This is to certify that the
dissertation entitled

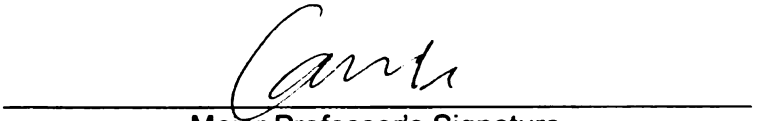
X-RAY CRYSTALLOGRAPHIC STUDIES OF BRANCHING ENZYME/POLYSACCHARIDE COMPLEX AND RNA POLYMERASE III TRANSCRIPTION FACOTOR TFIIIB COMPLEX

presented by

Lei Feng

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Chemistry



Major Professor's Signature
August 26, 2009

Date

X-RAY CRYSTALLOGRAPHIC STUDIES OF BRANCHING ENZYME/POLYSACCHARIDE COMPLEX AND RNA POLYMERASE III TRANSCRIPTION FACOTOR TFIIIB COMPLEX

By

Lei Feng

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Chemistry

2009

ABSTRACT

X-RAY CRYSTALLOGRAPHIC STUDIES OF BRANCHING ENZYME/POLYSACCHARIDE COMPLEX AND RNA POLYMERASE III TRANSCRIPTION FACOTR TFIIIB COMPLEX

By

Lei Feng

Glycogen and starch are major carbon sources and the carbohydrate storage molecule in living organisms. These two highly branched polysaccharides play a very important role in glucose cycles in nature. Branching enzyme is one of the biosynthetic pathway enzymes. It cleaves the α-1,4 glucosidic bond and transfers the oligosaccharide to the α-1,6 position to form α-1, 6 branch points. This action determines the final structure of glycogen and starch, which is very important in nature and industrial application.

The *E.coli* branching enzyme has been crystallized and its structure has been solved. In order to gain insight into its branching mechanism, the binding between the enzyme and its substrate needs to be investigated. Linear and cyclic oligosaccharides were used to bind with *E. coli* branching enzyme. The three dimensional crystal structures were obtained by substrate soaking experiments and x-ray diffraction. From the data it is clear that *E.coli* branching enzyme binds with α, β and γ-cyclodextrins on the surface of the protein. Four binding sites were found and the residues involved were identified. The cyclodextrins bind with the protein through hydrogen bonds and aromatic stacking interactions at the binding sites. Furthermore, the binding between *E.coli* branching

enzyme and linear oligosaccharides such as maltohexaose and maltoheptase identified binding sites near the catalytic region, and lead to a hypothesis for the mechanism of substrate and branch chain specificity..

Branching enzyme II from maize endo sperm was also purified and screened for crystallographic study. The full length protein did not crystallize. Further study unveiled a stable truncated version of the protein. Crystallization attempts on this variant are ongoing.

The structure of the TFIIIB/DNA transcription initiation complex can lead to the understanding of how transcription initiates. This is an extremely important step in the transfer of genetic information. The TBP, BRF and B" protein are the three subunits of TFIIIB, the central RNA polymerase III cofactor. Based on previous studies, different modified TATA box containing promoters were selected for the study. Different constructs of the subunits were designed and purified to make a variety of TBP/BRF/B"/DNA complexes *in vitro*. The quaternary complexes were screened for crystallization, but crystals were not obtained.

To my dear mother, Mrs. Ren, Sumei; my dear father, Mr. Feng, Mengxue; my lovely daughter Helen and my wife

ACKNOWLEDGMENTS

First of all, I would like to thank my advisors, Professor James H. Geiger for mentoring me in my adventure into science research. He showed enormous patience and support when I got lost in my projects. Without his help and encouragement I could not have finished my graduate study at Michigan State University. His knowledge and passion in science guided me through the difficult times. I sincerely thank him for all his help.

I also would like to thank our collaborator, Professor Jack Preiss. As a well-known pioneer in the area of the starch biosynthesis pathway, his support for our project was tremendous. I got lots of invaluable help from discussions with Dr. Preiss about the branching enzyme project.

I owe a very special thank to Dr. Stacy Hovde. She is the one who taught me everything in the lab and helped me on all my research. She actually worked closely with me for long time in Geiger's lab before she went to the Department of Biochemistry and Molecular Biology. Without her kind help, I could not achieve anything. Thank you, Stacy!

I also want to say thank you to my friends and group members Xiaofei Jia, Xiangshu Jin, Fang Sheng, and Susan Wortas. Xiaofei helped me with everything in my study and life; I learned a lot from the discussion with Fang. Their help and friendship have been

the source of my energy, courage and happiness.

I also would like to thank our former group members Aimee Brooks and Katie Strong for their efforts in the project.

Finally I am also in great debt to my beloved parents and my wife. Their love and support has always been with me. My parents' unconditional love, support and understanding is the most important encouragement to me when I encounter all kind of challenges in the pursuit of my Ph.D degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Images in this thesis/dissertation are presented in color

LIST OF ABBREVIATIONS

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	Phenylanaline
Pro, P	Proline
Ser, S	Serine
Thr, T	Threonine
Trp, W	Tryptophan

Tyr, Y	Tyrosine
Val, V	Valine

Other Symbols and Abbreviations

BE	branching enzyme
WT	wild type
CD	cyclodextrin
E.coli	*Escherichia coli*
LB	Luria broth
UV	Ultraviolet light
Vis	visible light
APS	Advanced Photon Source
b.p	Base pair
Da	Dalton
RNA	Ribonucleic acid
DNA	Deoxyribonucleic acid
FPLC	Fast protein liquid chromatography
hr	Hour
K_D	Dissociation constant
λ_{max}	Wavelength of the maximum absorption
Å	Ångström

L	Liter
M	Molar
µl	microliter
ml	milliliter
nm	Nanometer
mm	millimeter
MW	Molecular weight
nM	Nanomolar
DEAE	Diethylaminoethyl cellulose
IPTG	Isopropyl-1-thio-β-D-galactopyranoside
Tris	2-Amino-2-(hydroxymethyl)-1,3-propanediol
DTT	Dithiothreitol
PEG	Polyethylene glycol
PCR	Polymerase chain reaction
PDB	Protein Data Bank
CCP4	Collaborative computational project, number 4
R-factor	Reliability factor
Rmsd	root mean square deviation
SDS-PAGE	Sodium dodecyl sulfate – polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography

CHAPTER 1: INTRODUCTION

1.1 Branching Enzyme

1.1.1 Glycogen

Glycogen is the major carbohydrate storage molecule in bacteria and animal cells. It is widely found in the muscles and livers of all higher animals. In fact, it is called animal starch because it is the primary reserve polysaccharide for animals and very close to starch(*1-3*).

Glycogen is usually found in the form of granules in the cytosol in cells. Glycogen forms an energy reserve that can be quickly broken down to meet a sudden need for glucose when the blood glucose level is low. Glycogen from liver can be degraded to glucose under the action of glycogen phosphorylase so that it can meet the need of the rest of the body. Abnormal glycogen storage can cause various diseases(*4*).

Glycogen is a highly branched non-reducing water-soluble polysaccharide. In 1936, the structures of glycogen from different organism were determined (*5-6*). Depending on the source of glycogen, it consists of more than sixty thousand glucose units with molecular weights up to 10^7 Daltons with 8-12% branches. Chemistry studies showed glycogen is a polysaccharide composed of α-1,4-linked glucans and branched by α-1,6-glycosidic linkages (*1, 6-7*). In glycogen, the branches usually occur at intervals of 8-10 glucose units. The branched polysaccharide can be accumulated in cells in a more

efficient way, which makes the glycogen granules denser.

1.1.2 Starch

Different from glycogen, starch is another form of energy and carbohydrate storage found only in photosynthetic eukaryotes or their non-photosynthetic derivatives. It is the second most abundant polysaccharide in nature after cellulose. Starch is also the primary nutrition source in the human diet and is becoming increasingly important as a renewable industrial biomaterial (**Figure 1.1**). It is widely found in higher plants such as wheat, maize, potato and tapioca (*8-9*).

Starch is synthesized in plants as a result of photosynthesis, a process during which energy from sunlight is converted into chemical energy and stored in starch (*10*). It is synthesized in both plastids of leaves and amyloplasts in tubers, seeds and roots of plants. Large amounts of starch accumulate in the latter case for long-term energy storage.

Unlike glycogen, which is basically homologous and water-soluble, starch is essentially an insoluble semi-crystalline material. Starch is a huge (from 0.1 to over 50 μ m in diameter) complex made of two distinct polysaccharide fractions : amylopectin and amylose (**Figure 1.3**). Native starch granules typically contain around 20-25% amylose and 75-80% amylopectin (*10*). Amylopectin , the major compound, is composed of moderate size of α-1,4 linked glucans that are clustered together and hooked to longer spacer glucans by α-1,6 linkages. Nevertheless amylose consists almost exclusively of linear a-1,4-linked polymers (with less than 0.6% branches). Compared with glycogen,

starch has significantly less branches on average. The α-1,6 linkages occur only every 24-30 glucose units.

The semi-crystalline structure of the starch granule is represented by **Figure 1. 2**. The short amylopectin chains form double helices and associate into clusters. Then the clusters pack together to form a structure of alternating crystalline and amorphous lamellar composition (*11-12*).

The linear amylose is found to exist in a left-handed helical structure in its crystalline state (*13-15*). In each turn there are 6 glucose units, which make it very similar to α-cyclodextrin. Figure 1.4 is the schematic depiction of the helical structure(*2*).

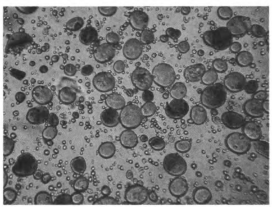


Figure 1. 1 Granules of wheat starch, stained with iodine, photographed through a light microscope(*16*)

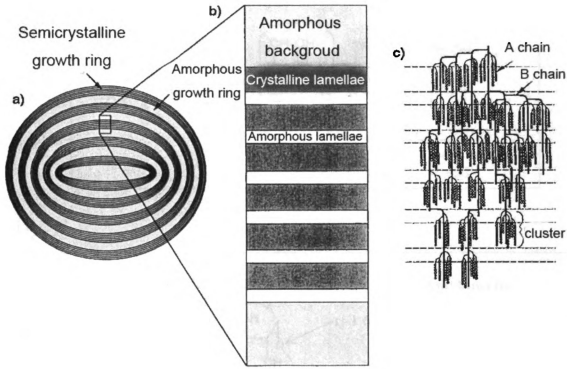
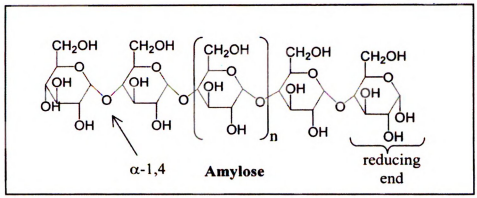


Figure 1. 2 Schematic diagram of starch granule structure. (a) A single granule, comprising concentric rings of alternating amorphous and semicrystalline composition. The semicrystalline growth rings contain stacks of amorphous and crystalline lamellae. (b) Predicting the scattering for a model structure allows us to fit the SAXS data. The model consists of a paracrystalline lamellar stack, embedded in an amorphous background medium. (c) The currently accepted racemose structure for amylopectin within the semicrystalline growth ring. A chain sections of amylopectin form double helices, which are regularly packed into crystalline lamellae. B chains of amylopectin provide intercluster connections. Branching points for both A and B chains of amylopectin are predominantly located within the amorphous lamellae(*12*).

a)



b)

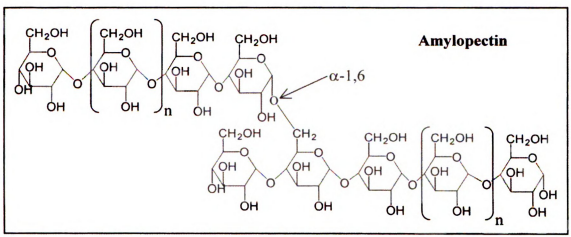


Figure 1. 3 Starch is made of two distinct polysaccharide components: amylose **a)** and **b)** amylopectin (*17*)

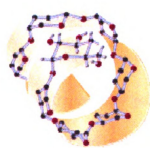


Figure 1. 4 The scale drawing of the structure of the amylose. The glucose units assume a tightly coiled helical structure. Each turn of the helix has 6 glucose units(*2*).

1.1.3 Glycogen / Starch Biosynthesis

Glycogen and starch as major forms of energy storage have also been drawing more and more interest in industrial usage. Starch is one of the most important products synthesized by plants that are used in industrial processes. It is widely used in food, papermaking and the pharmaceutical industries to produce starch hydrolysate, glucose syrups, fructose and cyclodextrins. It is also being researched to replace non-renewable petroleum based energy sources and materials (*18*). But the use of starch is still limited, because its properties are not ideal for a variety of applications. So altering the properties of starch is an important research subject. Although some of the properties can be changed by chemical modification, using chemical processes is not desired or successful under certain circumstances. As an alternative, bioengineering methods provide a proven and fertile avenue for the achievement of these goals (*19-21*). For example, the production of glucose from starch has been converted from an acid hydrolysis method to enzymatic treatment with several different enzymes (**Figure 1.5)** (*22*). Furthermore, the modification of the starch biosynthetic enzymes has enabled the genetic modification of crops in a rational manner to produce novel starches with improved functionality(*23*).

By altering the starch biosynthetic enzymes, it is possible to change the amylose / amylopectin ratio, improve the freeze-thaw property and alter the phosphate content of starch. It can also improve the yield and make the production of starch cheaper for use by industry, or find new applications for the modified starches (*23-24*).

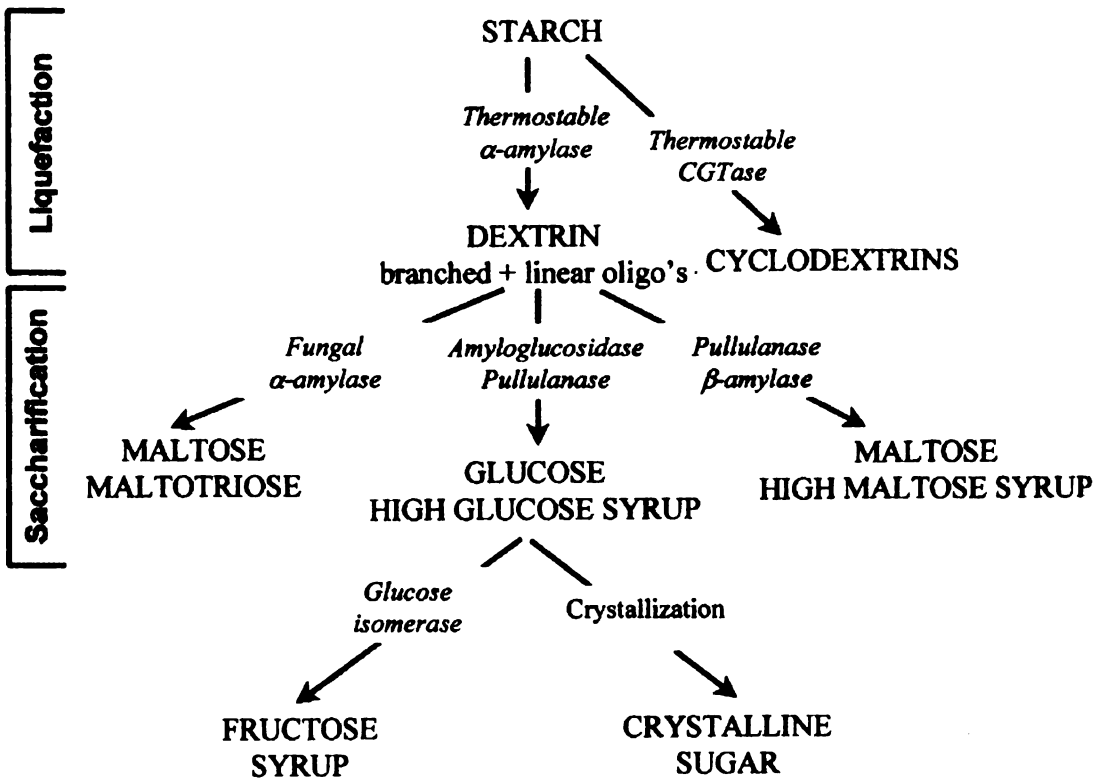


Figure 1. 5 Overview of the industrial processing of starch into cyclodextrins, maltodextrins, glucose or fructose syrups and crystalline sugars (*22*)

Understanding how the biosynthesis of glycogen / starch works *in vivo* was the major challenge. It was found that three enzymes are involved in the three steps of the biosynthetic pathway of bacterial glycogen and starch (**Figure 1.6**). The first step is the formation of ADP-glucose catalyzed by ADP-Glucose Pyrophosphorylase. This is the rate determining step. The formation of the α-1,4 glucosidic linkages occurs first by synthesis of ADP-glucose from ATP and alpha glucose 1-P and then transfer of the glucose moiety from the sugar nucleotide to a pre-existing glucan primer(*25*). The ADP-glucose molecule is the glucose donor in the bacterial and plant cells. Its formation is catalyzed by ADP-Glucose pyrophosphorylase (ADP-Glc PPase), which is an allosteric enzyme (*26-28*). The ADP-Glc PPase from *E.coli* and plants (especially potato tuber) are well

characterized (*26, 29-42*). The activation and the inhibition of the enzyme were well studied. The crystal structure of the enzyme was obtained (*35*).

