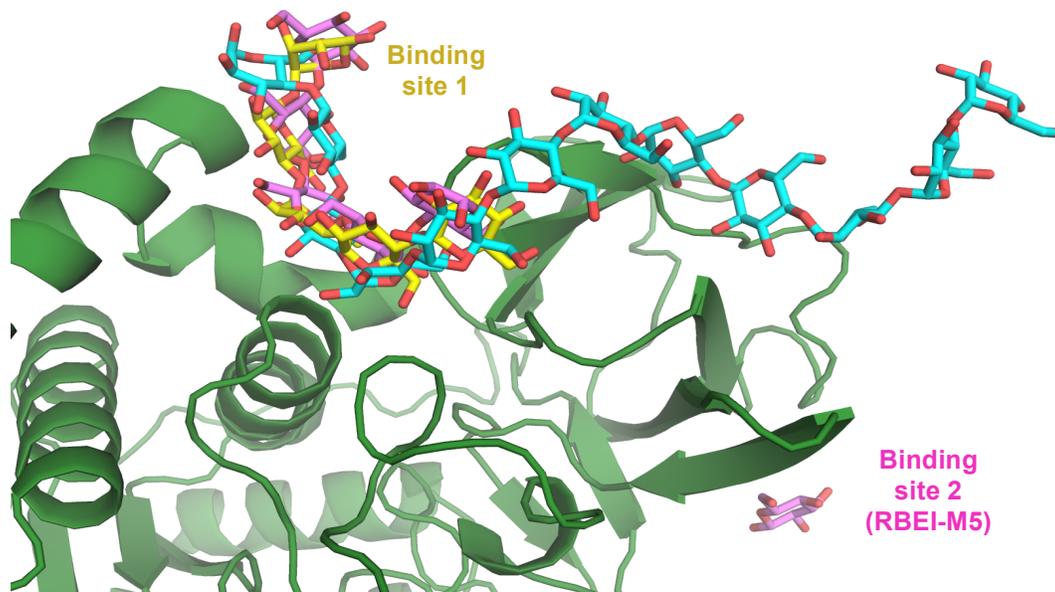
#### 4.3.4.4 Binding in site I of RBEI-M12 compared with RBEI-M5

The current rice branching enzyme I confirms once again the necessity of the CBM48 domain for carbohydrate binding and the cooperation with the N-terminal domain to accomplish this function. The maltopentaose occupying binding site 1 interacts with ten residues from the CBM48 domain and two residues exclusive to the N-terminal helices. RBEI-M5 also shows the five glucose units of a maltopentaose only in site 1, in a location identical to our structure (**Figure 4.9**), interacting with all the same residues with the exception of one, K97, which makes a hydrogen bond in RBEI-M12 with O6 of the non-reducing end sugar. Although the two maltopentaose-molecules share identical locations, it is ambiguous from our electron density map to deduce the exact orientation of M5 especially after several refinement rounds (**Figure 4.9**). Binding of the carbohydrate in the same site demonstrates the strong affinity this site has for sugars. The CBM of *Aspergillus Niger* (A.) glucoamylase is one of the best-characterized starch binding domains and belongs to the CBM20 family.<sup>55</sup> The three-dimensional structure of A. glucoamylase was determined in complex with beta-cyclodextrins and shows binding of substrate in the two distinct sites, one is the same as M5 in RBEI (*site 2*), and the other corresponds to site 1 of RBEI-M5, where a single glucose is visible in RBEI versus all units in A. glucoamylase (*site 1*).<sup>55</sup> This shared binding site among the three structures discussed here does not only have stronger affinity towards carbohydrates but also acts to set two substrate strands apart or to localize loose amylose regions as previously suggested.<sup>54</sup> This could also explain the fact that only five glucose units are

visible in this site although twelve were available for binding and that opposite orientations are seen between the two molecules of maltopentaose. When long substrates approach the protein or as the reaction occurs in the natural process, the floppy amylose helix requires localizing, binding site I, providing partial (only a handful of residues) yet strong (repeated in several structures) binding, is probably responsible for this substrate confinement. On the other hand, no binding was seen near Trp88 and Lys123, even though a single Glc was visible in RBEI-M5 in this site, although the protein's activity was reduced to null with the mutation W88A. In other complexes, *A. glucoamylase*<sup>55</sup> and AMPK<sup>91</sup> for example, a beta-cyclodextrin occupies this site, interacting with both of the two conserved residues mentioned. This binding site was suggested to be responsible for raw starch initial recognition.<sup>54</sup> Sorimachi, K. *et al.* also believe it is a separate site from binding site I (RBEI), since the two cyclodextrins bound in their structures adopt 90° relative orientations.<sup>55</sup> To verify this idea, we have modeled the dodecaose molecule in binding site I of RBEI (**Figure 4.10**). We can clearly conclude that the two binding sites are perpendicular in direction preventing simultaneous binding by a single carbohydrate molecule. In addition, we notice that M12 could potentially bind in site I and extend to as in the model as it is surrounded by several beta-sheets that could support the substrate; however, only five glucose units bind emphasizing again the high affinity of the residues interacting with M5.



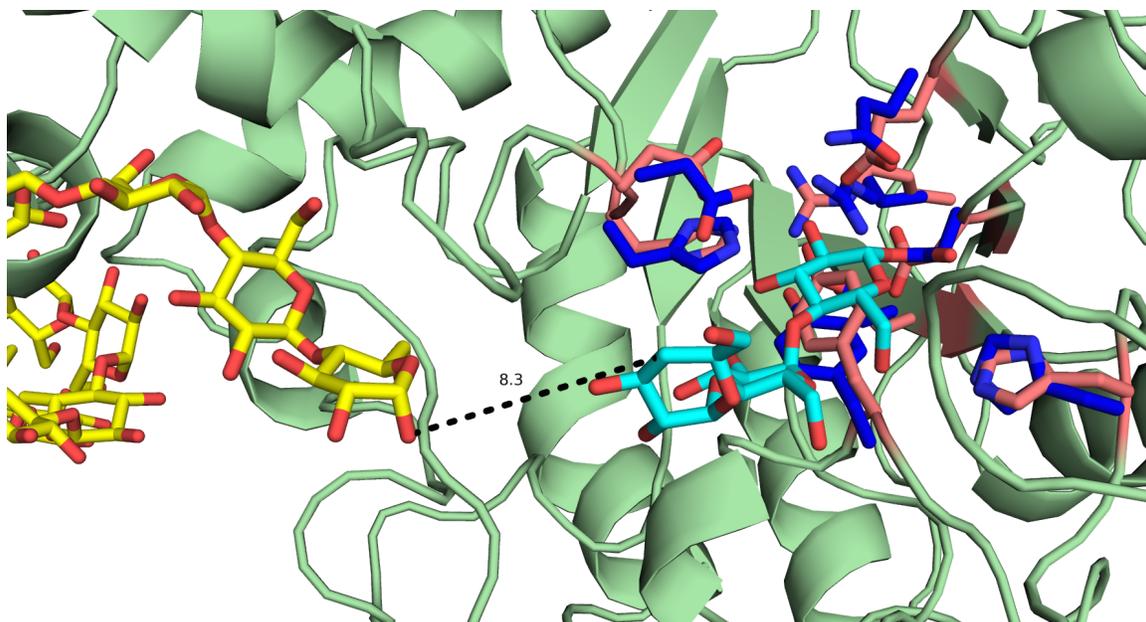
**Figure 4.9: Overlay of RBEI-M12 and RBEI-M5.** <sup>53</sup> RBEI-M12 is shown in light green, and its oligosaccharides in dark green, RBEI-M5 in light pink and corresponding carbohydrate in dark pink. The unordered loop in RBEI-M5 shown in brown is close enough to M12 in the current structure to become ordered and interact (colored in blue).



**Figure 4.10: Dodecaose molecule modeled in binding site I.** RBEI is shown in green cartoon, modeled molecule in sticks colored in cyan, M5 from RBEI-M12 in yellow and the RBEI-M5 oligosaccharides are shown in light magenta. All sticks representations are colored by atom type with O in red. Clearly, M12 molecule bound in site I is perpendicular in direction to site 2 of RBEI-M5.

#### 4.3.4.5 Binding in sites IV and V and comparison to EcBE-M8

In Binding site V, closest to the active site, a previously unordered loop, loop 469-473, does not only shows well ordered electron density in the present structure but also provides three sugar interacting residues, Ser470, Val472 and Gly473 (**Figure 4.9**). Despite positive density visualized for all twelve glucose units, binding in the substrate-binding active site could not be seen although BEI is known to transfer chains 6-15 units long.<sup>36</sup> In order to visualize how far the reducing end is from the active site, we overlaid the structure of CGT covalently bound to a triose intermediate in the active site with RBEI-M12 (**Figure 4.11**).<sup>92</sup> The residues that bind with sugars 1 and 2 inside the active site are conserved between the 2 proteins. The distance between sugar 3 – C4 and O1 of M12 is 8.3 Å, distance that fits two glucose units, although one could be sufficient depending on the orientation the entire carbohydrate adopts. Even though we see the reducing end of the sugar pointing towards the active site, M12 seems to have started binding at the non-reducing end as if the enzyme starts binding on the external sites and leads its substrate into the central groove for activity. This is in contrary to other bound structures excluding BE, where short oligosaccharides bind in the active site primarily even when external binding sites are unoccupied.<sup>66-78</sup>



**Figure 4.11: RBEI and CGT overlaid around the active site.** Residues of CGT covalently bound intermediate are shown in blue, RBEI cartoon in light green and overlaid residues in deep salmon. All residues are colored by element with O in red and N in dark blue. Residues shown in CGT are those interacting with the sugar bound. They overlap with their counterparts from RBEI. M12 of RBEI is shown in yellow and intermediate triose of CGT in cyan.