The elongation of the polysaccharide chain is the second step. It is catalyzed by glycogen or starch synthase. It forms α-1,4 glucosidic bonds between the primer glucose unit and the sugar nucleotide. Branching enzyme catalyzes the last step to make the α-1,6 linked branches.

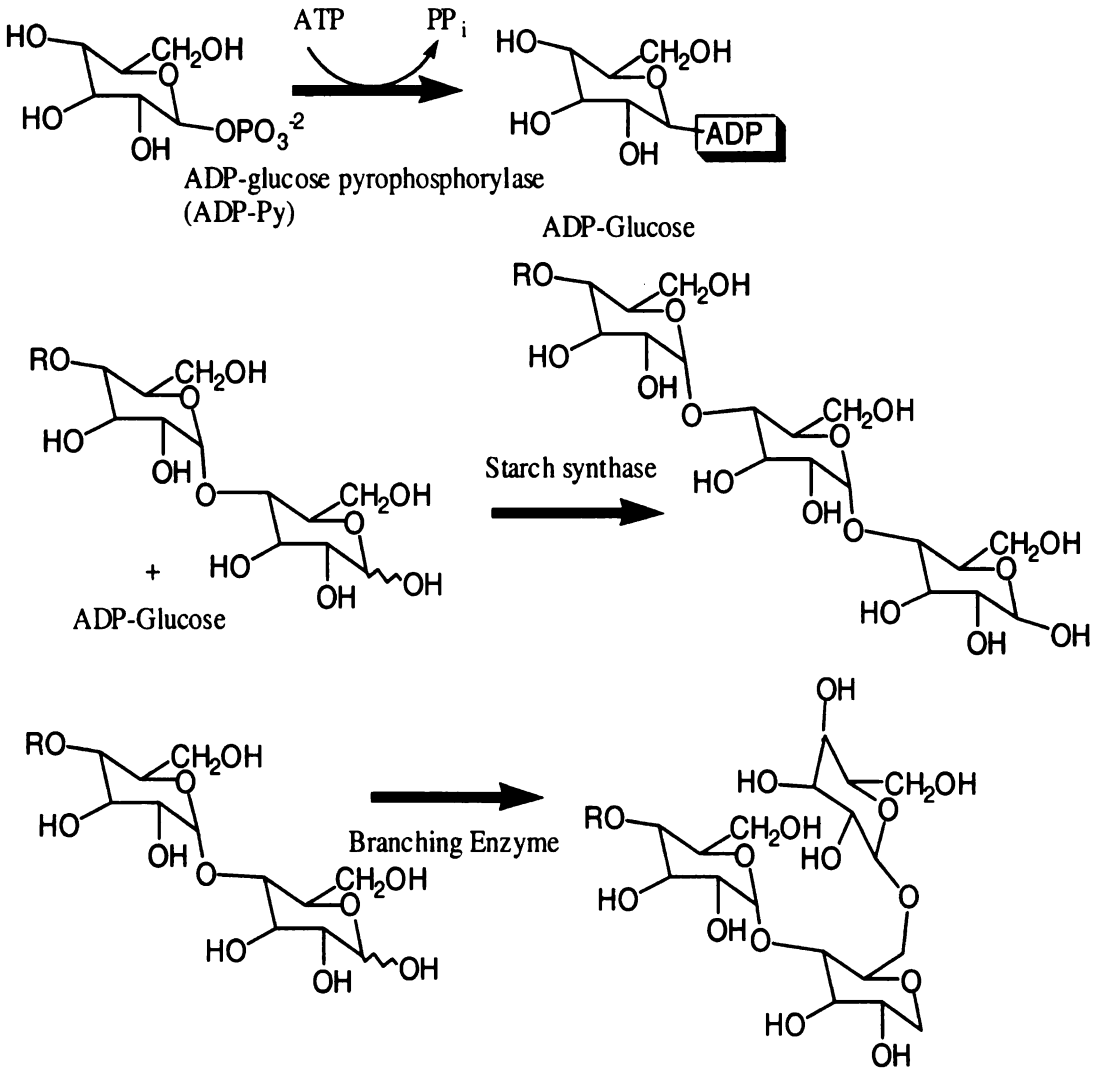


Figure 1. 6 Biosynthetic pathway of glycogen/ starch. There are three enzymes involved in the three consecutive steps: ADP-Glucose pyrophosphorylase, Glycogen or Starch Synthase and Branching Enzyme

Research on how the enzymes work in the pathway will shed light on the fundamental mechanism of glycogen / starch biosynthesis. Furthermore, it will give us more information towards potential starch bioengineering modification strategies.

1.1.4 Branching Enzyme (BE)

Branching Enzyme (1,4-α-glucan 6-glucosyltransferase; EC2.4.1.18) (BE) is found in all organisms that make either starch or glycogen. In the biosynthetic pathway of glycogen/starch, branching enzyme plays an important role in the determination of the final structure of either glycogen or starch. The rearrangement of the linear α-1,4 linked polysaccharide chain is achieved by the cleavage of the α-1,4 glucosidic bond and the subsequent transferring of the oligosaccharide to the α-1,6 position. Forming α-1,6 branch points gives more non-reducing ends to the polysaccharide, thus making it more reactive to synthesis and digestion. Branches are also critical for maintaining its solubility in the cell: exclusively linear glycogen / starch is prone to precipitate in cell while the branched product can maintain appropriate solubility.

BE belongs to the α-amylase family which consists of the amylases, isoamylase, cyclodextrin glucanotransferase (CGT), the pullulanases and branching enzyme. They all act on α-glycosidic bonds and hydrolyze this bond to produce α-anomeric mono- or oligosaccharides (hydrolysis), form α,1-4 or 1-6 glycosidic linkages (transglycosylation) , or a combination of both activities(*10*) (**Figure 1.7**)(*17*). All these enzymes share a common TIM-barrel fold consisting of an 8 β-sheet barrel surrounded 8 α-helices (*10*).

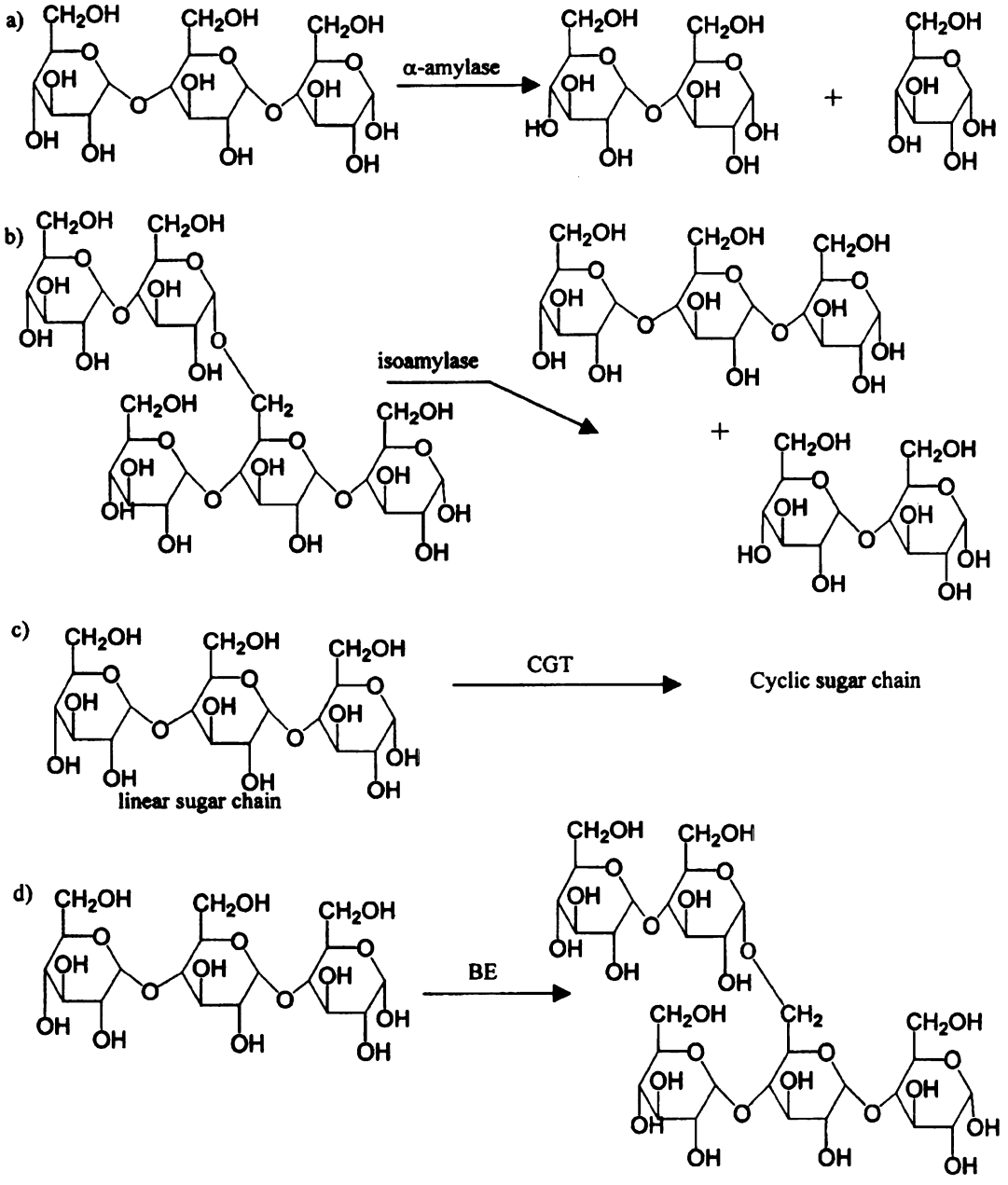


Figure 1. 7 The reactions catalyzed by the members of the α-amylase family. a) α-amylase hydrolyses α-1,4 bonds. b) isoamylase cleaves α-1,6 bonds. c) CGT catalyzes the formation of cyclodextrins and d) BE catalyzes the formation of α-1,6 branches (*17*).

In addition to the central fold containing the catalytic domain, BE also has three other domains not shared with most of the other enzymes of this family, two n-terminal domains and a c-terminal domain (*10, 43*).

BE exists as a single polypeptide in most glycogen-producing organisms (bacteria, fungi, higher mammals, etc.) and at least two forms of the enzyme are found in most plants, starch branching enzyme I (SBE1) and starch branching enzyme II (SBEII)(*44-47*). These enzymes differ in their substrate and branch chain length specificity. SBEII preferentially transfers shorter chains predominantly of lengths of 7 and 11, though there are significant populations of chains between these values (*48*). SBE1 on the other hand, transfers much longer chains. For example, Maize branching enzyme I (MBEI) predominantly transfers polysaccharide chain of 10-13 glucose units (*49-50*). Studies of chimeric forms of SBE1 and SBEII from maize indicated that the c-terminal domain is involved in substrate preference and catalytic capacity, while the n-terminal domain is involved in determining the size of the chain transferred (*43, 51*).

From the amino acid sequence alignment, *Escherichia coli* (*E.coli*) branching enzyme and starch branching enzyme share some critical residues that are involved in the catalytic function (**Figure 1.8**)(*43*)

E. coli	297	S	W	G	Y	Q	P	T	330	L	N	V	I
human	248	S	F	G	Y	Q	I	T	281	I	I	V	L
mBEI	303	S	F	G	Y	H	V	T	336	L	R	V	L
mBEII	277	S	F	G	Y	H	V	T	310	L	L	V	L
isoa	253	Y	W	G	Y	M	T	E	287	I	K	V	Y
a-Asp	79	Y	H	G	Y	W	N	D	112	M	Y	L	M
a-Por	59	W	E	R	Y	Q	P	V	91	V	R	I	Y
CGT	97	Y	H	G	Y	W	A	R	130	I	K	V	I

E. coli	398	G	I	D	A	L	R	V	D	A	V	A	454
human	350	R	F	D	G	F	R	F	D	G	V	T	408
mBEI	408	M	F	D	G	F	R	F	D	G	V	T	466
mBEII	379	K	F	D	G	F	R	F	D	G	V	T	437
isoa	368	G	V	D	G	F	R	F	D	L	A	S	431
a-Asp	199	S	I	D	G	L	R	I	D	T	V	K	226
a-Por	190	G	V	A	G	F	R	L	D	A	S	K	229
CGT	222	G	I	D	G	I	R	M	D	A	V	K	253

E. coli	520	V	L	P	L	S	H	D	E	V
human	475	A	Y	A	E	S	H	D	Q	A
mBEI	475	A	Y	A	E	S	H	D	Q	A
mBEII	503	T	Y	A	E	S	H	D	Q	A
isoa	504	N	F	I	D	V	H	D	G	M
a-Asp	291	T	F	V	E	N	H	D	N	P
a-Por	294	V	F	V	D	N	H	D	N	Q
CGT	322	T	F	I	D	N	H	D	M	E

Figure 1.8 Conserved catalytic residues in *E.coli* branching enzyme, human branching enzyme and maize branching enzyme I and II. Some other enzymes from the α-amylase family are also aligned: isoa is isoamylase from *Pseudomonas amyloderamosa*, a-Asp and a-Por are α-amylase from *Aspergillus Oryzae* and *Porcine Pancreatic*. CGT is cyclodextrin glucanotransferase from *Bascillus Circulans*(*43*). The residues involved in catalysis are highlighted in boxes.

1.1.5 *Escherichia coli* branching enzyme (*E.coli* BE)

The full length wild type *E.coli* BE was over-expressed and purified. It has 728 amino acid residues with a molecular weight of 84KDa. In order to gain insight into the structure, mechanism and specificity of *E.coli* BE, crystallization was attempted on the wild type full length enzyme without success. Further study indicated a 70KDa limited proteolysis product still retained 40-60% of the branching enzyme activity(*52*). The sequence of the truncated enzyme was determined. It lacks the first 113 amino acid residues due to the limited proteolysis. This construct of enzyme (named as N113BE) was purified and crystallized successfully. The three dimensional structure of the protein was reported (*43, 53*).

The determination of the structure of *E. coli* BE was a milestone in the study of glycogen / starch biosynthetic pathway enzymes. It was the first structure of a BE from any organism and also was the first structure of any of the enzymes that make up the starch biosynthetic pathway.(*43*) Figure 1.9 shows the ribbon depiction of the truncated *E. coli* BE structure with the second n-terminal domain in red and the c-terminal domain in blue. Studies of the *E. coli* BE have shown that the enzyme predominantly transfers shorter chains of between 7-15 glucose units in length. However, removal of the n-terminal 113 amino acid residues had only a moderate effect on catalytic activity but it significantly altered the transfer-chain preference to longer chains containing 15-25 glucose units. Substantial amounts of chains as long as 30-40 residues were transferred

with this mutant, indicating that the far n-terminus of *E. coli* BE has a significant effect on branch chain specificity(*54-55*). The overall structure of the enzyme was discussed in detail; and an acarbose molecule was modeled into the active site of the enzyme to elucidate the possible interaction between the enzyme and its substrate(*43*). In our work, we obtained the substrate bound E.coli BE structures and proposed a hypothesis on how the enzyme binds with its substrate prior to the catalytic action. The result and discussion are in the following chapters.

Figure 1. 9 Ribbon depiction of the x-ray crystal structure of *E. coli* BE truncated at amino acid 113. Residues involved in BE catalysis are shown in *green*, with atoms colored by type: *red*, oxygen; *green*, carbon; *blue*, nitrogen. *Red* indicates the NH_2-terminal domain; *orange* indicates the central α/β barrel catalytic domain; and *blue* indicates the COOH-terminal domain.

1.2 Transcriptional complex TFIIIB/DNA

1.2.1 Transcription

Transcription is the process by which the genetic message in DNA is transcribed into RNA. DNA is the genetic material of all living organisms except certain viruses(*56*). The genetic information coded in the double stranded DNA is transferred to RNA in an accurate multi-step process. This process is accomplished by synthesis of RNA by RNA polymerases with the help of some cofactors. In eukaryotic cells there are three different transcription systems. Each consists of one kind of RNA polymerase and a unique set of transcription factors. RNA polymerase I (pol I) transcribes large ribosomal RNA (rRNA) genes (35s-45s depending on species); RNA pol II, located in the nucleoplasm, synthesizes messenger RNA (mRNA) genes and most of the small nuclear RNA (snRNA) (e.g. U1,U2) genes while RNA pol III is responsible for transcribing all transfer RNA (tRNA), 5s rRNA and some other snRNA genes(*56-60*). The accurate initiation of any transcription requires RNA polymerase in association with the transcription factors to be recruited to the specific start sequence of DNA that is called the promoter. In the three stages of transcription (initiation, elongation and termination), the study of initiation could reveal how transcription starts and give answers to how the polymerases and transcription factors interact on the specific promoters.

The genes transcribed by RNA polymerase III (pol III) encode a variety of small RNA molecules, many of which have essential functions in cellular metabolism. For

example, tRNA and 5sRNA are required in the synthesis of protein; 7sLRNA is involved in intracellular protein transportation while U6 and H1 RNAs are involved in the post-transcriptional process (*57, 59-60*). RNA polymerase III transcribes some genes with unknown functions as well. The genes transcribed by RNA polymerase III are mainly shorter than 400 base pairs. This limit is consistent with the elongation properties of RNA polymerase III, which recognizes a simple run of T residues as a termination signal(*57*).

So far TFIIIB mediated RNA pol III transcription is the only eukaryotic transcription system that can be reconstituted using only recombinant factors. And this is only achieved in *S. Cerevisiae* where TFIIIB binds with the TATA box directly (*57, 61*). So the TFIIIB mediated pol III transcription becomes a favorite topic.

1.2.2 RNA polymerase III promoters

There are three types of promoters for RNA polymerase III called type 1-3 (**Figure 1.8**)(*57*). The 5s RNA gene from *Xenopus laevis* is the only type 1 promoter. It is the primary binding site of TFIIIA. The structure characteristic of the promoter is the unique box A sequence, followed by the intermediate element (IE) and a unique conserved sequence box C. All of these elements constitute the internal control region (ICR)(*62-63*).

The Adenovirus 2 (Ad2) VAI gene and various tRNA genes from *Xenopus laevis* and *Drosophila melanogaster* are typical type 2 promoters. Type 2 promoters have two highly

conserved internal promoter elements, box A and B(*64*).

The type 3 promoters were found in mammalian U6 snRNA genes, which encode the snRNA component of the spliceosome(*65*). Also it is found in the human 7SK gene which is involved in regulation of the CDK9/cyclin complex(*66*).

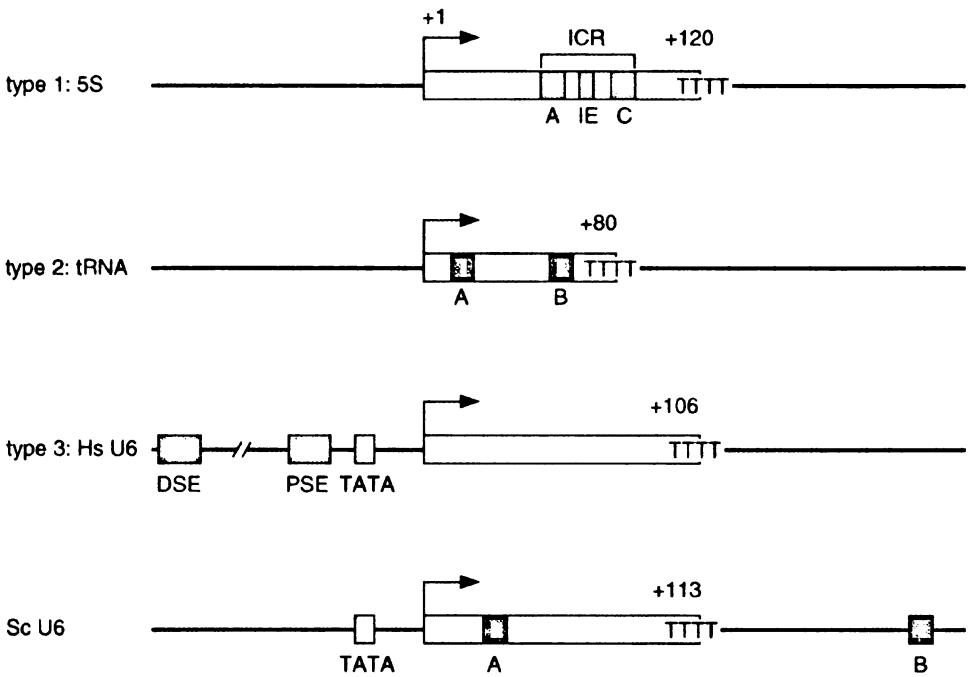


Figure 1. 10 Different types of RNA pol III promoters: Human U6 (Hs U6) and *S.Cerevisiae* U6 (Sc U6) are two examples of type 3 promoters. +1 indicates the starting point of transcription; TATA is the TATA box region; DSE and PSE stand for distal sequence element and proximal sequence element which are unique sequence regions(*57*).

1.2.3 TFIIIB transcription factor

In order to initiate transcription accurately and effectively, the RNA polymerases need the assistance of transcription factors to recruit them to the appropriate start sites of genes. The transcription factors recognize the specific promoter and bind with DNA, then load the corresponding polymerase onto the DNA so that the transcription can initiate. This process involves the interaction between transcription factor and DNA promoter, the interaction between transcription factor and polymerase and the interaction between DNA and polymerase. It is obviously a very complicated process. In the study of this process, different transcription factors were discovered. For example, the yeast RNA polymerase III transcription machinery consists of three transcription initiation factors: TFIIIA, which is a gene-specific factor, TFIIIB and TFIIIC which are general factors for RNA pol III (*67-68*).

Among all the transcription factors, TFIIIB is the central initiation factor since it alone can recruit pol III to the transcription starting site and initiate the transcription with polymerase III (*57, 67, 69*). Besides this function, TFIIIB was also found to participate in promoter opening steps (*70-71*). All of these make TFIIIB the research focus recently.

TFIIIC and TFIIIA, the other two components of the core transcription apparatus, were identified by the separation of cell extract through phosphocellulose chromatographic separation with TFIIIB together. They bind with DNA and sometimes serve as assembly factors for TFIIIB in the recognition of specific genes (*72-73*). TFIIIC's function of loading TFIIIB is already well known. But little is known about how

TFIIIA helps to load TFIIIC first(*67*).

An example of interactions between TFIIIB, TFIIIC and DNA in TFIIIC-dependent transcription is represented by **Figure. 1.9**

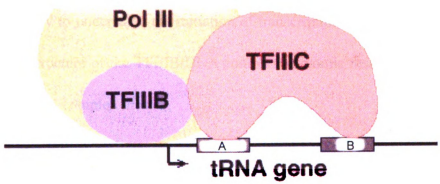


Figure 1. 11 A schematic model of TFIIIC-dependent TFIIIB binding to tRNA promoter(*74*).

It has been demonstrated that TATA-Binding Protein (TBP) can bind DNA segment with a TATAAA sequence (TATA box) by itself *in vitro (71, 75-77)*. And the recruitment of pol III could happen without the help of TFIIIC and TFIIIA. While TFIIIC and TFIIIA are unable to recruit pol III to a promoter without TFIIIB in any context, which also proves that TFIIIB is the core transcription factor for pol III(*68, 78-80*).

However in most budding yeast pol III transcribed genes do not have strong TATA boxes near the promoter region. In this case TFIIIC acts as the assembly factor for TFIIIB to load TFIIIB to the appropriate binding site (*75, 81*). This is the so-called TFIIIC-dependent binding. Whether assembled at the promoter by TFIIIC or

independently, TFIIIB alone suffices to recruit pol III for multiple rounds of transcription and plays an essential role in formation of the open pol III initiation complex. The structure of the pre-initiation complex of TFIIIB and DNA can reveal how the transcription factor interacts with the promoter at the very first moment of transcription. It is the key to understand the initiation of transcription. In our project we try to get the detailed structure of the TFIIIB/DNA complex at atomic resolution by reconstituting the pre-initiation complex *in vitro.*

1.2.4 Architecture of TFIIIB:

TFIIIB transcription factor is a multi-subunit protein. It consists of three subunit: TATA-binding protein (TBP), TFIIB-related factor (Brf1, also called TFIIIB70) and B" (also called Bdp1 or TFIIIB90). Each subunit has been identified and sequenced.

In order to understand how TFIIIB initiates transcription on promoter, the relationship and individual functions of the TBP, Brf and B" have to be investigated. As the pol III transcription core apparatus, TFIIIB undergoes both TFIIIC-dependent and TFIIIC-independent binding to DNA. **Figure 1.10** illustrates the binding process in both cases(*57*).

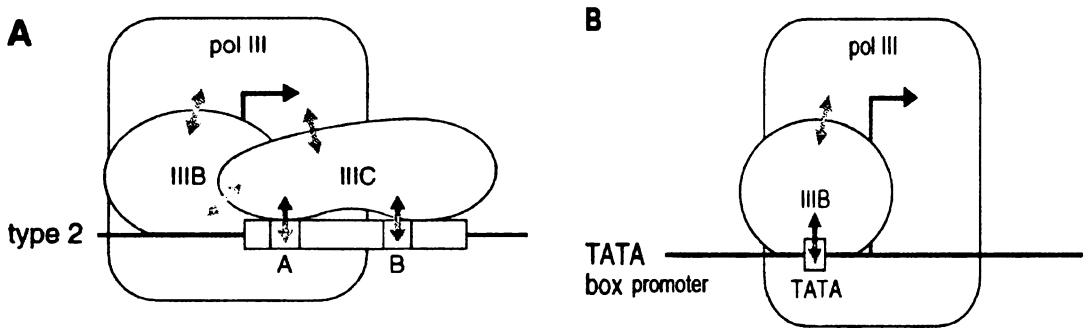


Figure 1. 12 A schematic representation of TFIIIB transcription complexes. There are two types of complex formation: **A)** TFIIIC-dependent pathway on a type 2 promoter; **B)** TFIIIC-independent pathway on an artificial TATA box containing promoter. Arrows stand for interaction among different components (*57*).

Fig 1.10 shows the mechanism of RNA pol III initiation *in vitro*. Figure A shows that TFIIIC recruits TFIIIB onto the TATA-less DNA sequence, followed by pol III binding and initiation. And Figure B depicts that TFIIIB is recruited by the binding of TBP subunit to TATA box, followed by pol III binding and initiation.

1.2.4.1 TBP protein

It is well known that the TATA-binding protein (TBP) is an essential component of the transcription machinery of all three RNA polymerase, pol I , pol II and pol III systems(*76-77, 81-82*). Thus it is required in the expression of all nuclear genes (*75, 83-84*). TBP protein binds the TATA box in a TFIIIC-independent binding situation, (*85*).

The C-terminal domain of TBP is highly conserved. This region shows transcriptional activity *in vitro*. And the site –directed mutagenesis or internal deletion in this region results in a dramatic decrease in biological activity (*86-87*). The N-terminus of TBP is irrelevant to the transcriptional activity in pol III transcription. In our project we are going to investigate the specific region that is absolutely required in forming stable pre-initiation complex and directing subsequent transcriptions(*67*). The interactions between TBP and the other two subunits, Brf1 and B", are to be studied as well.

The full length TBP protein from *S. cerevisiae* has 240 amino acid residues. Truncated construct (61-240) of this protein was used in the crystallization with TATA box promoter(*88*). We also used this protein for our studies. **Figure 1.11** shows the protein sequence of TBP from *S. cerevisiae.*

```
  1 MADEERLKEFKEANKIVFDPNTRQVWENQNRDGTKPATTFQSEEDIKRAA

 51 PESEKDTSATSGIVPTLQNIVATVTLGCRLDLKTVALHARNAEYNPKRFA

101 AVIMRIREPKTTALIFASGKMVVTGAKSEDDSKLASRKYARIIQKIGFAA

151 KFTDFKIQNIVGSCDVKFPI RLEGLAFSHGTFSSYEPELFPGLIYRMVKP

201  KIVLLIFVSG KIVLTGAKQR EEIYQAFEAI YPVLSEFRKM
```

Figure 1. 13 Amino acid sequence of *S. cerevisiae* TBP protein.

1.2.4.2 Brf1 protein

Brf1 is the TFIIB-related factor of the TFIIIB complex. It is a multi-domain protein with 596 amino acid residues. The molecular weight of Brf1 is about 70 KDa. The architecture of Brf 1 is represented by **Fig 1.12**.

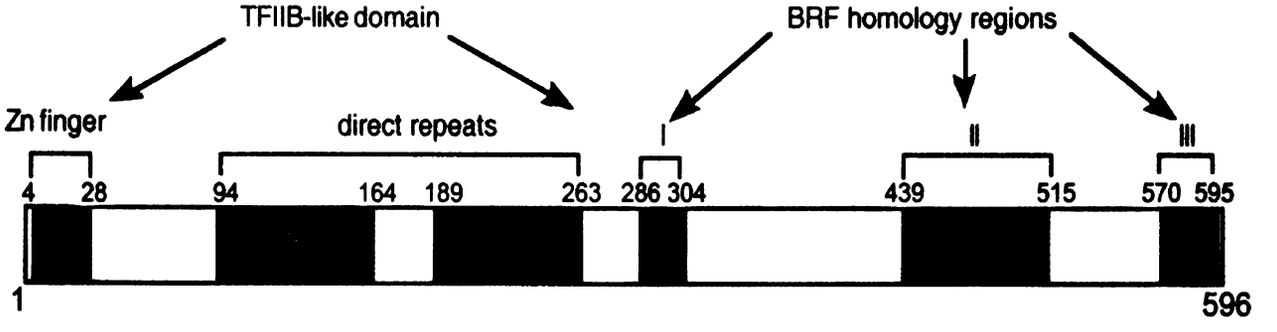


Figure 1. 14 A schematic graph of *S. Cerevisiae* Brf1. TFIIB homologous region is represented by orange color, blue region represents conserved region for Brf(*57, 74*)

The Brf1 protein has a putative Zn finger domain and two imperfect direct repeat domains which are homologous to TFIIB. The three conserved domains are only found in BRF proteins (*89-91*). The N-terminal half of the protein shows 19% amino acid identity to TFIIB. It was reported that the N-terminal region of Brf interacts with the C34 subunit of RNA polymerase III(*92*). Research shows that the N-terminus retains the ability to recruit B" to the TBP/DNA complex and to direct efficient transcription on supercoiled DNA. But the binding with TBP and B" is not mainly stabilized by this half of the protein(*79*). Although the N-terminal half can also form TFIIIB-DNA complexes at

strong TATA box promoters and recruit pol III and direct multi round transcription as well, the stability of the complexes is so low that the TFIIIB-DNA complexes assembled with the N-terminal half of Brf1 alone can not be detected by native electrophoresis gel(*84*). And in the case of TFIIIC-dependent binding of TFIIIB, the N-terminus of Brf1 also has contact with subunits of TFIIIC (*79, 91*).