#### 4.3.4.6 Comparison with EcBE-M8

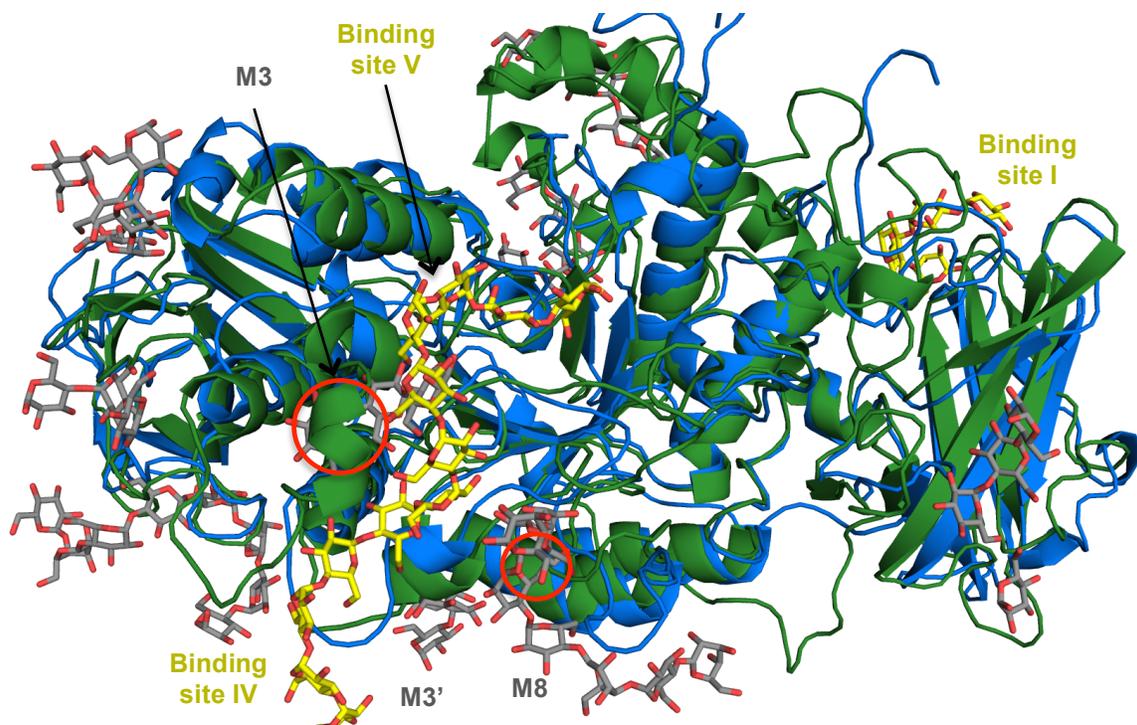
The glycogen branching enzyme (GBE) shares with SBEI several structural features including the four conserved regions,<sup>93</sup> and the TIM barrel domain,<sup>2</sup> both present in all alpha-amylase family enzymes. Despite this fact, *E. coli* BE and rice BEI are only about 23% identical.<sup>33</sup> We overlaid the structures of RBEI bound to M12 and that of EcBE in complex with M8 for comparison (**Figure 4.12**). Low sequence homology did not seem to prevent an overall similar topology with the only major difference in the N-terminal upstream of CBM48, where BEI forms three consecutive helices, one occupying most of this domain, while our truncated structure of EcBE misses this domain completely. One x-ray structure of a full-length bacterial BE, MycoBE, is available and shows the far N-terminal domain location away from the entire structure, and having a very similar topology to the CBM48 domain (more details on comparison of EcBE and MycoBE is found in chapter 3).

Binding, on the other hand, is quite unique for each enzyme. EcBE appears to communicate with the substrate using numerous external-binding sites spread on the protein's surface in all domains, while in BEI, binding is either in the N-terminal or central domain. This reality does not prohibit, however, binding of oligosaccharides in EcBE at near proximity of binding site IV and V of RBEI. Glucose units in binding site V cross over a maltotriose (M3) of EcBE and those of binding site IV run in an almost parallel orientation to the first four glucose units of a maltoheptaose (M8 in EcBE); both of these EcBE sugars direct their reducing ends towards the active site as in M12. The position of M3 in EcBE

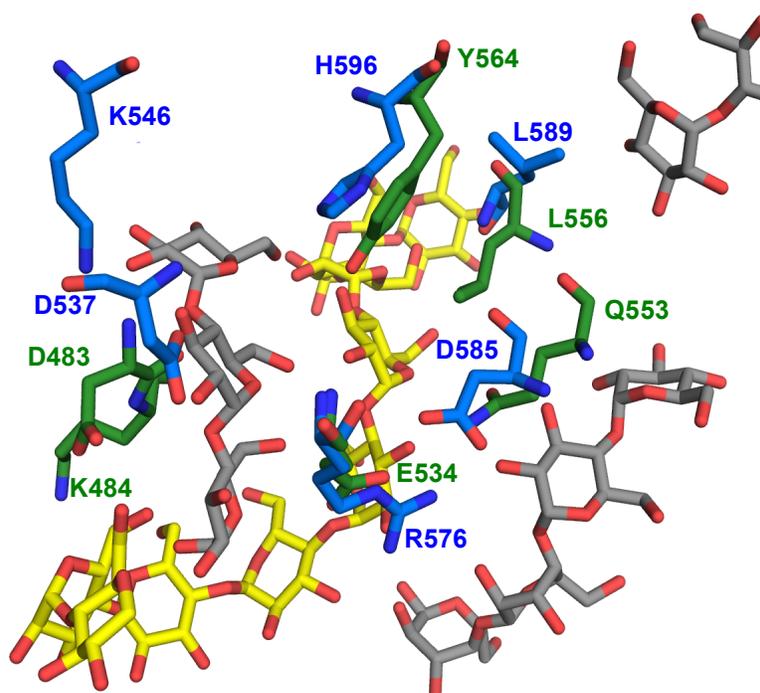
could not have been adopted in RBEI due to a clash with a loop ran by two short helices; as for M8, the case is different since only a single residue, located on an external loop in BEI, collides with one glucose unit in M8 (red circles in **Figure 4.12**). A second M3 of EcBE, labeled M3' in the figure, binds in between two molecules in the asymmetric unit of EcBE, Mol. A and Mol. B. Although this sugar orients itself in a nearly 90° angle compared to M12 (RBEI) and M8 (EcBE) (**Figure 4.12**), M3' shares, when binding to Mol. B, a similar interaction to that seen in M12 of BEI with Leu556.

Overlaying the two crystal structures also allowed us to compare binding residues in sites IV and V, revealing several conserved amino acids essential for binding the closest sugar to the active site to date available for either structure (**Figure 4.13.a**). Some of the conserved residues between the two structures overlap in location and orientation; these are, Gln553, Tyr564, and Glu534 of RBEI, and Asp585, His596 and Arg576 of EcBE, respectively. Another pair of residues interacting with the sugars, Leu556 (RBEI) and Leu589 (EcBE), have common locations but opposite side chain orientations (**Figure 4.13.a**). This is due to the differences in the topology in this area between the two proteins, depicted in **Figure 4.13.b**. The two loops in green (RBEI) and bright blue (EcBE) do not overlap and so is the case for their corresponding residues. A unique setting of conserved residues is that seen in Asp483 and Lys484 (RBEI) and their counterparts Asp537 and Lys546 (EcBE). The residues in this case do not share orientations pairwise but towards their carbohydrates (**Figure 4.13.a**). The

differences between the structures in this region are seen in **Figure 4.13.b**: RBEI loop in dark green and that of EcBE in dark blue.

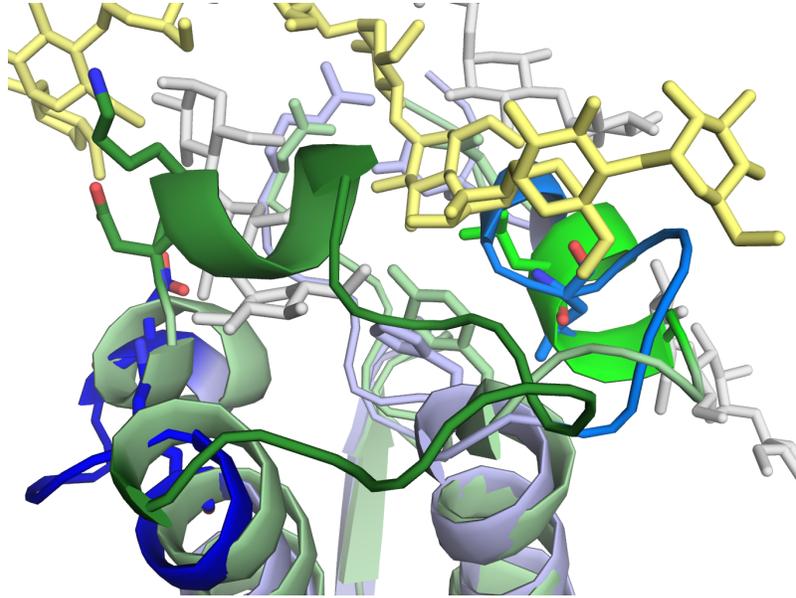


**Figure 4.12: Overlay of RBEI-M12 and EcBE-M8.** Oligosaccharides bound to each molecule are represented by sticks colored by type with O in red for both, C in yellow for RBEI-M12, and in grey for EcBE. Cartoon of RBEI is depicted in green and that of EcBE in blue. Only binding sites of BEI-M12 are labeled. Red circles represent areas of clash were BEI to show similar binding to EcBE.



**Figure 4.13: Comparison of RBEI and EcBE. a.** Conserved residues between RBEI and EcBE interacting in binding sites IV and V of RBEI. Coloring is the same as in figure 4.12. Conserved residues are represented, but only some superimpose between the two structures. K484, D483 and L556 of RBEI and their respective counterparts in EcBE, K546, D537 and L589 do not share orientations pairwise but towards their carbohydrates.

**Figure 4.13 (cont'd)**



**b.** Structural differences around the shared binding sites in RBEI and EcBE. RBEI is colored in light blue and EcBE in light green, and their carbohydrates in pale yellow and light grey, respectively. The loops containing conserved amino acids that do not overlap are shown in bright and dark green for RBEI, corresponding to those in bright and dark blue in EcBE, respectively.

#### 4.4 Conclusion and Future Research Plans

We have identified two new binding sites, IV and V, and reproduced binding site I, which is located in the N-terminal domain. In binding site I, we can not confirm the orientation of the sugar from the electron density map, which could mean that the sugar can bind in either direction (since it is shown in the opposite orientation in RBEI – M5) and possibly correlates with binding sites 2 and/or 3 of RBEI – M5 instead of directly interacting with the active site. It would be crucial to perform activity assays on point mutations of residues binding to each one of the sites, as well as combined mutations between sites to determine their relevance for activity. On the other hand, novel-binding sites IV and V in RBEI seem to show that the protein needs to bind its substrate to those sites before binding in the active site can happen. It is important to note that crystal packing does not block the closest sugar to the active site to allow another M12 to bind there, as Glc1 of M12 is over 6 Å away from side chain of mate Lys171, with all other distances over 9 Å away. So the ligand could have bound to the active site residues and emanate from there out; however, we see almost the opposite happening, since Glc12 that does not show direct interaction with the protein did not cause the oligosaccharide to move towards the active site, which confirms the necessity of the binding in site IV and V before entry into the active site of RBEI. In addition, the residues binding in these two sites have more affinity towards the carbohydrate than the active site amino acids. This is not surprising as the glycosidic bonds targets of the branching reaction are not reachable by

the reacting residues and require binding to the surface before directing into the active site.

Attempts at protein co-crystallization with a maltopentaose did not reveal any binding outside the N-terminal domain, despite using a high concentration of substrate. This proves weaker RBEI affinity to short sugars and suggests that soaking in shorter sugars will not identify more oligosaccharide binding sites. Overlaying with CGT confirmed that the reducing-end of M12 bound in the current structure is only 4 to 5 glucose units from binding to the active site residues, but authentic oligosaccharide binding might direct a 14 to 15 glucose-units sugar towards residues in the active site especially since Glc12 of M12 made little interaction with the protein. Whether a bit longer oligosaccharide than the present will finally be found bound in the active site or not could only be discovered by crystallizing the protein with an M14, M15 or M16, and this could only happen under two conditions: synthesizing the pure oligosaccharide required and surmounting carbohydrate dissolution problems (if soaking is tried).

Therefore for a shorter-term goal and in order to understand the effect of each residue on the protein activity and its involvement in determination of chain length transferred, it becomes important to perform a series of mutations starting with point mutations, and then after comparing results carrying double mutants of RBEI binding residues. The activity assays will reveal change in protein activity for each mutation and accordingly those residues causing moderate effect should be tested for chain length specificity. The following table (**Table 4.3**) lists these residues that would be crucial to assay by respective binding sites.