On the other hand, the ~ 30 KD C-terminal half which is only conserved among BRF is of great importance in the interaction with TBP and DNA. The C-terminal half of Brf1, which is not conserved in TFIIB, but homologous among BRF proteins, shows strong binding affinity with TBP, B" and DNA(*90-91*). The c-terminus itself can bind TBP, DNA and B" *in vitro* and the complex is stable enough for transcriptional assays(*93*). Internal deletion in Brf homology regions II or III eliminates the binding ability. But even the removal of the entire TFIIB homology region and the fungal homology region I of Brf1 doesn't entirely eliminate the ability to incorporate B" into a TFIIIB/DNA complex(*90*).

It has been reported that the N-terminus of Brf1 is required for directing transcription. But the C-terminal Brf1 itself (especially the region around 435-545) is sufficient for assembling the TFIIIB/DNA recombinant complex(*91*). When this domain is reconstituted with the N-terminal half(1-282) of Brf1, almost full wild-type Brf1 activity is recovered in both TFIIIC-dependent and TFIIIC-independent transcription. The protein interaction between Brf1 and TBP in the TBP/DNA/Brf complex is mainly between the C-terminus of Brf1 and the N-terminal top and stirrup of TBP (*90-91*).

The presence of Brf1 is also essential for the loading of B" to the TBP-DNA

complex. It has been observed that without Brf1 B" can not bind with TBP-DNA stably (*94-95*). The C-terminal mutant 284-596 can hold all of B" tightly to the TBP/DNA complex. So wild type Brf (and Brf 435-545 as well) serve as a 'bridge" between the TBP/DNA complex and B" in the reconstituted TFIIIB/DNA complex (*91, 96*).

1.2.4.3 B" protein

B" is the third subunit of TFIIIB. The subunit is also called TFIIIB90 or bdp1. In the TFIIIB complex, TBP and Brf1 by themselves are not competent to direct transcription initiation by RNA pol III either *in vitro* or *in vivo*. B" is required for transcription in duplex DNA or chromatin. But in order to form the pre-initiation complex, Brf has to be present. B" cross-links very weakly to DNA in TBP/DNA in the absence of Brf or Brf mutants(*96*). In other words, B" can bind with TBP/Brf/DNA, but not TBP/DNA. B" seems to serve as a "scaffold" for holding Brf in the TFIIIB/DNA complex (*95-96*). The assembly of the complete pre-initiation complex reconfigures Brf protein inside the complex. And the recruitment of B" also makes the TFIIIB/DNA complex much more stable (*71, 94*).

In yeast *S. cerevisiae*, B" consists of 594 amino acid residues. It is reported that two domains, 272-292 and 424-449, are required for TFIIIC-dependent transcription and at least one of them is required for TFIIIC-independent transcription. This is consistent

with the finding that B" without the N-terminal 262 amino acid residues or C-terminal 130 amino acid residues remain competent to assemble into stable TFIIIB-DNA complex via a TFIIIC-dependent assembly pathway on SUP4 genes(*96*). It was concluded that the central ~225 amino acid residues of B" appear to encompass the functional core of the protein(*57*). B" has a putative DNA binding SANT (SWI3, ADA2, N-CoR and TFIIIB) domain that is homologous among the families. **Figure 1.13** illustrates the SANT domain alignment in human and yeast *S.cerevisiae.* **Figure 1.14** shows the primary sequence of *S. cerevisiae* B" protein.

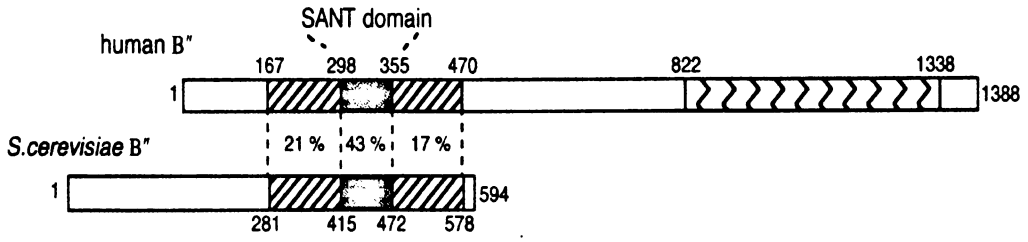


Figure 1. 15 The similarity in the human and *S. cerevisiae* B'' sequence. The orange box corresponds to the SANT domain.

```
  1 MSSIVNKSGTRFAPKVRQRRAATGGTPTPKPRTPQLFIPESKEIEEDNSD
 51 NDKGVDENETAIVEKPSLVGERSLEGFTLTGTNGHDNEIGDEGPIDASTQ
101 NPKADVIEDNVTLKPAPLQTHRDQKVPRSSRLASLSKDNESRPSFKPSFL
151 DSSSNSNGTARRLSTISNKLPKKIRLGSITENDMNLKTFKRHRVLGKPSS
201 AKKPAGAHRISIVSKISPPTAMTDSLDRNEFSSETSTSREADENENYVIS
251 KVKDIPKKVRDGESAKYFIDEENFTMAELCKPNFPIGQISENFEKSKMAK
301 KAKLEKRRHLRELPMRARQEFKPLHSLTKEEQEEEEEKRKEERDKLLNAD
351 IPESDRKAHTAIQLKLNPDGTMAIDEETMVVDRHKNASIENEYKEKVDEN
401 PFANLYNYGSYGRGGYTDPWTVEEMIKFYKALSMWGTDFNLISQLYPYRS
451 PKQVKAKFVNEEKKPPILIELALRSKLPPNFDEYCCEIKKNIGTVADFNE
501 KLIELQNEHKHHMKEIEEAKNTAKEEDQTAQRLNDANLNKKGSGGIMTND
551 LKVYRKTEVVLGTIDDLKRKKLKERNNDDNEDNEGSEEEPEIDQ
```

Figure 1. 16 The amino acid sequence of *S.cerevisiae* B". Region 415-472 is labeled in red and region 329-357 which is the SANT domain is labeled in blue color.

It has been shown that B" binds predominantly with ~10 base pairs upstream of the TBP binding site TATA box. In this region, B" and Brf share an extended overlapped interface with DNA. Upon binding with TBP/Brf/DNA, B" extended its binding ~15 base pairs upstream of TATA box. An additional 15-20 base pairs upstream will further stabilize the complex but is not absolutely required(*78, 97*). It is also suggested that the addition of B" in the TFIIIB-DNA complex induces a bend in the DNA conformation

between the TATA box and the transcriptional starting site. The bending of DNA is postulated to contribute to the stabilization of the TFIIIB-DNA complex through extended interaction between DNA and subunits of TFIIIB (*98-100*).

1.2.5 Oligonucleotide sequence

From the previous introduction on the three subunits of TFIIIB, it is clear that the DNA sequences to be investigated should have certain base pairs upstream of the TBP binding site (TATA box) and a segment downstream of the TATA box.

The TBP protein binds with the TATA box which is numbered around -25 to -30 on the promoter in most TATA containing DNA sequence(*101*).

Brf requires 12-15 base pairs downstream of TATA box to form a stable TBP/Brf/DNA complex (*78, 101*).

It is reported that two regions of B", 291-310 and 426-487, contact the DNA upstream and downstream of the TATA box respectively upon loading by Brf protein (*96, 101*). And at least 10 base pairs immediately downstream of the TATA box are needed for B" binding (*96*).

From the information listed above, we can see that formation of stable TFIIIB/DNA complexes prefers ~15 bp upstream and ~10 bp downstream of the TBP binding site. This is a quite long DNA strand which might be detrimental to crystallization of the complex

Also from the recent results, TBP/Brf/DNA complex was constructed with truncated version of TBP and Brf. It has been found that a very short DNA sequence was used to form the complex stably with the truncated TBP and Brf proteins(*102*). This gave us a hint that shorter oligonucleotides might be plausible in forming complexes with truncated TFIIIB subunits. So a series of oligonucleotides with different lengths will be screened in the TFIIIB-DNA complex study.

The structures of binary TBP/DNA and tertiary TBP/Brf/DNA complexes have been solved recently (*88, 102*). The detailed atomic structures helped us to understand the basic interaction among TBP, Brf protein and specific promoters. In order to shed light on the transcription mechanism which will benefit extensive biological and medical research (*103-104*), the structure of the complete pre-initiation quaternary complex TBP/Brf/B"/DNA must be obtained.

Reference cited

1. Beckmann, C. O. (1953) The structure of amylopectin and glycogen, *Ann N Y Acad Sci 57*, 384-397.

2. Lehninger, L. A. N. L. D. C., M. Michael (1993) *Principles of Biochemistry*, Second Edition ed., Worth Publishers, Inc., New York.

3. Ball, S. G., and Morell, M. K. (2003) From bacterial glycogen to starch: Understanding the biogenesis of the plant starch granule, *Annual Review of Plant Biology 54*, 207-233.

4. Ozen, H. (2007) Glycogen storage diseases: New perspectives, *World J Gastroentero 13*, 2541-2553.

5. Bell, D. J. (1936) The molecular structure of glycogen from the whole tissues of Mytilus edulis, *Biochem J 30*, 2144-2145.

6. Bell, D. J. (1937) Glycogen: The molecular structure of horse muscle glycogen, *Biochem J 31*, 1683-1691.

7. Dumazert, C. (1950) [New model of the structure of starch and glycogen.], *Bull Soc Chim Biol (Paris) 32*, 998-1008.

8. Preiss, J., and Sivak, M. N. (1998) Biochemistry, molecular biology and regulation of starch synthesis, *Genet Eng (N Y) 20*, 177-223.

9. Smith, A. M. (1999) Making starch, *Current Opinion in Plant Biology 2*, 223-229.

10. 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.

11. Jenkins, P. J., and Donald, A. M. (1995) The influence of amylose on starch

granule structure, *Int J Biol Macromol 17*, 315-321.

12. Jenkins, P. J., Cameron, R. E., Donald, A. M., Bras, W., Derbyshire, G. E., Mant, G. R., and Ryan, A. J. (1994) In-Situ Simultaneous Small and Wide-Angle X-Ray-Scattering - a New Technique to Study Starch Gelatinization, *J Polym Sci Pol Phys 32*, 1579-1583.

13. Hinrichs, W., Buttner, G., Steifa, M., Betzel, C., Zabel, V., Pfannemuller, B., and Saenger, W. (1987) An Amylose Antiparallel Double Helix at Atomic Resolution, *Science 238*, 205-208.

14. 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.

15. Helbert, W., and Chanzy, H. (1994) Single-Crystals of V-Amylose Complexed with N-Butanol or N-Pentanol - Structural Features and Properties, *Int J Biol Macromol 16*, 207-213.

16. Yuri, K. (2006) image of wheat starch granules , , (wheat starch granules stained with iodine, p. f. l. m., Ed.), http://en.wikipedia.org/wiki/File:Wheat_starch_granules.JPG.

17. Abad, M. (2002) the X-ray crystallographic structures of angiostatin and branching enzyme, In *Department of Chemistry*, p 127, Michigan State University, East Lansing.

18. Tally, S. (2002) Making biofuel from corn stover, *Biocycle 43*, 62-62.

19. Sticklen, M., Mei, C. S., Ransom, C., Park, S. H., Sabzikar, R., and Qi, S. (2008) Modifying the corn genome to improve its biomass biofuel production, *In Vitro Cell Dev-An 44*, S29-S29.

20. Smidansky, E. D., Clancy, M., Meyer, F. D., Lanning, S. P., Blake, N. K., Talbert, L. E., and Giroux, M. J. (2002) Enhanced ADP-glucose pyrophosphorylase activity in wheat endosperm increases seed yield, *P Natl Acad Sci USA 99*,

1724-1729.

21. 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, *Nat Biotechnol 20*, 1256-1260.

22. Crabb, W. D., and Shetty, J. K. (1999) Commodity scale production of sugars from starches, *Curr Opin Microbiol 2*, 252-256.

23. Jobling, S. (2004) Improving starch for food and industrial applications, *Current Opinion in Plant Biology 7*, 210-218.

24. Kossmann, J., and Lloyd, J. (2000) Understanding and influencing starch biochemistry, *Crit Rev Biochem Mol Biol 35*, 141-196.

25. Preiss, J., Yung, S. G., and Baecker, P. A. (1983) Regulation of bacterial glycogen synthesis, *Mol Cell Biochem 57*, 61-80.

26. Ballicora, M. A., Iglesias, A. A., and Preiss, J. (2004) ADP-Glucose Pyrophosphorylase: A Regulatory Enzyme for Plant Starch Synthesis, *Photosynth Res 79*, 1-24.

27. Preiss, J. (1996) ADPglucose pyrophosphorylase: basic science and applications in biotechnology, *Biotechnol Annu Rev 2*, 259-279.

28. Ballicora, M. A., Iglesias, A. A., and Preiss, J. (2003) ADP-glucose pyrophosphorylase, a regulatory enzyme for bacterial glycogen synthesis, *Microbiol Mol Biol Rev 67*, 213-225, table of contents.

29. Iglesias, A. A., Barry, G. F., Meyer, C., Bloksberg, L., Nakata, P. A., Greene, 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.

30. Greene, T. W., Chantler, S. E., Kahn, M. L., Barry, G. F., Preiss, J., and Okita, T.

W. (1996) Mutagenesis of the potato ADPglucose pyrophosphorylase and characterization of an allosteric mutant defective in 3-phosphoglycerate activation, *Proc Natl Acad Sci U S A 93*, 1509-1513.

31. Ballicora, M. A., Fu, Y., Nesbitt, N. M., and Preiss, J. (1998) ADP-Glucose pyrophosphorylase from potato tubers. Site-directed mutagenesis studies of the regulatory sites, *Plant Physiol 118*, 265-274.

32. Fu, Y., Ballicora, M. A., Leykam, J. F., and Preiss, J. (1998) Mechanism of reductive activation of potato tuber ADP-glucose pyrophosphorylase, *J Biol Chem 273*, 25045-25052.

33. Ballicora, M. A., Frueauf, J. B., Fu, Y., Schurmann, P., and Preiss, J. (2000) Activation of the potato tuber ADP-glucose pyrophosphorylase by thioredoxin, *J Biol Chem 275*, 1315-1320.

34. 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 Lett 573*, 99-104.

35. 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.

36. Bejar, C. M., Ballicora, M. A., Iglesias, A. A., and Preiss, J. (2006) ADPglucose pyrophosphorylase's N-terminus: structural role in allosteric regulation, *Biochem Biophys Res Commun 343*, 216-221.

37. Copeland, L., and Preiss, J. (1981) Purification of Spinach Leaf ADPglucose Pyrophosphorylase, *Plant Physiol 68*, 996-1001.

38. Morell, M. K., Bloom, M., Knowles, V., and Preiss, J. (1987) Subunit Structure of Spinach Leaf ADPglucose Pyrophosphorylase, *Plant Physiol 85*, 182-187.

39. Ball, K., and Preiss, J. (1994) Allosteric sites of the large subunit of the spinach leaf ADPglucose pyrophosphorylase, *J Biol Chem 269*, 24706-24711.

40. Dickinson, D. B., and Preiss, J. (1969) ADP glucose pyrophosphorylase from maize endosperm, *Arch Biochem Biophys 130*, 119-128.

41. Plaxton, W. C., and Preiss, J. (1987) Purification and Properties of Nonproteolytic Degraded ADPglucose Pyrophosphorylase from Maize Endosperm, *Plant Physiol 83*, 105-112.

42. Giroux, M., Smith-White, B., Gilmore, V., Hannah, L. C., and Preiss, J. (1995) The large subunit of the embryo isoform of ADP glucose pyrophosphorylase from maize, *Plant Physiol 108*, 1333-1334.

43. 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, *J. Biol. Chem. 277*, 42164-42170.

44. Guan, H. P., Baba, T., and Preiss, J. (1994) Expression of Branching Enzyme-Ii of Maize Endosperm in Escherichia-Coli, *Cell Mol Biol 40*, 981-988.

45. 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.

46. Guan, H. P., and Preiss, J. (1993) Differentiation of the Properties of the Branching Isozymes from Maize (Zea mays), *Plant Physiol 102*, 1269-1273.

47. Kuriki, T., Guan, H. P., and Preiss, J. (1994) Structure and Function of Starch Branching Enzyme, *Nippon Nogeik Kaishi 68*, 1581-1584.

48. Hernandez, J. M., Gaborieau, M., Castignolles, P., Gidley, M. J., Myers, A. M., and Gilbert, R. G. (2008) Mechanistic investigation of a starch-branching enzyme using hydrodynamic volume SEC analysis, *Biomacromolecules 9*, 954-965.

49. Guan, H. P., Li, P., ImparlRadosevich, 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.

50. 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.

51. 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, *J Biol Chem 272*, 28999-29004.

52. Binderup, K., Mikkelsen, R., and Preiss, J. (2000) Limited proteolysis of branching enzyme from Escherichia coli, *Archives of Biochemistry and Biophysics 377*, 366-371.

53. Abad, M. C., Binderup, K., Preiss, J., and Geiger, J. H. (2002) Crystallization and preliminary X-ray diffraction studies of Escherichia coli branching enzyme, *Acta Crystallogr D 58*, 359-361.

54. Binderup, M., 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.

55. 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.

56. lewin, B. (2000) *Genes VII*, Oxford University Press.

57. Schramm, L., and Hernandez, N. (2002) Recruitment of RNA polymerase III to its target promoters, *Gene Dev 16*, 2593-2620.

58. Klekamp, M. S., and Weil, P. A. (1987) Properties of Yeast Class-Iii Gene-Transcription Factor Tfiiib - Implications Regarding Mechanism of Action, *J Biol Chem 262*, 7878-7883.

59. Reznikoff, W. S., Siegele, D. A., Cowing, D. W., and Gross, C. A. (1985) The Regulation of Transcription Initiation in Bacteria, *Annu Rev Genet 19*, 355-387.

60. Willis, I. M. (1993) Rna Polymerase-Iii - Genes, Factors and Transcriptional

Specificity, *Eur J Biochem 212*, 1-11.

61. Huang, Y., and Maraia, R. J. (2001) Comparison of the RNA polymerase III transcription machinery in Schizosaccharomyces pombe, Saccharomyces cerevisiae and human, *Nucleic Acids Res 29*, 2675-2690.

62. Bogenhagen, D. F., Sakonju, S., and Brown, D. D. (1980) Control Region in the Center of the 5s Rna Gene Directs Specific Initiation of Transcription .2. 3' Border of the Region, *Cell 19*, 27-35.

63. Sakonju, S., Bogenhagen, D. F., and Brown, D. D. (1980) Control Region in the Center of the 5s Rna Gene Directs Specific Initiation of Transcription .1. 5' Border of the Region, *Cell 19*, 13-25.

64. Galli, G., Hofstetter, H., and Birnstiel, M. L. (1981) 2 Conserved Sequence Blocks within Eukaryotic Transfer-Rna Genes Are Major Promoter Elements, *Nature 294*, 626-631.

65. Krol, A., Carbon, P., Ebel, J. P., and Appel, B. (1987) Xenopus-Tropicalis U6-Snrna Genes Transcribed by Pol-Iii Contain the Upstream Promoter Elements Used by Pol-Ii Dependent U-Snrna Genes, *Nucleic Acids Res 15*, 2463-2478.

66. Nguyen, V. T., Kiss, T. S., Michels, A. A., and Bensaude, O. (2001) 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes, *Nature 414*, 322-325.

67. Kassavetis, G. A., Letts, G. A., and Geiduschek, E. P. (1999) A minimal RNA polymerase III transcription system, *Embo J 18*, 5042-5051.

68. Kassavetis, G. A., and Geiduschek, E. P. (2006) Transcription factor TFIIIB and transcription by RNA polymerase III, *Biochem Soc T 34*, 1082-1087.

69. Joazeiro, C. A. P., Kassavetis, G. A., and Geiduschek, E. P. (1994) Identical Components of Yeast Transcription Factor-Iiib Are Required and Sufficient for Transcription of Tata Box-Containing and Tata-Less Genes, *Mol Cell Biol 14*, 2798-2808.

70. Kassavetis, G. A., Kumar, A., Letts, G. A., and Geiduschek, E. P. (1998) A post-recruitment function for the RNA polymerase III transcription-initiation factor IIIB, *P Natl Acad Sci USA 95*, 9196-9201.

71. Kumar, A., Grove, A., Kassavetis, G. A., and Geiduschek, E. P. (1998) Transcription factor IIIB: The architecture of its DNA complex, and its roles in initiation of transcription by RNA polymerase III, *Cold Spring Harb Sym 63*, 121-129.

72. Deprez, E., Arrebola, R., Conesa, W., and Sentenac, A. (1999) A subunit of yeast TFIIIC participates in the recruitment of TATA-binding protein, *Mol Cell Biol 19*, 8042-8051.

73. Rozenfeld, S., and Thuriaux, P. (2001) Genetic interactions within TFIIIC, the promoter-binding factor of yeast RNA polymerase III, *Mol Genet Genomics 265*, 705-710.

74. Jin, X. (2002) X-ray crystallographic studies of RNA Polymerase III transcription factor TFIIIB and 1L-*MYO*-Inositol 1-phosphate synthase, In *Department of Chemistry*, Michigan State University, East Lansing.

75. Huet, J., and Sentenac, A. (1992) The Tata-Binding Protein Participates in Tfiiib Assembly on Transfer-Rna Genes, *Nucleic Acids Res 20*, 6451-6454.

76. Kassavetis, G. A., Joazeiro, C. A. P., Pisano, M., Geiduschek, E. P., Colbert, T., Hahn, S., and Blanco, J. A. (1992) The Role of the Tata-Binding Protein in the Assembly and Function of the Multisubunit Yeast Rna Polymerase-Iii Transcription Factor, Tfiiib, *Cell 71*, 1055-1064.

77. Taggart, A. K. P., Fisher, T. S., and Pugh, B. F. (1992) The Tata-Binding Protein and Associated Factors Are Components of Pol-Iii Transcription Factor Tfiiib, *Cell 71*, 1015-1028.

78. Colbert, T., Lee, S., Schimmack, G., and Hahn, S. (1998) Architecture of protein and DNA contacts within the TFIIIB-DNA complex, *Mol Cell Biol 18*, 1682-1691.

79. Chaussivert, N., Conesa, C., Shaaban, S., and Sentenac, A. (1995) Complex Interactions between Yeast Tfiiib and Tfiiic, *J Biol Chem 270*, 15353-15358.

80. Dieci, G., Percudani, R., Giuliodori, S., Bottarelli, L., and Ottonello, S. (2000) TFIIIC-independent in vitro transcription of yeast tRNA genes, *J Mol Biol 299*, 601-613.

81. White, R. J., and Jackson, S. P. (1992) Mechanism of Tata-Binding Protein Recruitment to a Tata-Less Class-Iii Promoter, *Cell 71*, 1041-1053.

82. Lobo, S. M., Tanaka, M., Sullivan, M. L., and Hernandez, N. (1992) A Tbp Complex Essential for Transcription from Tata-Less but Not Tata-Containing Rna Polymerase-Iii Promoters Is Part of the Tfiiib Fraction, *Cell 71*, 1029-1040.

83. Cormack, B. P., and Struhl, K. (1992) The Tata-Binding Protein Is Required for Transcription by All 3 Nuclear-Rna Polymerases in Yeast-Cells, *Cell 69*, 685-696.

84. Schroder, O., Bryant, G. O., Geiduschek, E. P., Berk, A. J., and Kassavetis, G. A. (2003) A common site on TBP for transcription by RNA polymerases II and III, *Embo J 22*, 5115-5124.

85. Juo, Z. S., Chiu, T. K., Leiberman, P. M., Baikalov, I., Berk, A. J., and Dickerson, R. E. (1996) How proteins recognize the TATA box, *J Mol Biol 261*, 239-254.

86. Shen, Y. H., Kassavetis, G. A., Bryant, G. O., and Berk, A. J. (1998) Polymerase (Pol) III TATA box-binding protein (TBP)-associated factor Brf binds to a surface on TBP also required for activated Pol II transcription, *Mol Cell Biol 18*, 1692-1700.

87. Vilalta, A., Trivedi, A., and Johnson, D. L. (1996) Mutation in the tata-binding protein (TBP) selectively abolishes both RNA polymerase III transcription and the interaction of TBP with a TFIIIB subunit., *Mol Biol Cell 7*, 2704-2704.

88. Kim, Y. C., Geiger, J. H., Hahn, S., and Sigler, P. B. (1993) Crystal-Structure of a Yeast Tbp Tata-Box Complex, *Nature 365*, 512-520.

89. Khoo, B., Brophy, B., and Jackson, S. P. (1994) Conserved Functional Domains of the Rna-Polymerase-Iii General Transcription Factor Brf, *Gene Dev 8*, 2879-2890.

90. Kassavetis, G. A., Bardeleben, C., Kumar, A., Ramirez, E., and Geiduschek, E. P. (1997) Domains of the Brf component of RNA polymerase III transcription factor IIIB (TFIIIB): Functions in assembly of TFIIIB-DNA complexes and recruitment of RNA polymerase to the promoter, *Mol Cell Biol 17*, 5299-5306.

91. Kassavetis, G. A., Kumar, A., Ramirez, E., and Geiduschek, E. P. (1998) Functional and structural organization of Brf, the TFIIB-related component of the RNA polymerase III transcription initiation complex, *Mol Cell Biol 18*, 5587-5599.

92. Ferri, M. L., Peyroche, G., Siaut, M., Lefebvre, O., Carles, C., Conesa, C., and Sentenac, A. (2000) A novel subunit of yeast RNA polymerase III interacts with the TFIIB-related domain of TFIIIB70, *Mol Cell Biol 20*, 488-495.

93. Werner, M., Chaussivert, N., Willis, I. M., and Sentenac, A. (1993) Interaction between a Complex of Rna Polymerase-Iii Subunits and the 70-Kda Component of Transcription Factor-Iiib, *J Biol Chem 268*, 20721-20724.

94. Ishiguro, A., Kassavetis, G. A., and Geiduschek, E. P. (2002) Essential roles of Bdp1, a subunit of RNA polymerase III initiation factor TFIIIB, in transcription and tRNA processing, *Mol Cell Biol 22*, 3264-3275.

95. Kumar, A., Kassavetis, G. A., Geiduschek, E. P., Hambalko, M., and Brent, C. J. (1997) Functional dissection of the B" component of RNA polymerase III transcription factor IIIB: A scaffolding protein with multiple roles in assembly and initiation of transcription, *Mol Cell Biol 17*, 1868-1880.

96. Shah, S. M. A., Kumar, A., Geiduschek, E. P., and Kassavetis, G. A. (1999) Alignment of the B " subunit of RNA polymerase III transcription factor IIIB in its promoter complex, *J Biol Chem 274*, 28736-28744.

97. Andrau, J. C., and Werner, M. (2001) B "-associated factor(s) involved in RNA polymerase III preinitiation complex formation and start-site selection, *Eur J*

Biochem 268, 5167-5175.

98. Grove, A., Kassavetis, G. A., Johnson, T. E., and Geiduschek, E. P. (1999) The RNA polymerase III-recruiting factor TFIIIB induces a DNA bend between the TATA box and the transcriptional start site, *J Mol Biol 285*, 1429-1440.

99. Braun, B. R., Kassavetis, G. A., and Geiduschek, E. P. (1992) Bending of the Saccharomyces-Cerevisiae 5s Ribosomal-Rna Gene in Transcription Factor Complexes, *J Biol Chem 267*, 22562-22569.

100. Soragni, E., and Kassavetis, G. A. (2008) Absolute gene occupancies by RNA polymerase III, TFIIIB, and TFIIIC in Saccharomyces cerevisiae, *J Biol Chem 283*, 26568-26576.

101. Persinger, J., Sengupta, S. M., and Bartholomew, B. (1999) Spatial organization of the core region of yeast TFIIIB-DNA complexes, *Mol Cell Biol 19*, 5218-5234.

102. Juo, Z. S., Kassavetis, G. A., Wang, J. M., Geiduschek, E. P., and Sigler, P. B. (2003) Crystal structure of a transcription factor IIIB core interface ternary complex, *Nature 422*, 534-539.

103. White, R. J. (2004) RNA polymerase III transcription and cancer, *Oncogene 23*, 3208-3216.

104. White, R. J. (2004) RNA polymerase III transcription - a battleground for tumour suppressors and oncogenes, *Eur J Cancer 40*, 21-27.

CHAPTER 2: X-RAY CRYSTALLOGRAPHIC STUDY OF *E. coli* BE/OLIGOSACCHARIDE COMPLEXES

2.1 *E.coli* Branching enzyme(BE) and polysaccharide

BE cleaves the α-1,4 linkage of natural linear polysaccharide and transfers the fragment to form α-1,6 branches. In order to cleave the glycosidic bonds, BE has to bind with substrate. The natural substrates for BE are long polysaccharide chains (either glycogen or starch) usually with more than 30 glucose units.

The active site of BE has been identified, it is located inside the $(\alpha / \beta)_8$ barrel region. (**Figure 2.1**) Several catalytic residues have also been identified by sequence alignment and site directed mutagenesis: Tyr300, Asp335, His340, Arg403, Glu458, His525 and Asp526 are involved in the branching function (*1-3*). This is also consistent with the active site location of α-amylase family enzymes. How these residues interact with substrate, and what is the detailed branching mechanism is still unknown. Although the investigation of some other α-amylase family enzymes showed binding between enzymes and substrate or pseudo-substrate (usually polysaccharide or mimics) (*4-6*), there is still no report about BE binding with substrate.

Given the fact that long polysaccharide chains are flexible, it is very likely there are

other binding sites outside the $(\alpha / \beta)_8$ barrel region. As mentioned before, so far there is no data about the binding between branching enzyme and polysaccharide (or natural substrate) either inside or outside the active site.

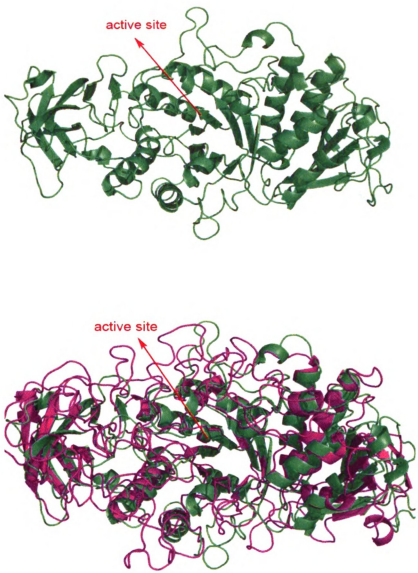


Figure 2. 1 ***E.coli*** **BE (PDB databank # 1M7X) overall structure** (*3, 7*). First figure shows the active site is inside the central α/β barrel region and the second figure shows *E.coli* BE overlaid with isoamylase from *Pseudomonas anyloderamosa* (PDB databank # 1BF2) (in magenta color).

In order to understand how BE catalyzes the formation of α-1, 6 branches, first we want to know how BE interacts with its substrate.

It is known that the *E.coli* BE lacking its first 113 amino acids (N113BE) still retains the majority of its branching enzyme activity, which means it still has the ability to form the enzyme-substrate interaction. And also the crystal structure of this construct is known, so the N113BE was selected to investigate the interaction.

Several kinds of oligosaccharides were selected for the study: maltose, maltotriose, maltotetraose, maltopentaose, maltohexaose, maltoheptaose, α-cyclodextrin, β-cyclodextrin and γ-cyclodextrin (Figure **2.2**). We used homogeneous oligosaccharides instead of natural starch/glycogen to simplify the problem.

2.2 Why were these oligosaccharides selected?

Cyclodextrins are cyclic oligosaccharides naturally produced from starch by enzymatic conversion (*8-9*). There are three major natural cyclodextrins: α-cyclodextrin (α-CD) (6 glucose units), β-cyclodextrin (β-CD) (7 glucose units) and γ-cyclodextrin (γ-CD) (8 glucose units). Figure 2.2 A shows their structures.