Binding site	Residues involved in binding
I	H44, E45, W72, P74, K97, K99, F100, W319, H294, E295, E320, R323
IV	H561, S491, Y487, T488, M490, Y564, M486, Q553, E534, L493, L562, L556, P533
V	K484, D483, K475, G473, S470, D537, F538, R540, F479, W535, Y487, V472

**Table 4.3: Residues involved in the interactions in binding sites I, IV and V of RBEI-M12.**

## REFERENCES

## REFERENCES

1. Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities., *Biochemical Journal* 280, 309-316.
2. Jespersen, H. M., MacGregor, E. A., Henrissat, B., Sierks, M. R., and Sevansson, B. (1993) Starch- and Glycogen-Debranching and Branching Enzymes: Prediction of Structural Features of the Catalytic ( $\alpha/\beta$ )8-Barrel Domain and Evolutionary Relationship to Other Amylolytic Enzymes, *Journal of Protein Chemistry* 12, 791-805.
3. Kuriki, T., Takata, H., Okada, S., and Imanaka, T. (1991) Analysis of the active center of *Bacillus stearothermophilus* neopullulanase, *Journal of bacteriology* 173, 6147-6152.
4. Takata, H., Kuriki, T., Okada, S., Takesada, Y., Iizuka, M., Minamiura, N., and Imanaka, T. (1992) Action of Neopullulanase: Neopullulanase catalyzes both hydrolysis and transglycosylation at  $\alpha$ -1,4 and  $\alpha$ -1,6-glycosidic linkages., *Journal of Biological Chemistry* 267, 18447-18452.
5. Baba, T., Kimura, K., Mizuno, K., Etoh, H., Ishida, Y., Shida, O., and Arai, Y. (1991) Sequence conservation of the catalytic regions of amylolytic enzymes in maize branching enzyme-I *Biochem. Biophys. Res. Commun.* 181, 87-94.
6. van der Maarel, M. J. E. C., van der Veen, B., Uitdehaag, J. C. M., Leemhuis, H., and Dijkhuizen, L. (2002) Properties and applications of starch-converting enzymes of the  $\alpha$ -amylase family, *Journal of Biotechnology* 94, 137-155.
7. Larner, J., Illingworth, B., Cori, G. T., Cori, C. F., 199, and 641-51. (1952) Structure of glycogens and amylopectins - II. Analysis by stepwise enzymatic degradation. , *Journal of Biological Chemistry* 199, 641-651.
8. Spilatro, S. R., and Preiss, J. (1987) Regulation of Starch Synthesis in the Bundle Sheath and Mesophyll of *Zea mays L.*: Intercellular Compartmentalization of Enzymes of Starch Metabolism and the Properties of ADP-Glucose Pyrophosphorylases, *Plant Physiology* 83, 621-627.
9. Imberty, A., Chanzy, H., and Perez, S. (1988) The Double-helical Nature of the Crystalline Part of A-Starch, *J. Mol. Biol.* 201, 365-378.
10. Manners, D. J. (1989) Recent Developments in Our Understanding of Amylopectin Structure, *Carbohydrate Polymers* 11, 87-112.

11. Manners, D. J. (1991) Recent Developments in Our Understanding of Glycogen Structure, *Carbohydrate Polymers* 16, 37-82.
12. Preiss, J., Ball, K., Smith-White, B., Iglesias, A. A., Kakefuda, G., and Li, L. (1991) Starch Biosynthesis and its Regulation, *Biochem Soc Trans* 19, 539-547.
13. Iglesias, A. A., and Preiss, J. (1992) Bacterial Glycogen and Plant Starch Biosynthesis, *Biochemical Education* 20, 196-203.
14. Martin, C., and Smith, A. M. (1995) Starch Biosynthesis, *Plant Cell* 7, 971-985.
15. Buleon, A., Colonna, P., Planchot, V., and Ball, S. (1998) Starch granules: structure and biosynthesis, *International Journal of Biological Macromolecules* 23, 85-112.
16. Ellis, R. P., Cochrane, M. P., Dale, M. F. B., Duffus, C. M., Lynn, A., Morrison, I. M., Prentice, R. D. M., Swanston, J. S., and Tiller, S. A. (1998) Starch Production and Industrial Use, *J. Sci. Food Agric.* 77, 289-311.
17. Morell, M. K., Samuel, M. S., and O'Shea, M. G. (1998) Analysis of starch structure using fluorophore-assisted carbohydrate electrophoresis, *Electrophoresis* 19, 2603-2611.
18. Frances, H., and Bligh, J. (1999) Genetic Manipulation of Starch Biosynthesis: Progress and Potential, *Biotechnology and Genetic Engineering Reviews* 16, 177-202.
19. Slattery, C. J., Kavakli, I. H., and Okita, T. W. (2000) Engineering starch for increased quantity and quality, *Trends in Plant Science* 5, 291-298.
20. Jobling, S. (2004) Improving starch for food and industrial applications, *Current opinion in plant biology* 7, 210-218.
21. Wang, L., and Wise, M. J. (2011) Glycogen with short average chain length enhances bacterial durability, *Die Naturwissenschaften* 98, 719-729.
22. Boyer, C. D., and Preiss, J. (1981) Evidence for Independent Genetic Control of the Multiple Forms of Maize Endosperm Branching Enzymes and Starch Synthases, *Plant Physiology* 67, 1141-1145.
23. Gao, M., Fisher, D. K., Kim, K. N., Shannon, J. C., and Gultinan, M. J. (1996) Evolutionary conservation and expression patterns of maize Starch Branching Enzyme I and IIb genes suggests isoform specialization, *Plant Molecular Biology* 30, 1223-1232.

24. Guan, H., Li, P., Imparl-Radosevich, J., Preiss, J., and Keeling, P. (1997) Comparing the Properties of *Escherichia coli* Branching Enzyme and Maize Branching Enzyme, *Archives of biochemistry and biophysics* 342, 92-98.
25. Guan, H. P., Baba, T., and Preiss, J. (1994) Expression of Branching Enzyme I of Maize Endosperm in *Escherichia coli*, *Plant Physiology* 104, 1449-1453.
26. Guan, H. P., Kuriki, T., Sivak, M., and Preiss, J. (1995) Maize Branching Enzyme Catalyzes synthesis of Glycogen-like Polysaccharide in GlgB-Deficient *Escherichia coli*, *Proceedings of the National Academy of Sciences in the United States of America* 92, 964-967.
27. Guan, H. P., and Preiss, J. (1993) Differentiation of the Properties of the Branching Isozymes from Maize (*Zea mays*), *Plant Physiology* 102, 1269-1273.
28. Kuriki, T., Guan, H., Sivak, M., and Preiss, J. (1996) Analysis of the Active Center of Branching Enzyme II from Maize Endosperm, *J Protein Chem* 15, 305-313.
29. Kuriki, T., Stewart, D. C., and Preiss, J. (1997) Construction of Chimeric Enzymes out of Maize Endosperm Branching Enzymes I and II: Activity and Properties, *Journal of Biological Chemistry* 272, 28999-29004.
30. Takeda, Y., Guan, H.-P., and Preiss, J. (1993) Branching of amylose by the Branching Isoenzymes of maize endosperm, *Carbohydr. Res.* 240, 253-263.
31. Kawasaki, T., Mizumo, K., Baba, T., and Shimada, H. (1993) Molecular analysis of the gene encoding a rice starch branching enzyme, *Molecular and General Genetics* 237, 10-16.
32. Mizuno, K., Kawasaki, T., Shimada, H., Satoh, H., Kobayashi, E., Okumura, S., Arai, Y., and Baba, T. (1993) Alteration of the Structural Properties of Starch Components by the Lack of an Isoform of Starch Branching Enzyme in Rice Seeds, *The Journal of biological chemistry* 268, 19084-19091.
33. Mizuno, K., Kimura, K., Arai, Y., Kawasaki, T., Shimada, H., and Baba, T. (1992) Starch Branching Enzymes from Immature Rice Seeds, *Journal of Biochemistry* 112, 643-651.
34. Mizuno, K., Kobayashi, E., Tachibana, M., Tsutomu, K., Fujimura, T., Funane, K., Kobayashi, M., and Baba, T. (2001) Characterization of an isoform of rice starch branching enzyme, RBE4, in developing seeds, *Plant Cell Physiol.* 42, 349-357.

35. Nakamura, Y., Takeichi, T., Kawaguchi, K., and Yamanouchi, H. (1992) Purification of two forms of starch Branching Enzyme (Q-enzyme) from developing rice endosperm, *Plant Physiology* 84, 329–335.
36. Nakamura, Y., Utsumi, Y., Sawada, T., Aihara, S., Utsumi, C., Yoshida, M., and Kitamura, S. (2010) Characterization of the reactions of Starch Branching Enzymes from rice endosperm, *Plant and Cell Physiology* 51, 776-794.
37. Morell, M. K., Blennow, A., Kosar-Hashemi, B., and Samuel, M. (1997) Differential Expression and Properties of Starch Branching Enzyme Isoforms in Developing Wheat Endosperm, *Plant Physiol.* 113, 201-208.
38. Rahman, S., Li, Z., Abrahams, S., Abbott, D., Appels, R., and Morell, M. K. (1999) Characterisation of a gene encoding wheat endosperm starch branching enzyme-I, *Theor. Appl. Genet.* 98, 156-163.
39. Repellin, A., Båga, M., and Chibbar, R. N. (2001) Characterization of a cDNA encoding a type I starch branching enzyme produced in developing wheat (*Triticum aestivum* L.) kernels, *Journal of Plant Physiology* 158, 91-100.
40. Sadequr Rahman\*, A. R., Zhongyi Li, Yasuhiko Mukai, Maki Yamamoto, Behjat Kosar-Hashemi, Sharon Abrahams, and Matthew K. Morell. (2001) Comparison of Starch-Branching Enzyme Genes Reveals Evolutionary Relationships Among Isoforms. Characterization of a Gene for Starch-Branching Enzyme IIa from the Wheat D Genome Donor *Aegilops tauschii*1, *Plant Physiology* 125.
41. Sun, C., Sathish, P., Ahlandsberg, S., Deiber, A., and Jansson, C. (1997) Identification of four starch-branching enzymes in barley endosperm-partial purification of forms I, IIa and IIb, *New Phytol.* 137, 215-222.
42. ChuanXin, S., Sathish, P., Deiber, A., and Jansson, C. (1995) Multiple starch branching enzymes in barley endosperm, *Photosynthesis: from light to biosphere Volume V. Proceedings of the Xth International Photosynthesis Congress, Montpellier, France*, 317-320.
43. Sun, C., Sathish, P., Ahlandsberg, S., and Jansson, C. (1998) The Two Genes Encoding Starch-Branching Enzymes IIa and IIb Are Differentially Expressed in Barley, *Plant Physiol.* 118, 37–49.
44. Rydberg, U., Andersson, L., Andersson, R., Åman2, P., and Larsson, H. (2001) Comparison of starch branching enzyme I and II from potato, *Eur. J. Biochem.* 268, 6140-6145.