The natural substrates of BE, glycogen and starch, actually are not straight linear polymers. Naturally synthesized glucan adopts a helical structure instead of a random flexible structure in the glycogen /starch granule (*10-12*). The structure of amylose, the linear component of natural starch, was determined to be helical (*10, 13-16*). Amylose has 6 glucose units in one turn, which is very similar to α-CD. We overlaid the structures of amylose and α-CD in Figure 2.2B. It is clear that amylose has similar size and curvature as α-CD; two of the six glucose units of the one turn of the amylose overlay with α-CD perfectly. This indicates that cyclodextrins have similar structures to natural helical amylose. So cyclodextrins may serve as substrate mimics in substrate binding studies.

A)

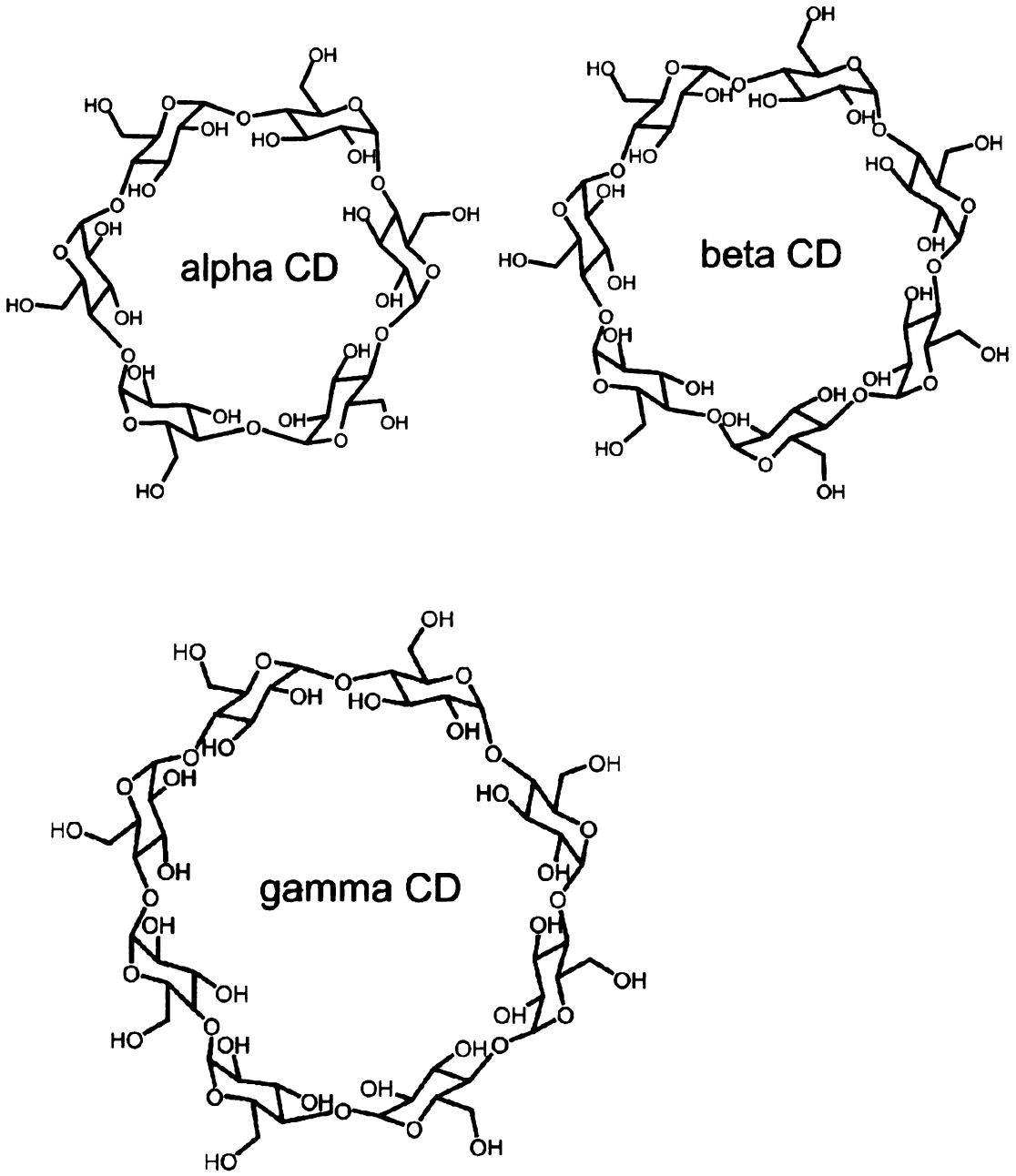


Figure 2. 2 **A)** Structures of α-cyclodextrin, β-cyclodextrin and γ-cyclodextrin. These three cyclodextrins are cyclic oligosaccharides with 6, 7 and 8 glucose units respectively.

B)

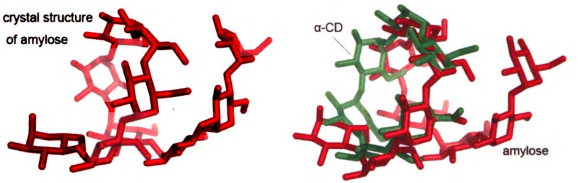


Figure 2.2 B) The left figure shows the crystal structure of amylose (*16*); the right figure shows the overlay of α -CD and amylose

There is no research regarding the binding between *E.coli* BE and oligosaccharides, but α-glucan binding to potato tuber starch branching enzyme I (SBEI) has been studied and shows that some cyclodextrins bind with SBEI stronger than linear oligosaccharides (*18*). SBEI and *E.coli* BE share similarities in sequence, function and overall structure. This gave us a hint that maybe cyclodextrin could bind with *E.coli* BE as well. Our research confirmed the binding between *E.coli* BE and several of these glucans.

Compound	Dissociation constant (mM)
Maltose	>50
Maltotriose	11.7
Maltotetraose	1.1
Maltopentaose	0.75
Maltohexaose	0.25
Maltoheptaose	0.16
α-cyclodextrin	6.0
β-cyclodextrin	0.25
γ-cyclodextrin	0.00067

Table. 1 Dissociation constants for different α-glucans and Potato tuber SBEI (*18*)

2.3 Materials and methods

2.3.1 Experimental design

X-ray crystallography was used to study the interaction between the truncated *E.coli* BE protein and selected oligosaccharides. After purification of the truncated *E.coli* BE protein, co-crystallization and substrate soaking were both attempted. Many crystals were used to collect diffraction data at the Advanced Photon Source (APS). Diffraction data was examined to find out if the substrate was present. Final structures of substrate-binding *E.coli* BE protein were refined with the CCP4 and COOT software packages. The experimental details will be elucidated.

2.3.2 Protein over-expression and purification

The pET 23d vector based plasmid containing the coding sequence of *E.coli* BE lacking the first 113 amino acids(*19*) at the N-terminus was obtained from our collaborator Dr. Jack Preiss. It was transformed into *E.coli* BL21 (DE3) cells and grown on ampicillin (0.05 mg/mL) resistant agar plate. Colonies appeared after overnight incubation at 37 °C. Cells were cultivated in 50 ml Luria Broth (L.B) overnight with ampicillin (0.1mg/mL) and then were transferred into 1 liter fresh L.B media with

ampicillin (0.1 mg/mL). Cells were grown at 37 °C until the optical density at 600nm ($O.D_{600}$) reached 0.5-0.6. Then IPTG (0.5mM) induced production of the protein. After cultivation at 25 °C for five more hours, the cells were harvested.

The cell pellet was re-suspended in buffer A (50mM Tris-Acetate, pH8.0, 10mM EDTA, 2.5mM DTT and 10% v/v glycerol) and lysed by sonication. Cell debris was spun down by centrifugation at 5000 rpm for 30 minutes. The supernatant was precipitated by adding 40% ammonium sulfate solution. After centrifugation the protein pellet was re-suspended in buffer A. The solution was dialyzed overnight against buffer A, then loaded into DEAE fractogel column. Protein fractions were eluted with 0-0.4 M KCl gradient in buffer A. The protein fractions were centrifuged to about 5 mg / ml. Source Q anion exchange chromatography was used to purify the protein to almost homogeneity. Pure protein was eluted with 0-1.0M KCl gradient in buffer A. Size exclusion chromatography (Superdex200 16/60 column from GE healthcare) was used as the final purification step. **Figure 2.3** shows the size exclusion chromatograph. Comparing the protein peak position and the protein standard chromatograph (shown in the graph), it is clear that the protein exists in solution as monomer. SDS-PAGE was used to verify the homogeneity and purity of the protein. Pure *E.coli* BE protein fractions were pooled together and concentrated to about 10 mg/ml for further use. **Figure 2.4** shows the SDS-PAGE gel of final pure truncated *E.coli* BE protein samples.

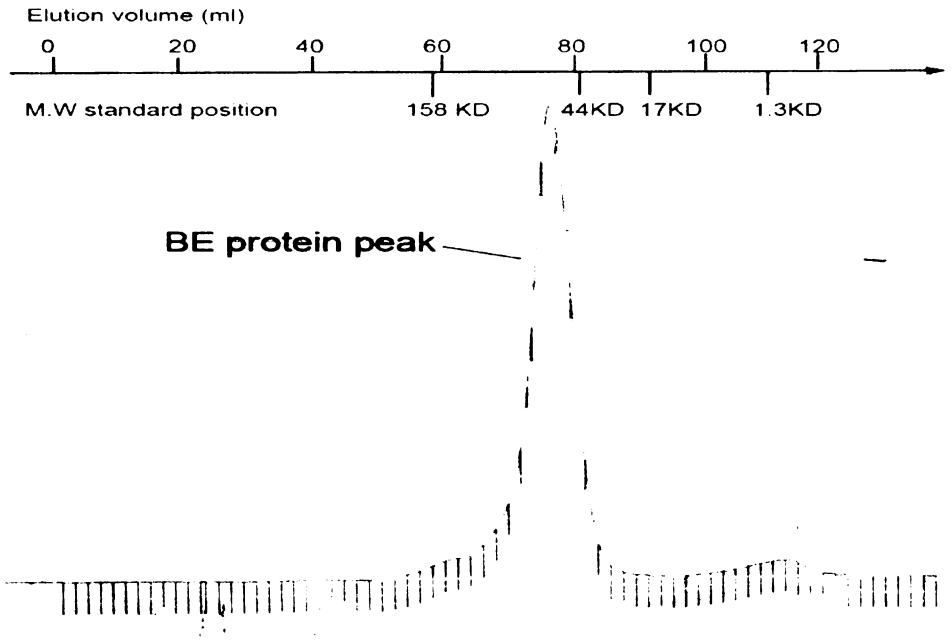


Figure 2. 3 Chromatograph of truncated *E.coli* BE protein after size exclusion chromatography purification (Absorption at 280 nm). Protein was eluted as a single sharp peak, which indicated the homogeneity of protein particle size.

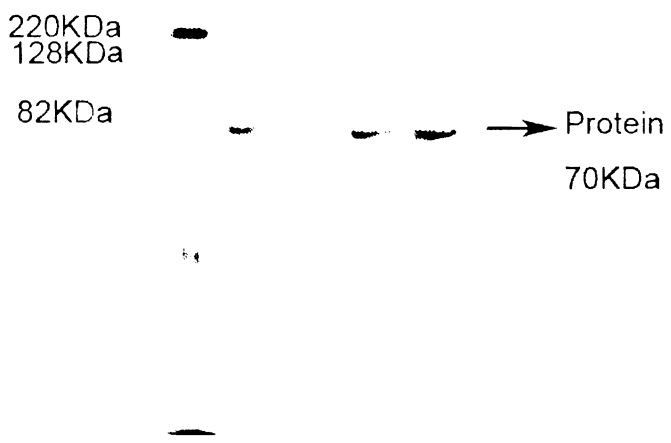


Figure 2. 4 SDS-PAGE gel of several pure truncated *E.coli* BE protein samples. The samples were taken from different batches of final protein solution.

2.3.3 Protein crystallization

Pure *E.coli* BE protein was buffer exchanged into Crystallization buffer (25 mM Na-HEPES, pH7.2). Then protein solution (about 10 mg /ml) was subject to sparse matrix crystallization screen (under condition: 100mM Na-HEPES, pH 7.2 at 4 °C) using the hanging drop vapor diffusion method(*3*). Protein crystals appeared after 2 weeks, and reached maximal size (0.3x0.15x0.1 mm) in 6 weeks. (See **Figure 2.5)**

Several batches of protein were used to grow crystals. Single crystals were reproduced for further substrate soaking experiments.

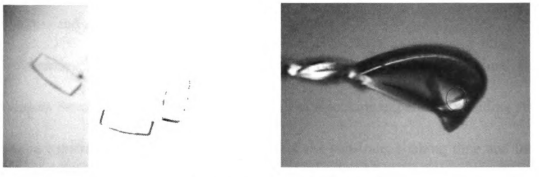


Figure 2. 5 Crystals of *E.coli* branching enzyme. The first picture shows crystals inside the crystallization screen drop (crystal size 0.05x0.1x0.2 mm), second picture shows a crystal in cryo-solution, ready for x-ray diffraction

2.3.4 Attempts at co-crystallization of BE with oligosaccharides

Co-crystallization of *E.coli* BE with various oligosaccharides was attempted. Oligosaccharides solutions (from 1 mM to 10 mM) were mixed with protein solution together and incubated at 4 °C for at least 1 hour for binding. Sparse Matrix crystallization screens were performed at both room temperature and 4 °C. No crystals were found.

2.3.5 Soaking BE crystals with oligosaccharide solutions

The complex structures were obtained by substrate soaking. The truncated *E.coli* BE protein was crystallized first, then the protein crystals were picked up using a cryo loop (Hampton Research, Ltd), and dropped into oligosaccharide solutions at 4 °C for various time. Crystal cracking was carefully monitored. Then crystals were flash frozen with cryo protection and stored in liquid nitrogen.

Extensive soaking experiments were carried out to search for the optimal soaking condition. Several factors affect the final result of a soaking experiment: binding strength between substrate and protein, concentration of the substrate, soaking time and inherent nature of the protein crystal.

In this research, we managed to overcome the low quality diffraction and soaking induced crystal cracking problem. Extensive soaking conditions were screened and soaking induced cracking was carefully monitored.

The substrates were prepared in solutions at the highest concentrations possible.

Soaking was carried out for maximal time period before significant cracking was found on the surface of crystals.

For linear oligosaccharides, the concentration of substrate ranged from 50 mM to 200 mM. A 70 mM solution was used for α and γ-cyclodextrin. β-cyclodextrin was prepared at a concentration of 15 mM due to its low solubility.

Substrate	M.W. (Dalton)	Conc. (mM)	Soaking time before crystal cracking (hrs)	Electron density of the oligosaccarhde
Maltose	342	200	72	No
Maltotriose	504.4	100	72	No
Maltotetraose	666.6	140	12	No
Maltopentaose	828.7	100	6	Yes
Maltohexaose	990.9	50	3	Yes
Maltoheptaose	1153	90	3.5	Yes
α-CD	972.8	70	1	Yes
β-CD	1135	15	12	Yes
γ -CD	1297	70	15	Yes

Table 2. 2 Summary of substrates used in soaking experiment and soaking conditions. The concentrations are the highest concentration of oligosaccharides used in the experiments. The soaking times are the longest possible time before the crystals crack. All oligosaccharides were from Sigma.

* The x-ray diffraction data were examined to find if there is electron density of the substrate. But this method can not decide if the substrate binds with *E.coli* BE protein or not. There is possibility that the substrate binds with protein, but is flexible to some extent that its density can not show up in diffraction data.

2.3.6 X-ray diffraction data collection and processing

The x-ray diffraction data were collected at beam line #21 (LS-CAT) at Advanced Photo Source, Argonne National Lab.

Diffraction data were processed using the HKL2000 software package. The MOLREP program in CCP4 program suite(*20*)) was used to position the *E.coli* BE / oligosaccharide complex in the unit cell using the *E.coli* BE structure as the search model. CCP4 and the Coot program suite were used to refine the structures.

Table 1 and 2 in Appendix list the detailed data collection and refinement statistics for all the structures obtained. All the ligand-bound structures retain the same cell dimension as the apo structure (1M7X). All the diffraction data have acceptable completeness. Ramachandran plots showed that the numbers of disallowed residues were within the reasonable limit for the resolution of the structure (see Ramachandran plots in the Appendix).

2.4 Three dimensional structure of *E.coli* BE/cyclodextrin complexes

The *E.coli* BE structure was previously reported by our group(*3*). The protein was crystallized in 100mM Na-HEPES (pH 7.2) solution. It belongs to the P21 space group with cell dimension of 91, 103 and 185 Å. There are 4 monomers in one asymmetric unit (numbered as chain A, B, C and D respectively). The overall structure is shown in Figure 2.1. We obtained x-ray diffraction data sets of *E.coli* BE in complex with α-cyclodextrin, β-cyclodextrin and γ-cyclodextrin.

We found 4 cyclodextrin binding sites in the *E.coli* BE /cyclodextrin complex. The binding between α, β, and γ-cyclodextrins and BE are present in many data sets(we collected multiple data sets for each complex) . **Table 2.4** lists the binding sites and the residues involved in the binding.

Binding site	**Bound CDs**	**BE residues involved in the binding**
I (on ABCD chains)	α and β-CDs	ARG255, ASN259, ASN260, PHE261 and TRP262
II(on ABCD chains)	α, β and γ-CDs	ASP505,PHE508, ILE511, LEU512 and TRP628
III(only on D chain)	α and γ- CDs	TRP159, LYS189, LEU201,GLN211 and GLU215
IV(only on C chain)	α, β and γ-CDs	ASP542, TRP544, GLN545, PRO659 and SER689

Table 2. 3 Detailed binding sites in *E.coli* BE /cyclodextrin complex structures

2.4.1 Three dimensional structure of *E.coli* BE/α- cyclodextrin complex.

α- cyclodextrin (α-CD) binds with *E.coli* BE at all 4 binding sites. α- cyclodextrin binds with binding site I and II on all of the 4 chains in the asymmetric unit. Binding site III is only occupied with α- cyclodextrin on Chain C. Also α- cyclodextrin only binds with binding site IV on chain D of the protein.

α- cyclodextrin is a six-glucose cyclic oligosaccharide. The molecule forms a truncated cone shape. Cyclodextrins are well-known for their ability to carry guest molecules inside its "cavity" (*8, 21*). Since the inside of the molecule is hydrophobic while the outside is hydrophilic, it tends to host small hydrophobic guest molecule inside. This characteristic makes cyclodextrins suitable for drug delivery. It can carry a small hydrophobic guest molecule and deliver it into an aqueous environment (*22-23*). Upon binding with a protein, it can incorporate a hydrophobic residue side chain inside its cone-shape cavity, which stabilizes the binding further. **Figure 2.6** shows the electron density map of α- cyclodextrin from one of our α- CD/BE structures.

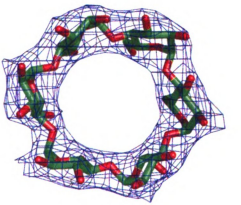


Figure 2. 6 1.0 σ $2F_0$-F_c electron density map of α-cyclodextrin binding with *E.coli* BE at binding site IV.

At binding site I, we found the side chain of PHE261, the benzyl group, extends into the cavity, forming a hydrophobic interaction with the cyclodextrin. This is a important stabilizing force in the binding.

TRP262 also plays an important role in stabilizing α- cyclodextrin binding: Its aromatic indole side chain is parallel to the +1 glucose moiety. The aromatic stacking pattern is one of the major types of sugar-protein interaction. Furthermore, the aromatic side chain of PHE261 is also parallel to the +1 glucose unit. The two aromatic side chains from the two consecutive residues clamp the glucose unit between them. This interesting interaction pattern is not very common in sugar-protein binding.

The hydrogen bonds between a hydroxyl group of cyclodextrin and ARG255, ASN259 and ASN260 also contribute to the binding. **Figure2.7** shows the detailed interaction between α- cyclodextrin and BE residues at the binding site.

By overlaying the local residues with the un-bound *E.coli* BE structure (1M7X). We found there was basically no conformational change upon binding with cyclodextrin except the side chain of ASN259. It moved up towards the cyclodextrin a little to form a hydrogen bond (3.48 Å). Considering that this hydrogen bond is very weak and the ASN259 side chain was flexible in the 1M7X structure, we can conclude that there is little energy needed to distort the conformation of any residue upon binding with α-cyclodextrin. This site can only bind with the end of amylose, otherwise the rest of the amylose chain will clash with the protein.

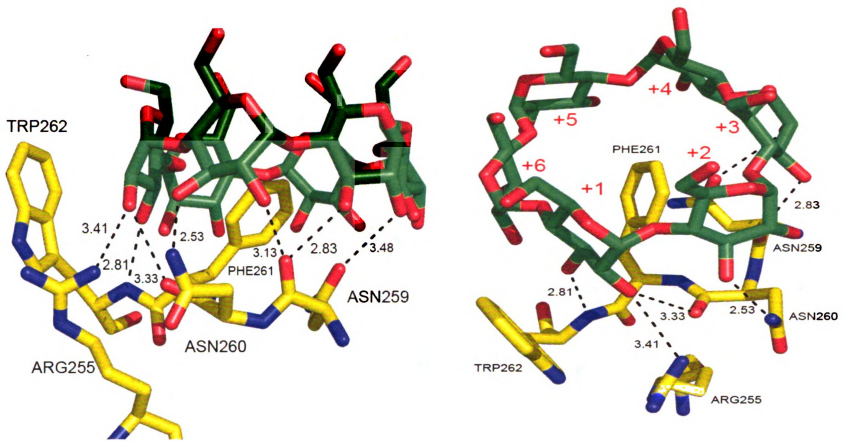


Figure 2. 7 Left shows α-cyclodextrin binding with binding site I (ARG255, ASN259, ASN260, PHE261 and TRP 262). Hydrogen bonds are shown in the figure. Right figure shows the interaction from a different view angle, the glucose units are numbered from +1 to +6.

Another binding site is also on the surface of the protein, quite far away from the active site (about 20 Å away from the edge of the beta barrel). Binding site II includes ASP505, THR508, LEU512, and TRP628. The cyclodextrin is bound to the surface of protein through mainly hydrogen bonds. (see **Figure 2.8**)

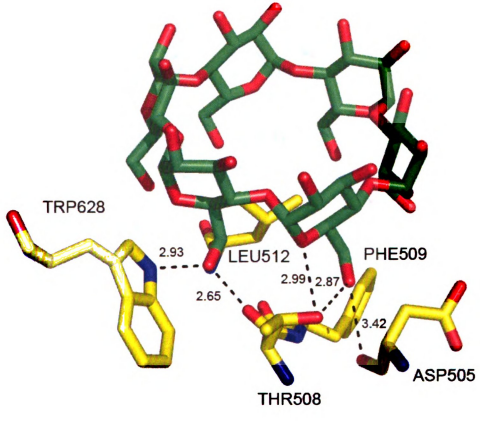


Figure 2. 8 α-cyclodextrin binds with binding site II, residue 505, 508, 512 and 628 have interactions with cyclodextrin. Detailed hydrogen bonds are indicated.

The side chain of LEU512 sticks to the cavity of α-cyclodextrin, but can not go further into the hydrophobic pocket, that is because the steric hindrance of PHE509: the side chain of PHE509 blocks the way of a glucose unit, this can be seen in **Figure 2.9**(It shows the interaction from a different angle), otherwise we can imagine that the hydrophobic side chain of leucine will extend deeply into the cyclodextrin pocket like binding site I.

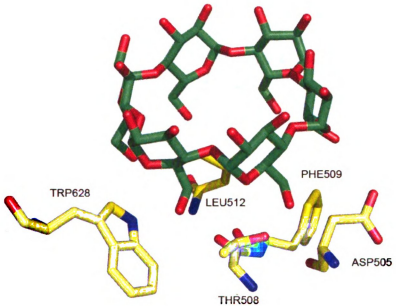


Figure 2. 9 Binding between α-cyclodextrin and *E.coli* BE binding site II, from another angle. It shows that side chain of PHE509 blocks the glucose units from moving closer to incorporate LEU512 into the cyclodextrin pocket

Binding sites I and II are found on all 4 chains in the asymmetric unit. The binding is quite strong considering some data sets were obtained by soaking the crystal in CD solution for only 30 minutes. Also, by overlaying with apo structure 1M7X, we did not find significant conformational changes around binding sites I and II,

We also found two binding sites (III and IV) which bind with CD on one of the four

chains in the asymmetric units. Binding site III is only on the D chain and binding site IV is found only on the C chain. Binding site III consists of TRP159, GLN 211 and GLU 215 on the D chain.

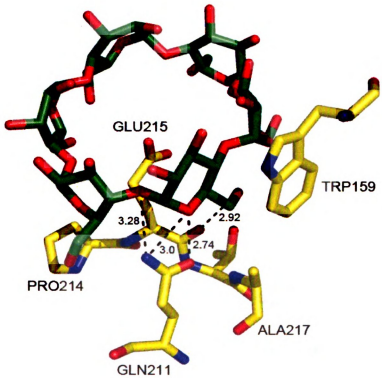


Figure 2. 10 α-cyclodextrin binds with binding site III. GLN211 and GLU215 on D Chain of the protein have hydrogen bonds with CD.

GLN211 and GLU 215 have hydrogen bonds with CD. Also there is an aromatic stacking interaction between the indole side chain of tryptophan 159 and CD. As we mentioned before, this type of interaction between the aromatic side chain of tryptophan and the glucose plane is very typical of protein-sugar binding (*24-29*).

The apo structure of *E.coli* BE has some "disordered" loops (the electron density of the residues are absent because they are flexible inside the crystal). Residue 213-215 is one of the disordered loops in the *E.coli* BE apo structure. While upon binding with ligand, the interaction between the ligand and certain residues of the missing loop helped to stabilize the loop. As a result, the former disordered loop appeared in the ligand-bound structure. In our α-cyclodextrin-bound BE structure, GLU215 not only contributes by means of hydrogen bonds but also extends its side chain into the cyclodextrin pocket to form an extra interaction. (see **Figure 2.11**).

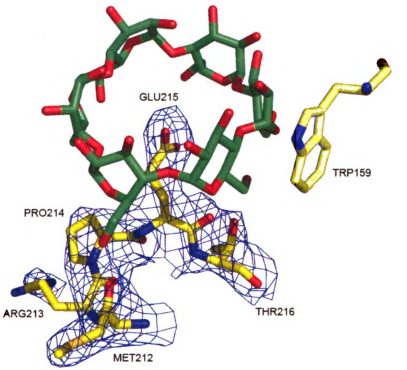


Figure 2. 11 $2F_0$-F_c electron density map of loop 213-215 on D chain of *E.coli* BE /α-cyclodextrin

The last binding site consists of ASP542, TRP544, GLN545, PRO659 and SER689.

Figure 2.12 shows the detailed interaction between the residues and α-CD.

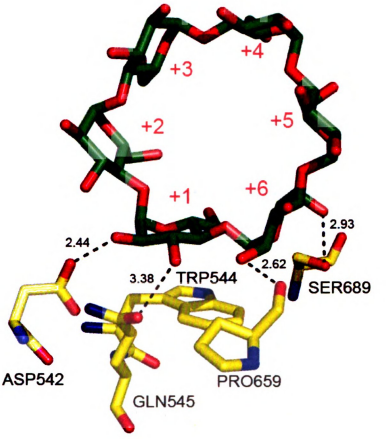


Figure 2. 12 *E.coli* BE binds with α-CD at binding site IV. α-CD is stabilized by hydrogen bonds from ASP542, GLN545, PRO659 and SER689. TRP544 contributes to the binding through aromatic stacking with the glucose unit.

Binding sites III and IV only exist in one molecule of the 4 in the asymmetric unit. Binding site III is on the D chain. Its corresponding position on the A chain is blocked by both the cyclodextrin bound to binding site II on the D chain and another symmetry mate due to crystal packing. Although the same site on the B and C chains are not directly blocked, the pathway to the binding site is not as open as it is in the D chain, so its

binding with substrate may be more difficult, which maybe why we did not see the binding in those molecules.

Binding site IV is only found on the C chain. That is because the same positions on the A, B and D chains are completely blocked by symmetric molecules due to crystal packing, making the binding impossible.

2.4.2 Three dimensional structure of *E.coli* BE / β- cyclodextrin complex

β- cyclodextrin (β-CD) has 7 glucose units. It is bigger than α-cyclodextrin, internal diameter increases from 57 Å to 78 Å (*9*). We expected β- CD to bind with *E.coli* BE more strongly, since the size of inner cavity seems to be an important factor for binding site I and II. Due to the unusual low solubility of β- CD (*9, 23*), we had to increase the soaking time to get proper binding. the experiment was carried out at low substrate concentration such as 1mM and 5 mM for longer times (over 10 hrs) , but did not yield any cyclodextrin bound protein data.

We also tried the soaking experiment with saturated substrate for shorter times (30 minutes, 1 hrs, 2 hrs, and 4 hrs), still no substrate binding was found. The β- CD bound protein data sets were obtained by soaking experiments with a saturated β- cyclodextrin solution for a longer time (at least 10 hrs). From the X-ray diffraction data, we found that β- CD still binds at binding site II and IV, but the binding with binding sites I and III are weaker than those seen in the *E.coli* BE/ α-CD structure.

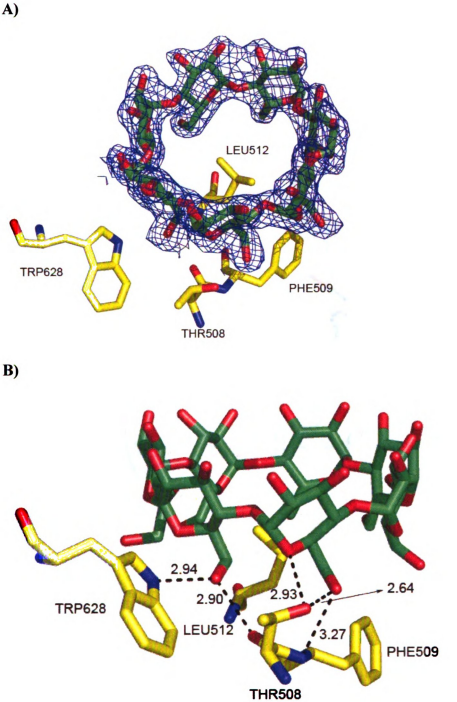


Figure 2. 13 β- CD binds with *E.coli* BE at binding site II. TRP508, PHE509, LEU512 and TRP628 are involved in the interaction by both hydrogen bonds and hydrophobic interaction. A: the 1.0 σ 2Fo-F_c density map of the β- CD; B: The detailed interaction between the residues and the CD.

Although β-CD is bigger than α-CD, there is basically no difference in terms of the binding pattern at binding site IV except that ASP542 could donate two hydrogen bonds instead of one , and tiny changes in hydrogen bond lengths (**Figure 2.14**).

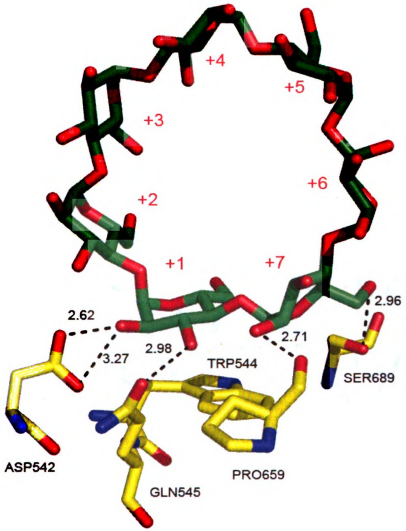


Figure 2. 14 The interaction between *E.coli* BE binding site IV and β-CD. Like α-CD binding, ASP542, GLN545, PRO659 and SER689 contribute by hydrogen bonds and TRP544 through an aromatic stacking force.

From Figure 2.14, we can see that β-CD is bigger; the part bound to protein is flatter, so that part has more contact with the protein. In our structure, it indicates that the protein stabilizes the β-CD better through one more hydrogen bond coming from ASP542. By comparing the β-CD bound BE structure with the α-CD bound structure, we noticed that ASP542 adopted a tiny side chain conformational change to accommodate the hydrogen bonds.