45. Visko-Nielsen, A., and Blennow, A. (1998) Isolation of starch branching enzyme I from potato using gamma-cyclodextrin affinity chromatography, *J. Chromatogr.* **800**, 382-385.
46. Salehuzzaman, S. N. I. M., Jacobsen, E., and Visser, R. G. F. (1992) Cloning, partial sequencing and expression of a cDNA coding for branching enzyme in cassava, *Plant Molecular Biology* **20**, 809-819.
47. Amith, A. M. (1988) Major differences in isoforms of starch-branching enzyme between developing embryos of round- and wrinkled-seeded peas (*Pisum sativum* L.), *Planta* **175**, 270-279.
48. Tomlinson, K. L., Lloyd, J. R., and Smith, A. M. (1997) Importance of isoforms of starch-branching enzyme in determining the structure of starch in pea leaves, *The Plant Journal* **11**, 31-43.
49. Zhu, Z. P., and Chen, S. Q. (1992, 20-24 July) Starch accumulation and branching enzyme gene expression during seed development of soybean and pea, *Proceedings of the Fourth International Workshop on Seeds: basic and applied aspects of seed biology, Angers, France 1*, 127-133.
50. Noguchi, J., Chaen, K., Vu, N. T., Akasaka, T., Shimada, H., Nakashima, T., Nishi, A., Satoh, H., Omori, T., Kakuta, Y., and Kimura, M. (2011) Crystal structure of the Branching Enzyme I (BEI) from *Oryza sativa* L with implications for catalysis and substrate binding, *Glycobiology* **21**, 1108-1116.
51. Nakamura, Y. (2002) Towards a better understanding of the metabolic system for amylopectin biosynthesis in plants, *Plant Cell Physiol.* **43**, 718-725.
52. Vu, N. T., Shimada, H., Kakuta, Y., Nakashima, T., Ida, H., Omori, T., Nishi, A., Satoh, H., and Kimura, M. (2008) Biochemical and Crystallographic Characterization of the Starch Branching Enzyme I (BEI) from *Oryza sativa* L, *Biosci. Biotechnol. Biochem.* **72**, 2858-2866.
53. Chaen, K., Noguchi, J., Omori, T., Kakuta, Y., and Kimura, M. (2012) Crystal structure of the rice branching enzyme I (BEI) in complex with maltopentaose, *Biochemical and biophysical research communications* **424**, 508-511.
54. Rodriguez-Sanoja, R., Oviedo, N., and Sanchez, S. (2005) Microbial starch-binding domain, *Current opinion in microbiology* **8**, 260-267.
55. Sorimachi, K., Le Gal-Coëffet, M.-F., Williamson, G., Archer, D. B., and Williamson, M. P. (1997) Solution structure of the granular starch binding domain of *Aspergillus niger* glucoamylase bound to beta-cyclodextrin, *Structure* **5**, 647-661.

56. Blennow, A., Visko-Nielsen, A., and Morell, M. K. (1998) Alpha-Glucan binding of potato-tuber starch-branching enzyme I as determined by tryptophan fluorescence quenching, affinity electrophoresis and steady-state kinetics, *Eur. J. Biochem.* 252, 331-338.
57. Borovsky, D., Smith, E. E., and Whelan, W. J. (1975) Purification and Properties of Potato 1,4-alpha-D-Glucan: 1,4-alpha-D-Glucan 6-alpha-(1,4-alpha-Glucano)-Transferase, *Eur. J. Biochem.* 59, 615-625.
58. Boyer, C. D., and Preiss, J. (1978) Multiple Forms of (1,4)-alpha-D-Glucan, (1,4)-alpha-D-Glucan-6-Glycosyl Transferase From Developing *Zea mays* L. Kernels, *Carbohydrate Research* 61, 321-334.
59. National Institute of Agrobiological Sciences, J. *Rice Genome Resource Center* [http://cdna01.dna.affrc.go.jp/cDNA/CDNA\\_main\\_front.html](http://cdna01.dna.affrc.go.jp/cDNA/CDNA_main_front.html).
60. Otwinowski, Z., and Minor, W. (1997) Processing of X-Ray Diffraction Data Collected in Oscillation Mode, *Methods in Enzymology* 276, 307-326.
61. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures, *Acta Crystallographica D* 67, 355-367.
62. Murshudov, G. N., Vagin, A. A., and J., D. E. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method, *Acta Crystallographica D, EU Validation contract: BIO2CT-92-0524* 53, 240-255.
63. Boraston, A. B., Healey, M., Klassen, J., Ficko-Blean, E., Lammerts van Bueren, A., and Law, V. (2006) A structural and functional analysis of alpha-glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition, *The Journal of biological chemistry* 281, 587-598.
64. Brzozowski, A. M., and Davies, G. J. (1997) Structure of the *Aspergillus oryzae* Alpha-Amylase Complexed with the Inhibitor Acarbose at 2.0 Å Resolution, *Biochemistry* 36, 10837-10845.
65. Uitdehaag, J. C. M., van Alebeek, G. J., Van Der Veen, B., Dijkhuizen, L., and Dijkstra, B. W. (2000) Structures of Maltohexaose and Maltoheptaose Bound at the Donor Sites of Cyclodextrin Glycosyltransferase Give Insight into the Mechanisms of Transglycosylation Activity and Cyclodextrin Size Specificity, *Biochemistry* 39, 7772-7780.
66. Dumbrepatil, A. B., Choi, J. H., Park, J. T., Kim, M. J., Kim, T. J., Woo, E. J., and Park, K. H. (2010) Structural features of the *Nostoc punctiforme* debranching enzyme reveal the basis of its mechanism and substrate specificity, *Proteins* 78, 348-356.

67. Kanai, R., Haga, K., Akiba, T., Yamane, K., and Harata, K. (2004) Biochemical and Crystallographic Analyses of Maltohexaose-Producing Amylase from Alkalophilic *Bacillus* sp. 707, *Biochemistry* 43, 14047-14056.
68. Knegtel, R. M. A., Strokopytov, B., Penninga, D., Faber, O. G., Rozeboom, H. J., Kalk, K. H., Dijkhuizen, L., and Dijkstra, B. W. (1995) Crystallographic Studies of the Interaction of Cyclodextrin Glycosyltransferase from *Bacillus circulans* Strain 251 with Natural Substrates and Products, *Journal of Biological Chemistry* 270, 29256-29264.
69. Larson, S. B., Day, J. S., and McPherson, A. (2010) X-ray crystallographic analyses of pig pancreatic alpha-amylase with limit dextrin, oligosaccharide, and alpha-cyclodextrin, *Biochemistry* 49, 3101-3115.
70. Lin, J., Deng, H., Jin, L., Pandey, P., Quinn, J., Cantin, S., Rynkiewicz, M. J., Gorga, J. C., Bibbins, F., Celatka, C. A., Nagafuji, P., Bannister, T. D., Meyers, H. V., Babine, R. E., Hayward, N. J., Weaver, Benjamin, H., Stassen, F., Abdel-Meguid, S. S., and Strickler, J. E. (2006) Design, Synthesis, and Biological Evaluation of Peptidomimetic Inhibitors of Factor XIa as Novel Anticoagulants, *J Med. Chem.* 49, 7781-7791.
71. Mikami, B., Adachi, M., Kage, T., Sarikaya, E., Nanmori, T., Shinke, R., and Utsumi, S. (1999) Structure of Raw Starch-Digesting *Bacillus cereus* beta-Amylase Complexed with Maltose, *Biochemistry* 38, 7050-7061.
72. Mikami, B., Iwamoto, H., Malle, D., Yoon, H. J., Demirkan-Sarikaya, E., Mezaki, Y., and Katsuya, Y. (2006) Crystal structure of pullulanase: evidence for parallel binding of oligosaccharides in the active site, *Journal of molecular biology* 359, 690-707.
73. Payan, F., and Qian, M. (2003) Crystal Structure of the Pig Pancreatic 'Alpha-Amylase Complexed with Malto-Oligosaccharides, *J Protein Chem* 22, 275-284.
74. Qian, M., Haser, R., Buisson, G., Duee, E., and Payan, F. (1994) The Active Center of a Mammalian a-Amylase. Structure of the Complex of a Pancreatic alpha-Amylase with a Carbohydrate Inhibitor Refined to 2.2-A Resolution, *Biochemistry* 33, 6284-6294.
75. Strokopytov, B., Penninga, D., Rozeboom, H. J., Kalk, K. H., Dijkhuizen, L., and Dijkstra, B. W. (1995) X-ray Structure of Cyclodextrin Glycosyltransferase Complexed with Acarbose. Implications for the Catalytic Mechanism of Glycosidases, *Biochemistry* 34, 2234-2240.
76. Tan, T. C., Mijts, B. N., Swaminathan, K., Patel, B. K., and Divne, C. (2008) Crystal structure of the polyextremophilic alpha-amylase AmyB

- from *Halothermothrix orenii*: details of a productive enzyme-substrate complex and an N-domain with a role in binding raw starch, *Journal of molecular biology* 378, 852-870.
77. Vester-Christensen, M. B., Abou Hachem, M., Svensson, B., and Henriksen, A. (2010) Crystal structure of an essential enzyme in seed starch degradation: barley limit dextrinase in complex with cyclodextrins, *Journal of molecular biology* 403, 739-750.
  78. Yokota, T., Tonozuka, T., Shimura, Y., Ichikawa, K., Kamitori, S., and Sakano, Y. (2001) Structures of *Thermoactinomyces vulgaris* R-47 alpha-amylase II complexed with substrate analogues, *Bioscience, Biotechnology, and Biochemistry* 65, 619-626.
  79. Jespersen, H. M., MacGregor, E. A., Sierks, M. R., and Svensson, B. (1991) Comparison of the domain-level organization of starch hydrolases and related enzymes, *Biochemistry J.* 280, 51-55.
  80. Svensson, B. (1994) Protein engineering in the  $\alpha$ -amylase family- catalytic mechanism, substrate specificity, and stability, *Plant Molecular Biology* 25, 141-157.
  81. Katsuya, Y., Mezaki, Y., Kubota, M., and Matsuura, Y. (1998) Three-dimensional Structure of *Pseudomonas* Isoamylase at 2.2 Å Resolution, *Journal of molecular biology* 281, 885-897.
  82. Malle, D., Iwamoto, H., Katsuya, Y., Utsumi, S., and Mikami, B. (TBA) Crystal structure of pullulanase type I from *Bacillus subtilis* str. 168 in complex with maltose and alpha-cyclodextrin, *TBA*.
  83. Wu, L., Yang, D., Li, J., Zhou, J., and Wu, J. (TBA) The Crystal Structure of Gamma-Cgtase from Alkalophilic *Bacillus Clarkii* at 1.65 Å Resolution., *TBA*.
  84. Binderup, K., and Preiss, J. (1998) Glutamate-459 Is Important for *Escherichia coli* Branching Enzyme Activity, *Biochemistry* 37, 9033-9037.
  85. Kumar, V. (2010) Analysis of the key active subsites of glycoside hydrolase 13 family members, *Carbohydr Res* 345, 893-898.
  86. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition, *Biochemistry J.* 382, 769-781.
  87. Tomme, P., Warren, R. A., Miller, R. C., Jr., Kilburn, D. G., and Gilkes, N. R. (1995) Cellulose-binding domains: classification and properties. , *Enzymatic Degradation of Insoluble Polysaccharides* (Saddler, J.N. & Penner, M., eds.) American Chemical Society, Washington, 142-163.

88. Penninga, D., van der Veen, B. A., Knegt, R. M. A., van Hijum, S. A. F. T., Rozeboom, H. J., Kalk, K. H., Dijkstra, B. W., and Dijkhuizen, L. (1996) The Raw Starch Binding Domain of Cyclodextrin Glycosyltransferase from *Bacillus circulans* Strain 251, *Journal of Biological Chemistry* 271, 32777-32784.
89. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics, *Nucleic acids research* 37, D233-238.
90. Palomo, M., Kralj, S., van der Maarel, M. J., and Dijkhuizen, L. (2009) The unique branching patterns of *Deinococcus* Glycogen Branching Enzymes are determined by their N-terminal domains, *Applied and environmental microbiology* 75, 1355-1362.
91. Polekhina, G., Gupta, A., van Denderen, B. J., Feil, S. C., Kemp, B. E., Stapleton, D., and Parker, M. W. (2005) Structural basis for glycogen recognition by AMP-activated protein kinase, *Structure* 13, 1453-1462.
92. Uitdehaag, J. C. M., Mosi, R., Kalk, K. H., van der Veen, B., Dijkhuizen, L., Withers, S. G., and Dijkstra, B. W. (1999) X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the  $\alpha$ -amylase family, *Nature Structural Biology* 6, 432-436.
93. Nakajima, R., Imanaka, T., and Aiba, S. (1986) Comparison of amino acid sequences of eleven different  $\alpha$ -amylases, *Appl. Microbiol. Biotechnol.* 23, 355-360.

## CHAPTER 5: CLONING, EXPRESSION AND PURIFICATION OF POTATO

### TUBER ADP-GLUCOSE PYROPHOSPHORYLASE

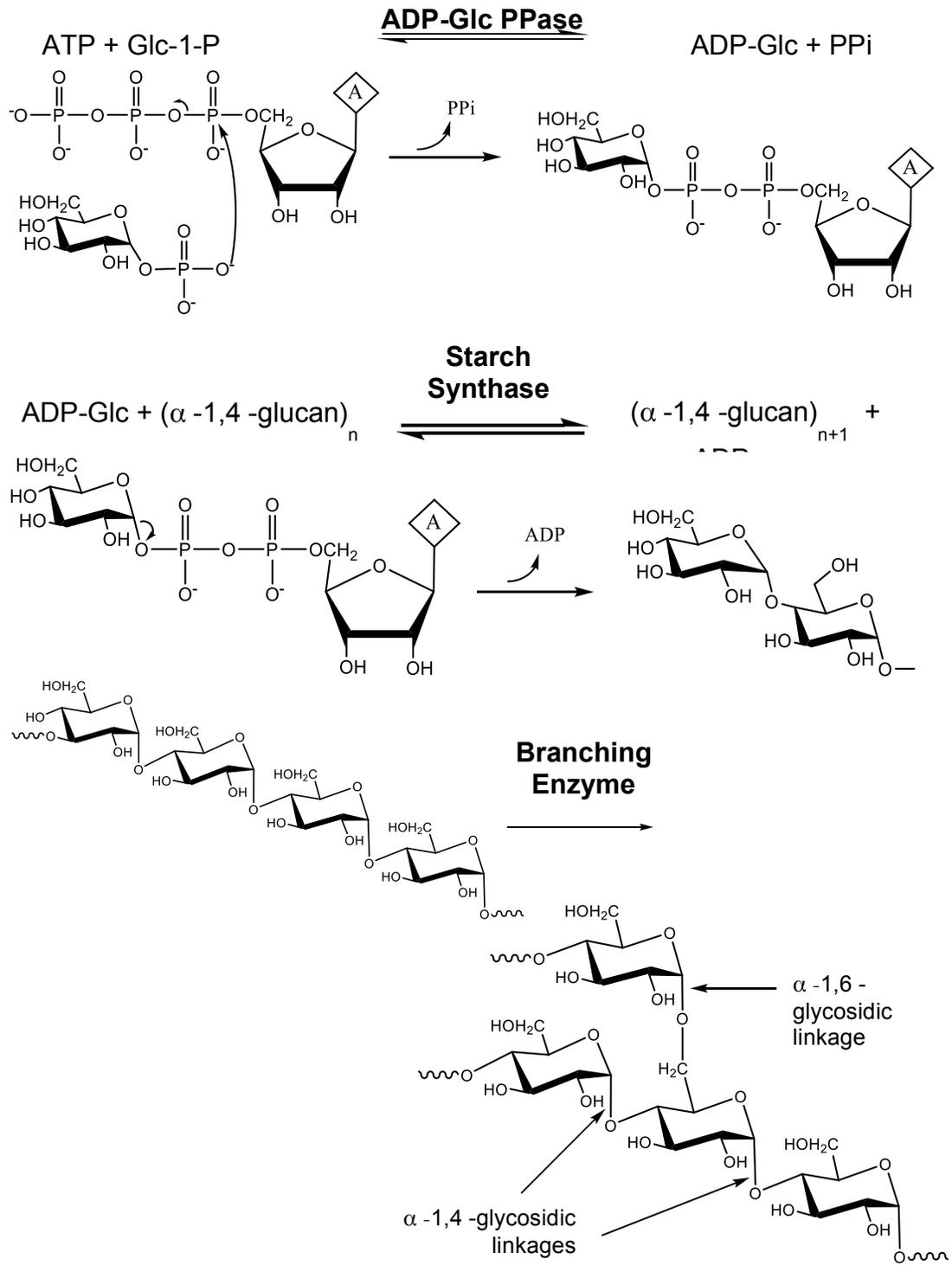
#### 5.1 Introduction

Starch and glycogen are biosynthesized in three major steps in photosynthetic eukaryotes and bacteria, and in mammals, fungi and eukaryotic heterotrophic micro-organisms, respectively. <sup>1</sup> While both groups use an activated form of glucose, we are interested in the first group, plants and bacteria, which utilize ADP-Glucose Pyrophosphorylase (ADP-Glc PPase), an enzyme that catalyzes and regulates the first step of the polyglucan biosynthesis (**Figure 5.1**). This step involves a reaction between  $\alpha$ -glucose-1-phosphate ( $\alpha$ -Glc-1-P) and an ATP molecule to produce a molecule of ADP-Glc and a pyrophosphate (**Figure 5.1**). This reaction is rate-limiting to the process of synthesis and is a reversible one, *in vitro*. <sup>2</sup> The reaction requires a metal ion,  $Mg^{2+}$ , which along with ATP binds to the enzyme first, before Glc-1-P binding, making the reaction sequential. <sup>3</sup> The enzyme is allosterically regulated by small effector molecules, which are key metabolites indicating either high or low carbon and energy content inside the cell. <sup>4</sup> Many higher plant tissues have been studied thus far, and all share a common major activator for the enzyme, 3-phosphoglycerate (3-PGA), and a common major inhibitor, orthophosphate ( $P_i$ ). <sup>5,6</sup> The ratio  $[3\text{-PGA}/P_i]$  regulates starch synthesis where starch accumulation has a parallel relationship with the concentration of 3-PGA, and an inverse reciprocity to the concentration of  $P_i$ . <sup>5,6</sup> Experiments on a mutant, *Arabidopsis thaliana*, where the plant lacked both

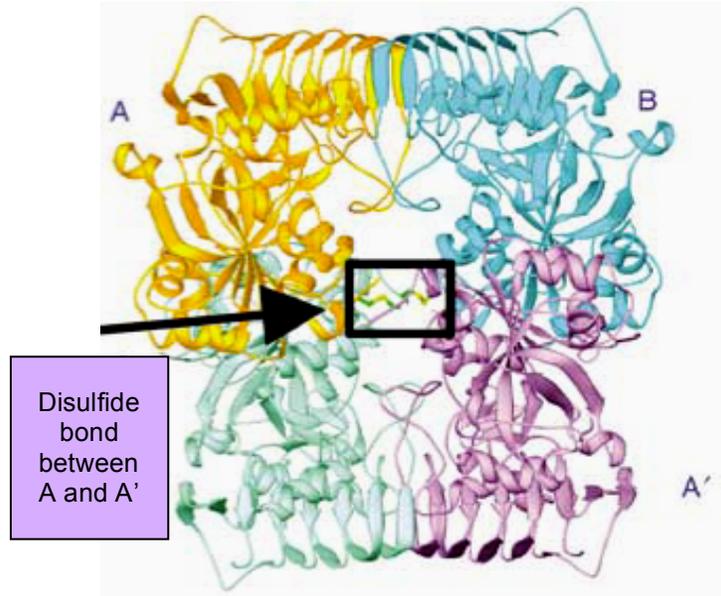
subunits of the enzyme ADP-Glc PPase, showed inability to make starch (with only 2%) compared to normal plants, which confirms that starch biosynthesis can not thrive without synthesis of the precursor, ADP-Glc, by ADP-Glc PPase.<sup>7</sup>

ADP-Glc PPase is a tetrameric protein formed of two small subunits, called  $\alpha$ , and two large subunits, called  $\beta$ .<sup>8</sup> The  $\alpha$  subunit is highly conserved along the higher plants, while the  $\beta$  subunit is less conserved. In potato tuber, the two subunits  $\alpha$  and  $\beta$  share 53% identity.<sup>9</sup> The protein was isolated and characterized.<sup>10</sup> It was demonstrated that the  $\alpha$  subunit, which expresses as a homotetramer, is highly active especially when present with the activator 3-PGA, for which it has a lower affinity than the heterotetramer, whereas the  $\beta$  subunit does not turn into an active homotetramer.<sup>10</sup>

The crystal structure of the homotetramer  $\alpha_4$  form ADP-Glc PPase was determined by Jin et al. in 2005 (**Figure 5.2**).<sup>11</sup> The protein adopted the inhibited conformation, which prompted us to try determining the crystal structure of the activated form of the protein.



**Figure 5.1: The three step process of starch / glycogen biosynthesis.**



**Figure 5.2: ADP-Glc PPase homotetramer.** The disulfide bond between monomers A and A' is boxed in black.

## 5.2 Experimental Procedure

### 5.2.1 Potato small subunit ADP-Glc PPase

#### 5.2.1.1 Vectors

The small subunit ADP-Glc PPase gene (SSpyro) was initially in a pet 28 vector but was overexpressed in extremely low quantities and was not soluble, so we decided to insert it into a pMAL-Sumo vector making the following sequence: pMAL-Sumo-His-SSpyro. The pMAL-c4X vector, purchased from New England BioLabs, is designed to produce the maltose binding protein (MBP) fusions. More details about this vector can be found by searching pMAL-c4X vector on [www.neb.com](http://www.neb.com). MBP tightly binds to amylose resin and is expected to increase overexpression and solubilize the insoluble protein of interest. The Sumo<sup>12</sup> tag was inserted in the Geiger laboratory into the pMAL-c4X vector using NdeI and BamHI restriction enzymes as follows: pMAL-NdeI-Sumo-BamHI. Sumo is a small protein that binds specifically to Sumo protease and can be later cleaved off of the expressed protein. SSpyro was then inserted into the vector downstream of the Sumo tag, keeping the Histidine tag upstream of SSpyro, using the restriction enzymes BamHI and SalI, making the final version of this vector: pMAL-Sumo-BamHI-SSpyro-SalI.

### **5.2.1.2 Over-expression**

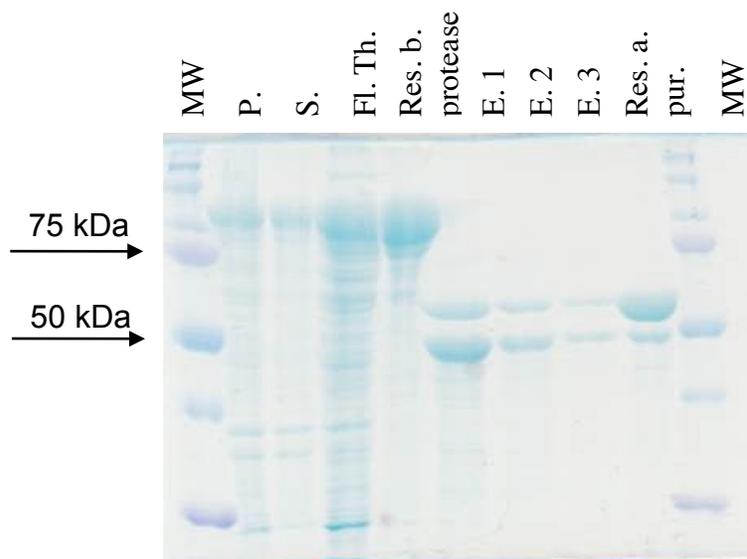
The new vector was transformed into the knock out expression cells, AC70R1-504,<sup>13</sup> which are deficient in the gene coding for ADP-Glc PPase in an effort to eliminate background proteins. Cells were plated on 50 µg/mL ampicillin resistant plates and heated overnight. Colonies were then picked and grown in 50 mL Luria broth (LB), and stocks were flash frozen in liquid nitrogen and stored at -70°C. 1 mL of stock cells was divided into 6 small 50 mL LB flasks. Starting at this step, a multitude of combinations were tried in an attempt to obtain a pure soluble ADP-Glc PPase. The 1 L flasks were grown for a series of different time lengths as follows: ½ hour, 1, 2, 4, 6, 8, 10, and 16 hours. The growth temperatures tried were 5, 10, 16, and 25°C. Shaking speeds in the incubator also varied between 40, 100 and 200 RPM. The media was allowed to shake until cells' OD<sub>600</sub> reached 0.1, 0.2, 0.4, 0.6, or 1.

### **5.2.1.3 Purification**

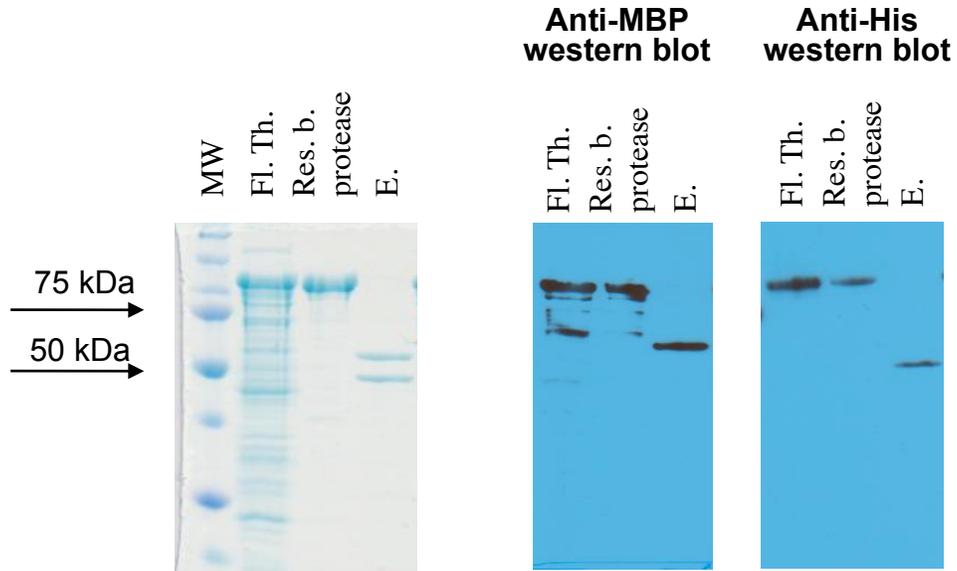
Cells obtained from 1.5 L of media were suspended and lysed into a 50 mL buffer (200 mM NaCl, 10 mM BME, 1 mM EDTA, 10% glycerol, 20 mM Tris-HCl, pH 7.4). Cells were centrifuged at 7000 RPM for 20 min and lysate was transferred to amylose resin and let bind for 3 hours at 4°C with occasional tube rotation. Next, the resin would be washed with the same buffer. Then, 5 mL of wash buffer and 100 units of Sumo protease were added per tube of resin and allowed to cut for 30 min at 0°C; purification column was inverted frequently. The solution was then spun down for 10 min to collect the supernatant buffer

containing the protein, then another 5 mL of buffer were added and solution centrifuged to extract more protein. This step was repeated until all protein has been eluted from resin. Individual steps of the purification are observed on SDS-gel (**Figure 5.3**) and if protein elutions are similar, they were pooled together then protein was either flash frozen at  $-70^{\circ}\text{C}$  or concentrated without freezing. In **figure 5.3**, the pellet, supernatant, flow through, and resin before applying Sumo protease, all showed much of MBP-Sumo-ADP-Glc PPase fusion band at around 200 kDa. Elutions after the protease action showed two major bands, one at 50 kDa and another heavier band at around 70 kDa, while the resin after purification showed the same two major bands, with much more of the heavier band, which is the MBP-Sumo. To confirm the protein identity, we ran western blots with anti-MBP and anti-His on several purifications (**Figure 5.4**).

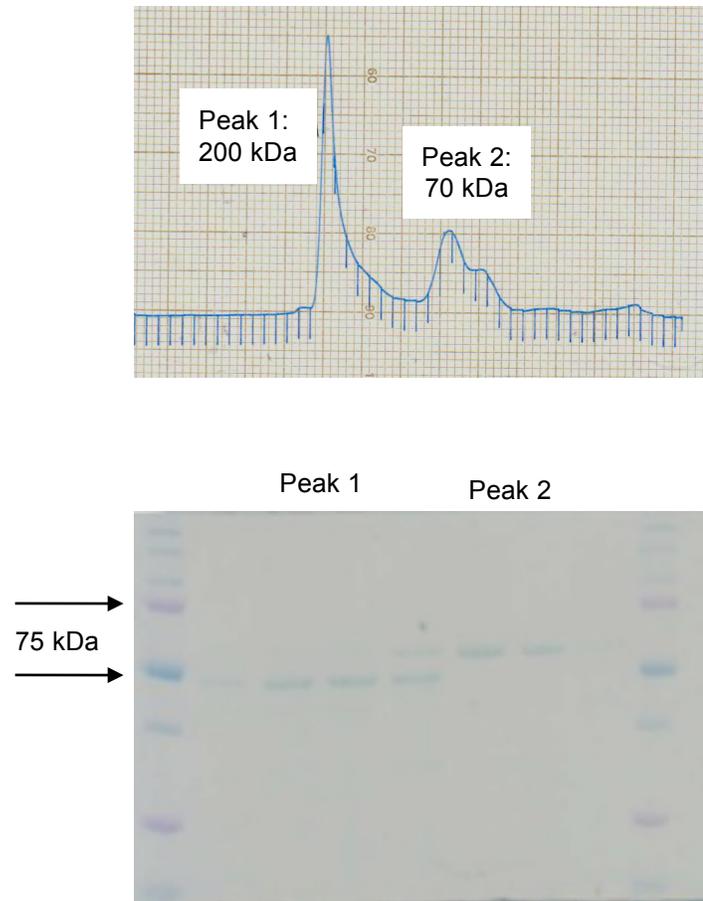
To further purify SSpyro, we concentrated the protein to 5 mg/mL before running the gel filtration chromatography. We have tried a variety of buffers for protein concentration including the purification buffer above, another one where sucrose 10% replaces glycerol, a third one containing 2 mM  $\text{MgCl}_2$ , 10% sucrose, and 10 mM Hepes, pH 7.4, and many others; however, the nature of the buffer did not seem to make a difference in terms of protein loss during concentration. **Figure 5.5** shows gel filtration chromatography diagram. Two proteins were eluted separately, one at 200 kDa and the other at around 70 kDa as determined by comparison to molecular weight standards.



**Figure 5.3: Sample purification of ADP-Glc PPase.** Lanes from left to right are MW, pellet, supernatant, flow through, resin before protease, elutions 1, 2 and 3, resin after purification and MW again.



**Figure 5.4: Anti-MBP and anti-His western blots of ADP-Glc PPase purification.** Loaded in order are the MW, flow through, resin before protease, and elutions. The MW was used for both western blots but is not shown here on the graph; the band at 50kDa always lights up in the anti-His western blot confirming the location of the protein bands identified.



**Figure 5.5: Gel filtration chromatography run of the elution sample.** It shows two peaks, one at 200 kDa and the other at about 70 kDa, representing ADP-Glc PPase tetramer and MBP-Sumo, respectively.

### 5.3 Discussion and Conclusion

MBP-Sumo-ADP Glc PPase was surely being overexpressed and purified as verified by the western blots; the elution lane reflects binding of the anti-MBP and anti-His antibodies to MBP-Sumo and His-SSpyro, respectively. The problem is these two were always both eluted, regardless of the concentration of the protein expressed and amount of resin. Our biggest struggle with this protein, however, was getting a soluble protein. As it is clearly observed in the gel filtration, the first peak represents a 200 kDa protein, which turns out to be the lower weight band on the SDS-gel in the elution band, while the second peak is the MBP-Sumo as determined by the gel filtration. This clearly indicates that SS-pyro is running as a tetramer in solution. The problem however, is that it was never concentrating but precipitating. The highest concentration we were able to get is 2 mg/mL with white flakes showing at the bottom of the centriprep concentrators. Endless trials to in an attempt to solubilize this protein were all unsuccessful.

## REFERENCES

## REFERENCES

1. Sivak, M. N., and Preiss, J. (1998) Starch: basic science to biotechnology, *Advances in Food and Nutrition Research, Taylor SL (ed) 41*, 1-199; San Diego, CA: Academic Press.
2. Ballicora, M. A., Frueauf, J. B., Fu, Y., Schürmann, P., and Preiss, J. (2000) Activation of the Potato Tuber ADP-glucose Pyrophosphorylase by Thioredoxin, *The Journal of biological chemistry* 275, 1315-1320.
3. Haugen, T. H., and Preiss, J. (1979) Biosynthesis of Bacterial Glycogen. The Nature of Binding of substrates and effectors to ADP-Glucose Synthase, *J. Biol. Chem.* 254, 127-136.
4. Preiss, J., and Sivak, M. N. (1998a) Biochemistry, Molecular Biology and Regulation of Starch Synthesis, *Genetic Engineering* 20, 177-223.
5. Preiss, J. (1982) Regulation of the Biosynthesis and Degradation of Starch, *Annu. Rev. Plant, Physiol.* 33, 431-454.
6. Preiss, J. (1988) Biosynthesis of starch and its regulation, *The Biochemistry of Plants. A Comprehensive Treatise. (Preiss, J. ed) 14*, 181-254, Academic Press, San Diego, CA.
7. Lin, T. P., Caspar, T., Somerville, C., and Preiss, J. (1988) Isolation and characterization of a starchless mutant of *Arabidopsis thaliana* L. Henyh lacking ADP-glucose pyrophosphorylase activity., *Plant Physiol.* 86, 1131-1135.
8. Okita, T. W., Nakata, P. A., Andersen, J. M., Sowokinos, J., Morell, M. K., and Preiss, J. (1990) The Subunit Structure of Potato Tuber ADPglucose Pyrophosphorylase, *Plant Physiol.* 93, 785-790.
9. Smith-White, B., and Preiss, J. (1992) Comparison of Proteins of ADP-Glucose Pyrophosphorylase from Diverse Sources, *J. Mol. Evol.* 34, 449-464.
10. Ballicora, M. A., Laughlin, M. J., Fu, Y., Okita, T. W., Barry, C. F., and Preiss, J. (1995) Adenosine 5'-Diphosphate-Glucose Pyrophosphorylase from Potato Tuber : Significance of the N Terminus of the Small Subunit for Catalytic Properties and Heat Stability, *Plant Physiology* 109, 245-251.
11. Jin, X., Ballicora, M. A., Preiss, J., and Geiger, J. H. (2005) Crystal structure of potato tuber ADP-glucose pyrophosphorylase, *EMBO J.* 24, 694-704.

12. Mossesso, E., and Lima, C. D. (2000) Ulp1-SUMO Crystal Structure and Genetic Analysis Reveal Conserved Interactions and a Regulatory Element Essential for Cell Growth in Yeast, *Molecular Cell* 5, 865-876.
13. Iglesias, A. A., Barry, G. F., Meyer, C., Blocksberg, L., Nakatan, P. A., Greenen, T., Laughlin, M. J., Okita, T. W., Kishore, G. M., and Preiss, J. (1993) Expression of the Potato Tuber ADP-glucose Pyrophosphorylase in *Escherichia coli*, *J. Biol. Chem.* 268, 1081-1086.

## CHAPTER 6: EXPERIMENTAL PROCEDURES – *Ec*BE AND RBE I

### 6.1 Vectors

The complete DNA sequence for *E. coli* branching enzyme (*Ec*BE) can be found on the website <http://www.ncbi.nlm.nih.gov>, under the GenBank accession code: M13751.1, and that of the rice branching enzyme I (RBEI), on the rice genome resource center website <http://cdna01.dna.affrc.go.jp/cDNA/>, under the keyword: AK119436. Both genes were truncated at the N-terminus. *Ec*BE used for the current study is missing the first 112 amino acids (N112BE) compared to the full protein and RBEI the first 65 residues. Both genes were engineered into the modified pet 28-Sumo vector, a pet 28a vector with a Sumo tag<sup>1</sup> insertion in order to allow for recognition by Sumo protease; this vector then provides Sumo and Histidine tags. The vectors are then designed as follows: pet28-His-Sumo-BamHI-N112BE-Sall (*E. coli* protein), and pet28-His-Sumo-HindIII-RBEI-Scal.

Previous attempts to crystallize RBEI truncated at the N-terminus were successful but the needle-form crystals were very clustery and resulted in a low-resolution diffraction.<sup>2</sup> In addition, the protein truncated at the C-terminus gave rise to a higher resolution diffraction pattern previously,<sup>2</sup> which directed our focus to the C-terminus truncated protein. This was accomplished by inserting a TEV protease recognition sequence (ENLYFQG) in the pet28-His-Sumo-HindIII-RBEI-Scal sequence at position 702 using the 3 primers listed below consecutively:

*Primer 1:*

5'-GAA GAT CGT GAA GAG CTA AGG AGG GAG AAC GGT GGA GCA GTT  
GCT TCT-3'

*Primer 2:*

5'-AGG AGG GAG AAC CTG TAT TTT CAA GGC GGT GGA GCA GTT GCT  
TC-3'

*Primer 3:*

5'-GAG CTA AGG AGG GAG AAC CTG TAT GGT GGA GCA GTT GCT T-3'

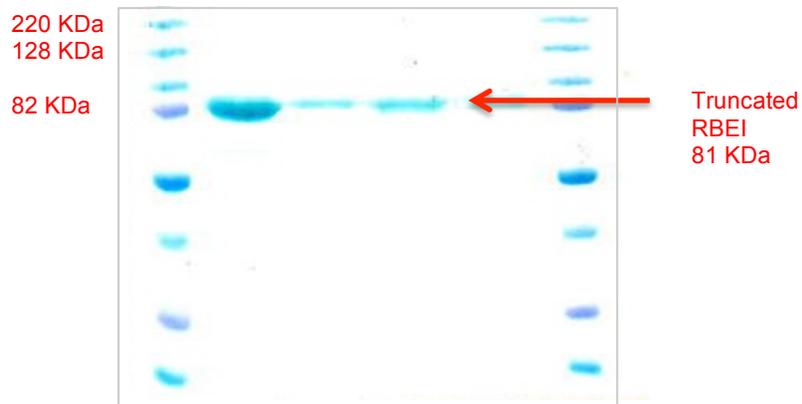
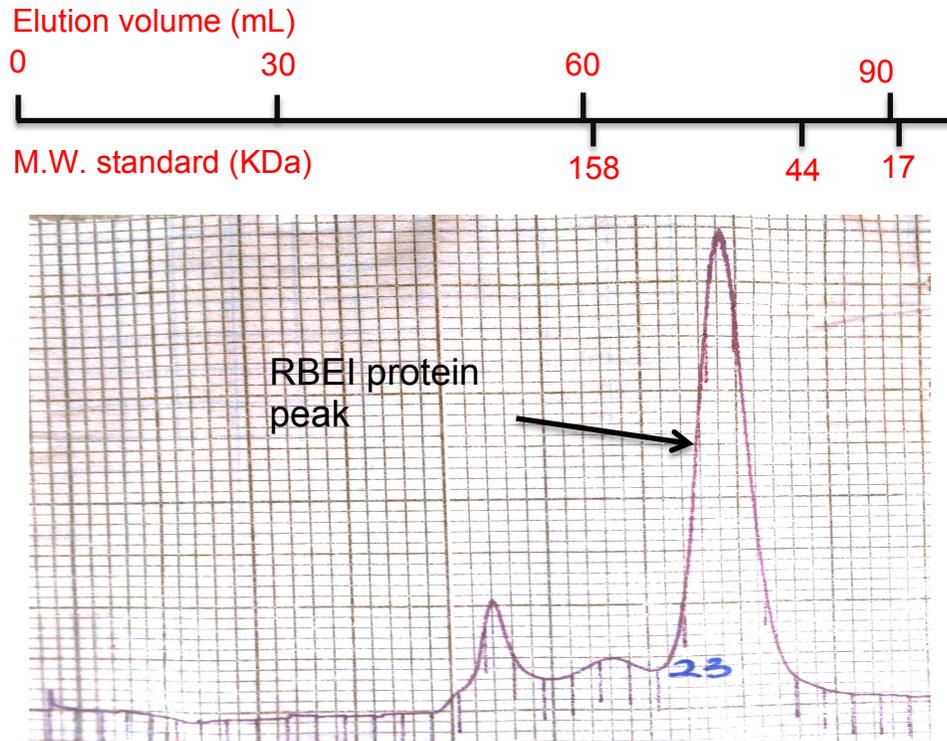
## 6.2 Overexpression and Purification

The vectors were transformed into BL21 codon plus cells purchased from Agilent Technologies (CAT No. 230280). A single colony is allowed to grow overnight in a 50 mL LB media using Kanamycin (Kan) at the ratio of 1:1000 (stock solution 50 mg/mL). Glycerol stocks are flash frozen and stocked at -70°C. To grow a batch of protein, 1 mL of glycerol stock is used. Protein is allowed to grow in 50 mL LB flasks overnight before transferring into 1 L flasks (adding Kan for both sizes). Those are grown to OD<sub>600</sub> = 0.5 at 25°C and 1 mL of a 0.5 mM stock solution IPTG is added for protein overexpression, which is then stopped after 5 hours. Cells are centrifuged and frozen at -20°C for future use.

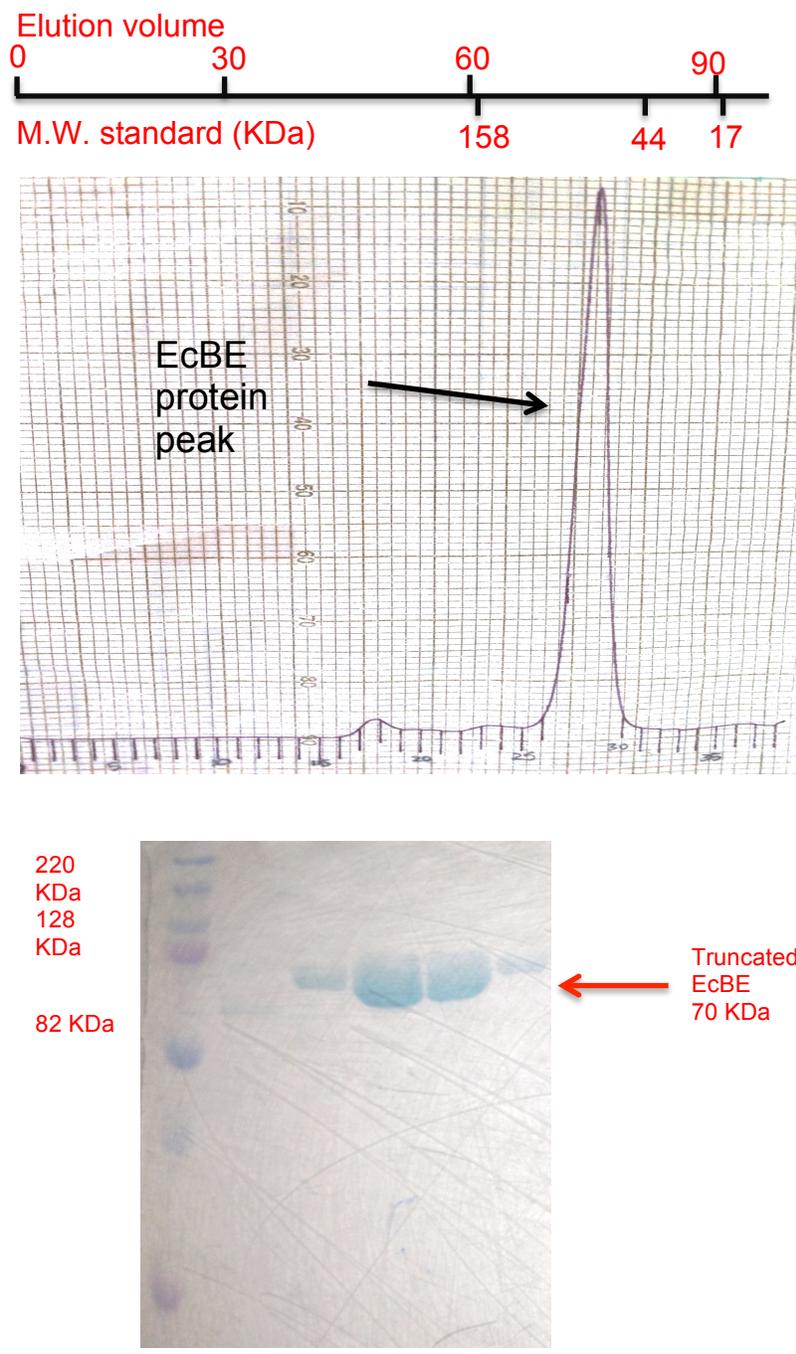
The purification of the protein was performed using the Ni-NTA (Qiagen) column according to the manual recommended protocol with the following modifications: the buffer used consists of 50 mM Tris, pH8.0, 150 mM NaCl and 1 mM BME, with 10 mM imidazole in the lysis buffer and 20 mM in the wash. To identify the protein, a western blot experiment was run on the eluted protein with 250 mM imidazole and the experiment repeated. For all later experiments, the Histidine-Sumo tag was cleaved off of the protein of interest with 100 µL of laboratory prepared Sumo protease for half-an-hour, eluting now with the wash buffer leaves the tag behind attached to the resin beads. Separate elution samples showed a nearly pure protein, with a single minimal impurity band for EcBE and two bands of impurities for RBEI.

The native protein was further purified on a gel-filtration (GF) chromatography column. To prepare the protein for GF, the solution was exchanged overnight into

a buffer containing 50 mM Tris-Cl, pH8.0, 150 mM NaCl for EcBE, while the C-terminus of RBEI was cleaved using the laboratory prepared TEV protease (a 1:100 ratio of TEV-protease to protein ratio is typically used) during the overnight buffer exchange into 50 mM Tris-Cl, pH 8.0, 0.5 mM EDTA and 1mM DTT. The protein was then concentrated in the new buffer with centrprep concentrators (MWCO 10 kDa) at a speed not exceeding 2500 rpm in the early steps and at 1600 rpm when the concentration reached 2 mg/mL. Close attention is required to prevent or catch any possible aggregation early in the concentration process. Aggregation-induced aggregation was common with this protein, and if the protein starts to aggregate, a white residual is noticed in the concentrator, which quickly builds up. The best way to prevent further aggregation is by spinning down the solution and carefully decanting the supernatant for subsequent concentration at lower speeds than previous steps. Checking protein concentration is key in every single step of protein concentration. This tendency towards aggregation is especially true for EcBE. GF as an additional step of purification gave a highly pure protein as shown in the **Figure 6.1** for RBEI and **Figure 6.2** for EcBE. These figures show the gel filtration chromatograph and the SDS-page gel of the final pure protein. The mutants of both proteins were assayed for activity after the Ni-NTA agarose purification without running on size exclusion chromatography column.



**Figure 6.1: RBEI purification (N and C termini truncated).** Top: GF chromatograph (absorption at 280 nm). RBEI is the predominant peak. Bottom: SDS-page gel of samples from tubes corresponding to RBEI peak. The impurity peaks are not shown here. Protein size perfectly matches the standard protein size.



**Figure 6.2: N112BE (EcBE truncated at N-terminus).** Top: GF chromatograph (absorption at 280 nm). EcBE is the predominant peak. Bottom: SDS-page gel of samples from tubes corresponding to EcBE peak. The impurity peaks are not shown here. Very concentrated protein (2 middle peaks) shows lighter than actual size.

### 6.3 Glucan-Iodine Complex Activity Assay

Amylose stock solution: potato amylose was purchased from Sigma Aldrich, with the catalog no. A0512-1G. To dissolve the amylose, 50 mg of solid amylose was mixed with 0.5 mL of 10% NaOH, and 2 mL of water in a 15 mL conical vial. The vial was set in boiling water and inverted frequently until complete amylose dissolution (visible to the eye). This solution was then set to cool down to room temperature before use.

Iodine/iodide stock solution: in a small beaker, 2.6 g of KI was mixed with 0.26 g I<sub>2</sub> and 10 mL of H<sub>2</sub>O.

Working KI / I<sub>2</sub> solution: 2.5 mL stock solution was mixed with 61.95 mL of H<sub>2</sub>O, this amount is sufficient for the 3 trials of one protein sample for the time interval 0 – 30 min. The mixture was stirred in a beaker and split into small test tubes (2.95 mL / tube), preparing a slight excess of this solution is ideal because of pipette error. Iodine and amylose solutions were made fresh before every use.

In small eppendorf tubes, 100 µL of a 1 M Na citrate, pH 7.8 was mixed with 820 µL of H<sub>2</sub>O, and 80 µL of amylose stock solution. The pH of each mixture was adjusted to 8.0 using the small strip pH papers and very small quantities of highly concentrated HCl. The eppendorf tubes were then spun down at 12000 rpm for 10 min to get rid of salt precipitation. 500 µL of the supernatant were then transferred into new labeled-eppendorf tubes now set to pH 8.0. Tubes were allowed to stabilize at 30°C for 10 min. Labeling and arranging the test tubes identically to the eppendorf tubes made work more convenient and practical.

The protein to be tested was thawed on ice and the volume corresponding to 30  $\mu\text{g}$  calculated based on its concentration. The following few steps were performed consecutively in a flash: 30  $\mu\text{g}$  of protein (corresponding volume) added to the first eppendorf tube, tube closed, briefly shook, 50  $\mu\text{L}$  of the tube aliquoted out and quickly added to the test tube of 2.95 mL iodine solution, and the eppendorf tube was put back at 30°C as soon as possible. This represents the first sample at time  $t = 0$  min; steps above were then repeated for all samples of this assay within a 5 min interval. At time  $t = 5$  min, the same steps were repeated for all samples in the same order and at the same time intervals as the first time.

When all seven time-points were completed for all samples, OD for the iodine solutions (test tubes) now containing the protein-amylose samples were checked at 660 nm and graphs of OD versus time were drawn. One unit of activity is defined as a decrease in absorbance of 1.0 absorbance unit per min at OD = 660 nm<sup>3</sup> and is measured in U/mg of protein.<sup>3, 4</sup> Chapter 2 explains the resulting data.

## 6.4 Chain Length Specificity Assay

- Prepare amylose stock solution by mixing 50 mg amylose (from Sigma, potato source) with 500  $\mu\text{L}$  of 10% NaOH and 2 mL double-distilled  $\text{H}_2\text{O}$ , boil to dissolve.
- Mix in Chain Length Specificity (CLS) assay tubes:
  - 10 mg of amylose (500  $\mu\text{L}$  of stock solution)
  - 25  $\mu\text{L}$  Mops (of a 1M stock at pH 7.5)
  - 475  $\mu\text{L}$  dd $\text{H}_2\text{O}$
- Adjust the pH of the above solution to 8.0 using a  $\sim$  6M HCl (which is 26-28  $\mu\text{L}$ ), check pH with pH papers.
- Spin down for 10 min at 10,000 rpm. Aliquot the top 500  $\mu\text{L}$  for the CLS assay.
- Let stabilize at 30°C.
- Add the following amounts of *E. coli* Branching Enzyme (EcBE) (full length (FL), truncated EcBE (N112BE) or buffer for blank) and incubate at 30°C.
  - 0.15 mg of FL-EcBE (1.68 mg/mL),
  - 0.21 mg of t-EcBE (6.7 mg/mL)
- Incubate at 30°C.
- After 1hr and 4 hrs, aliquot out a 200  $\mu\text{L}$  sample, boil for 3 minutes to stop the reaction.
- Let chill to 45°C for 5 min, and then add 20  $\mu\text{L}$  of 1M sodium acetate, pH 3.5, and 200 units of isoamylase (from Sigma, source *Pseudomonas* sp., diluted to 40units/  $\mu\text{L}$ ).

- Incubate at 45°C for 90 min.
- Boil for 3 min, then pH back to neutral with 10% NaOH. (at this step, samples not filtered nor spun down).
- Freeze samples with liquid nitrogen.
- The HPAEC was completed by Dr. Andrew Mort laboratory.

## 6.5 EcBE Mutants

Binding Site	Mutation
	<b>WT Truncated</b>
I	E590A
	E590R
	S583A
	D585A
	D585K
	R255A
	R255D
	R255S
	R576A
	R576D
	H587A
	H587F
II	W595A
	W595L
	D537A
	D537K
	K546A
III	R468A
	W478A
	F239A
	F477A
	P469A

**Table 6.1: Mutations in the N112BE protein.** All listed mutations, including combination mutations, were expressed, purified, and assayed. WT represents wild type truncated protein; CD is the catalytic domain.

**Table 6.1 (cont'd)**

<b>Binding Site</b>	<b>Mutation</b>
<b>IV</b>	K189A
	K189D
	Q211A
	Q211R
	W159A
	W159L
	E215A
	E215W
<b>V</b>	P659A
	S689A
	S689V
	D542A
	D542K
	W544A
	W544K
	Q545A
	Q545K
<b>VI</b>	L512A
	L512W
	D505A
	D505K
	W628A
	W628R
<b>VII</b>	W262A
	W262L
	W262K
	N260A
	N260K
	F261A
	F261W
<b>CD</b>	<b>D405A</b>

**Table 6.1 (cont'd)**

<b>Binding Sites</b>	<b>Mutation</b>
<b>II</b>	H596A+W595L+K546A
<b>IV</b>	K189A+Q211A
<b>IV+VII</b>	W262A+K189D
<b>III+VI</b>	W478A+W628A
<b>III+VI</b>	F477A+L512W

## 6.6 Proteins Crystallization and Substrate Soaking

After several attempts at crystallizing both EcBE and RBEI, we obtained crystals in multiple conditions, and we mention here those growth and soaking conditions that gave the best diffracting crystals:

Only native N112EcBE was crystallized. The final pure protein was concentrated to 5 mg/mL, buffer exchanged into 1.3 M ammonium tartrate, pH 7.6, solution in which very small crystals grew at room temperature within 2 weeks, but took 6 weeks to reach full size, so we decided to micro-seed in order to obtain bigger crystals. A stock solution of maltooctaose was prepared using the same solution. Micro-seeding was performed using the hanging drop method with 2  $\mu$ L of protein solution and 2  $\mu$ L of sugar solution when added. Crystals grew to full size within two weeks of setting the seeds, and the stabilizer solution used was 1.3 M ammonium tartrate, pH 7.6, 10% PEG4k. When crystals were ready, soaking in maltooctaose was tried for a variety of concentrations and times and the best result was accomplished when soaking in a final polysaccharide concentration of 150 mM and for 5 hrs 30 min then flash frozen in 1.3 M ammonium tartrate, pH 7.6, 10% PEG4k, 10% glycerol.

RBEI-core domain gave the best crystals, this was truncated both at the N- and C- termini. The protein was concentrated to about 6 mg/mL. The crystals grew in 30% PEG 8k, 550 mM sodium acetate, and 100 mM sodium cacodylate, pH 6.9, at room temperature. Dodecaose was only soluble at a maximum concentration of 45 mM in the same solution as the protein. The freezing solution contained 9.55% glycerol in addition to the ingredients of crystals' growth.

## REFERENCES

## REFERENCES

1. Mossesso, E., and Lima, C. D. (2000) Ulp1-SUMO Crystal Structure and Genetic Analysis Reveal Conserved Interactions and a Regulatory Element Essential for Cell Growth in Yeast, *Molecular Cell* 5, 865-876.
2. Noguchi, J., Chaen, K., Vu, N. T., Akasaka, T., Shimada, H., Nakashima, T., Nishi, A., Satoh, H., Omori, T., Kakuta, Y., and Kimura, M. (2011) Crystal structure of the Branching Enzyme I (BEI) from *Oryza sativa* L with implications for catalysis and substrate binding, *Glycobiology* 21, 1108-1116.
3. Boyer, C. D., and Preiss, J. (1978) Multiple Forms of (1,4)-alpha-D-Glucan, (1,4)-alpha-D-Glucan-6-Glycosyl Transferase From Developing *Zea mays* L. Kernels, *Carbohydrate Research* 61, 321-334.
4. Binderup, K., and Preiss, J. (1998) Glutamate-459 Is Important for *Escherichia coli* Branching Enzyme Activity, *Biochemistry* 37, 9033-9037.
5. Abad, M. C., Binderup, K., Rios-Steiner, J., Arni, R. K., Preiss, J., and Geiger, J. H. (2002) The X-ray crystallographic structure of *Escherichia coli* Branching Enzyme, *Journal of Biological Chemistry* 277, 42164-42170.
6. Otwinowski, Z., and Minor, W. (1997) Processing of X-Ray Diffraction Data Collected in Oscillation Mode, *Methods in Enzymology* 276, 307-326.
7. Collaborative Computational Project, N. (1994) The CCP4 suite: programs for protein crystallography., *Acta Crysta. D50*, 760-763.
8. Binderup, K., Mikkelsen, R., and Preiss, J. (2002) Truncation of the amino terminus of Branching Enzyme changes its chain transfer pattern, *Archives of biochemistry and biophysics* 397, 279-285.