We reasonably doubted that binding site I binds with β-CD much weaker than it does with α-CD since in our data sets we never have 4 CD bind to one structure. And usually the electron density of CD is also much weaker, which indicates the weaker interaction. (See **Figure 2.15 and Figure 2.16** for binding at binding site I)

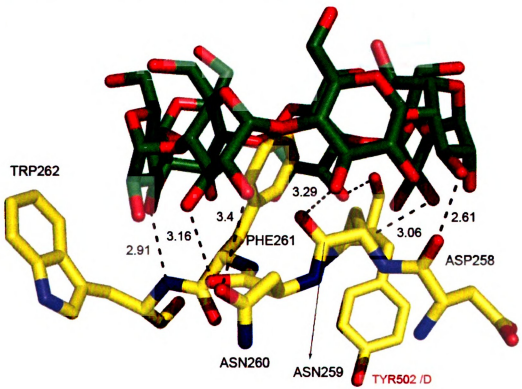


Figure 2. 15 The detailed interaction in binding site I of the A molecule with β-CD. ASP258, ASN259, ASN260, TRP261, PHE262 on A molecule and TYR502 on D molecule are involved in binding.

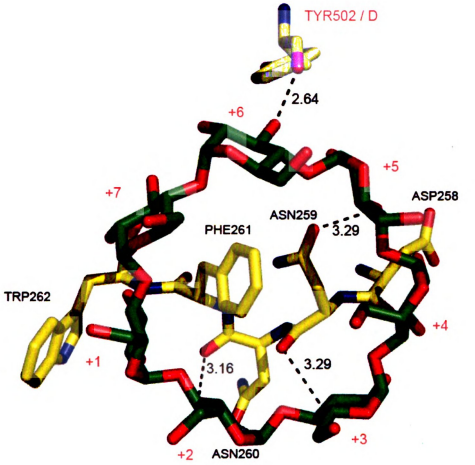


Figure 2. 16 The binding between β-CD and *E.coli* BE binding site I from top.

From Figures 2.15 and 2.16 it is clear that β-CD has more interaction with the protein than α-CD does. In addition to the α-CD interactions, ASP258, ASN259 and ASN260 are also involved in the hydrogen bonds. Moreover TYR502 from the D molecule also binds with β-CD through a hydrogen bond. It might contribute to the stabilization of the complex in the crystal.

There is no electron density for β-CD at binding site III. After examination of all of our β-CD bound *E.coli* BE structures, we only found very weak and incomplete electron density at this binding site in 2 data sets (out of 21 data sets collected).

2.4.3 Three dimensional structure of *E.coli* BE/ γ- cyclodextrin complex

γ-cyclodextrin (γ- CD) is the largest natural cyclodextrin; it also has the highest solubility of the three natural cyclodextrins. So it is easy to prepare a high concentration substrate solution. Also from the SBEI / cyclodextrin binding assay, γ- CD has the lowest K_m value (0.00067 mM)(*18*). All of this indicated that perhaps γ- CD was a good choice for soaking experiments. In our soaking experiments, 1 mM, 10 mM and 70 mM γ- CD solutions were used. Soaking time ranged from 30 minutes to 24 hours (beyond 24 hrs, the crystals cracked completely). The substrate bound data set was obtained by soaking *E.coli* BE apo crystals in 70 mM γ- CD for 15 hrs.

γ- CD only binds with *E.coli* BE at binding site II, III and IV. At binding site II, the binding residues include THR508, PHE509, LEU512 and TRP628. These residues have direct contact with γ- CD through hydrogen bonds.

Compared with α-CD and β-CD, γ- CD is larger, the glucose ring is much looser/flexible, so the electron density of the non-binding glucose units are not as rigid as in α-CD and β-CD.

Figure 2.17, **Figure 2.18** and **Figure 2.19** show the detailed interactions at binding site II, III and IV respectively.

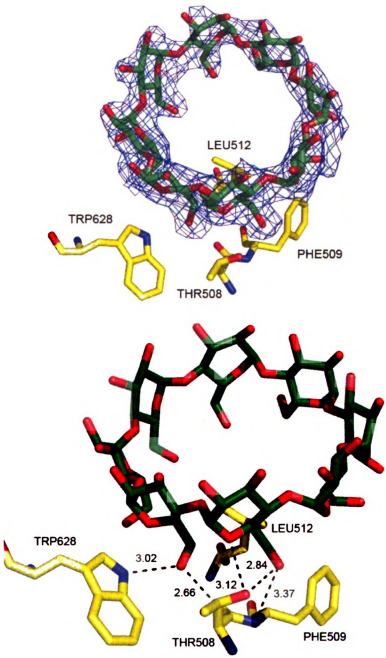


Figure 2. 17 γ- CD binds with *E.coli* BE at binding site II. THR508, PHE509, LEU512 and TRP628 are interacting with CD by hydrogen bonds. The top figure shows the 1.0 σ $2Fo\text{-}F_c$ electron density map of the CD and the bottom figure shows the detailed interaction.

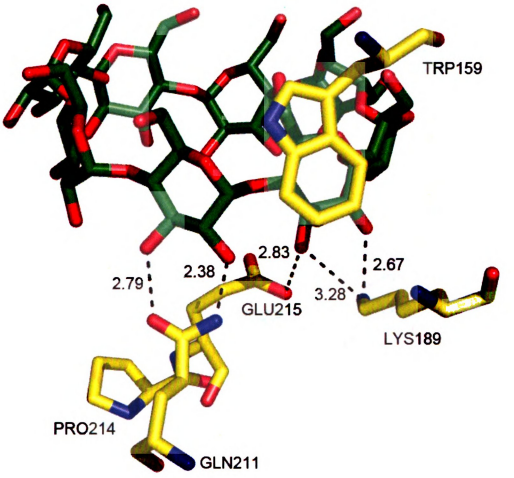


Figure 2. 18 γ- CD binds with *E.coli* BE at binding site III. LYS189, GLN211, PRO214, GLU215 and TRP159 interact with γ- CD through hydrogen bonds and aromatic stacking.

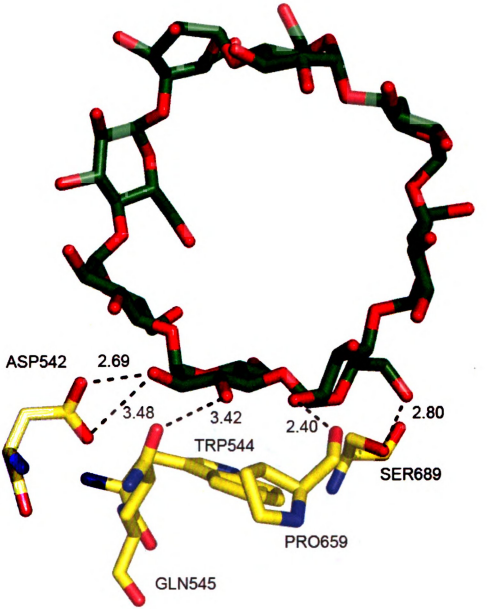


Figure 2. 19 γ- CD binds with *E.coli* BE at binding site IV. ASP542, GLN545, PRO659, SER689 and TRP544 bind with γ- CD by hydrogen bonds and aromatic stacking.

2.4.5 Overlay of three cyclodextrin bound BE structures

α –CD, β –CD and γ –CD have similar structures and chemical properties. The only difference is the size of the rings and the curvatures. From the previous discussion, we already know that they have similar binding patterns with the protein. The different sizes and geometry of the CD molecules result in slightly different interactions. For example, γ –CD interacts with more residues than the other 2 CDs at binding site II. These differences may or may not affect the binding strength.

The overlaying of three CD bound *E.coli* BE structures did give us some hint about the substrate preference of *E.coli* BE. We first overlaid the three structures at binding site II. (**Figure2.20** and **Figure 2.21**)

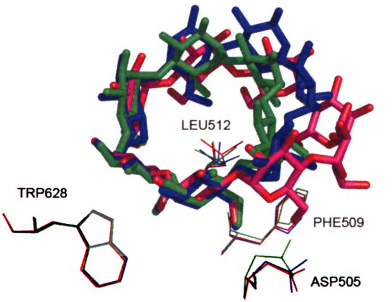


Figure 2. 20 Overlay of α–CD, β-CD, γ-CD and their bound residues at binding site II. Green color represents α–CD and its bound residues; blue represents β-CD and magenta for γ-CD

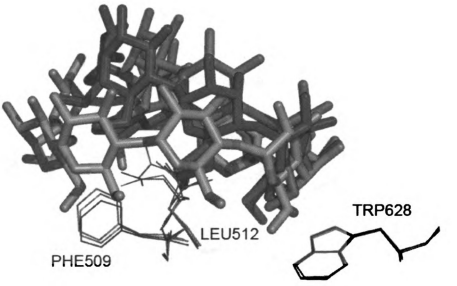


Figure 2. 21 Overlay of α–CD, β-CD, γ-CD and their bound residues at binding site II. Green color represents α–CD and its bound residues; blue represents β-CD and magenta stands for γ-CD

From the figures above, it is clear that the residues bound to the CDs remain in almost identical conformations. The two glucose units on TRP628 overlay very well among the three structures. This side is the major hydrogen bond source. As to the relative position between LEU512 and cyclodextrins, LEU512 itself did not adopt any conformational change. It remains at the same position in all three structures. But when the CD rings become bigger (from α–CD to β-CD to γ-CD), the CDs no longer sit on top of LEU512, they tend to include LEU512 further inside the hydrophobic pocket. This can be seen from Figure 2.21 more clearly.

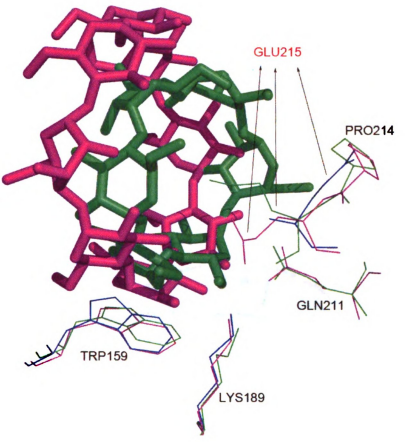


Figure 2. 22 Overlay of α–CD, γ-CD and their bound residues at binding site III. Green color represents α–CD and its bound residues; and magenta stands for γ-CD. There is no β-CD bound to this binding site. So we use blue color to represent unbound structure.

From **Figure 2.22** we can see at binding site III, α–CD and γ-CD binds with the protein similarly. The side chains of GLU215 participate in the interaction to stabilize both the CD and the disordered 213-215 loop although their side chains adopt different conformations. We were barely able to assign the GLU215 residue in the disordered loop in the unbound structure. But the position of its side chain is different from the α–CD and γ-CD bound structures since there is no interaction to lock its position. As a

result, the B factor of GLU215 indicates it is very flexible. Also from the overlaid structures, we can see that PRO214 blocks γ-CD from moving towards GLN211. **Figure 2.23** shows the three different CD-bound structures overlaid at binding site IV. We can see almost all the bound residues remain in the same position. The two glucose units closest to TRP544 also overlap completely.

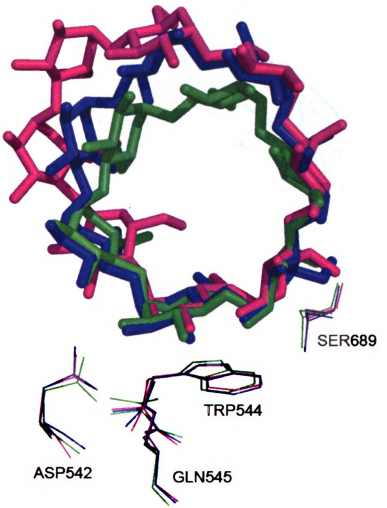


Figure 2. 23 Overlay of α–CD, β-CD, γ-CD and their bound residues at binding site IV. Green color represents α–CD and its bound residues; blue represents β-CD and magenta stands for γ-CD and its residues.

The last binding site we overlaid is binding site I. This site has strong binding with α –CD, relatively weak binding with β –CD and no binding with γ –CD. This trend seems related to the size of CDs.

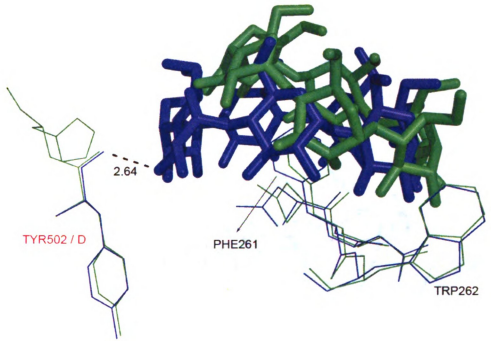


Figure 2. 24 Overlay of α–CD, β-CD and their bound residues at binding site I. Green color represents α–CD and its bound residues; blue represents β-CD

Binding site I binds with α–CD and β-CD. Among the 4 molecules in the asymmetric unit, the binding site on the B molecule cannot bind with CD because the space is blocked by a symmetric molecule. The same binding site on the A, C and D molecules is still open.

After examination of all of our data, we found that both α–CD and β-CD bind with binding site I on molecule A, but γ-CD does not bind. In order to explain this discrepancy, we overlaid the α–CD and β –CD bound structures and discovered the reason why γ –CD is not found on molecule A. From **Figure 2.24** we can see, binding site I holds α –CD like a tweezers. The side chains of TRP 262 and PHE261 are two tips of the tweezers. The benzyl group side chain of PHE261 probes into the hydrophobic inner cavity of CD. Part of the CD lies in the middle of TRP262 and PHE261, with its position quite fixed. So the bigger CD has to extend away from the "tweezers". And in that direction, we can see the TYR502 from the D molecule.

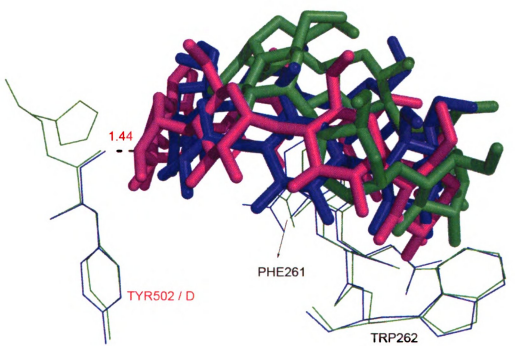


Figure 2. 25 Overlay of α–CD, β-CD and their bound residues at binding site I. Also a free γ –CD molecule is modeled on binding site. Green color represents α–CD and its bound residues; blue represents β-CD and its bound residues. Magenta represents a free γ –CD molecule.

α –CD has no contact with the D molecule, since it is the smallest CD. But β –CD already has interaction with TYR502 through a hydrogen bond (2.64 Å). **Figure 2.25** shows the overlay of α –CD and β –CD bound to binding site I on *E.coli* BE. We also modeled one γ –CD onto the binding site, with the right part of the CD (2 glucose units) positioning between side chain of TRP262 and PHE261. It is clearly seen from the overlaid structure that there is a clash between γ –CD and molecule D of *E. coli* BE (1.44 Å). This is probably the reason why γ –CD can not bind with this binding site by the soaking method. As we mentioned before, *E.coli* BE assumes a monomer state in solution (in *vivo / in vitro* situation), but forms a tetramer in its crystalline state. This indicates that γ –CD can not bind with this binding site of the BE in this crystal packing pattern. But it is possible γ –CD binds with *E.coli* BE binding site I in the solution since there will no be clash due to crystal packing.

In our β –CD bound structure; we can see very weak electron density at this binding site in the C and D molecules, but no density at all in the γ –CD bound structures. The reason is still unknown. Because of crystal packing, binding site I on molecule B is completely blocked by symmetry molecules, so neither β –CD nor γ –CD can bind to this site on molecule B.

2.4.6 Overall binding of cyclodextrins with *E.coli* BE

In order to show the overall binding site distribution, we overlaid the 4 molecules of the α-CD bound structure , and superimposed all the 4 bound α-CD onto one molecule (**Figure 2.26**) (in our crystal structures, they never bind to the same molecule: binding site III is only on molecule D and binding site IV only on molecule C).

From the overall composite figure, we can see that all the binding is on the surface of the protein, the binding sites are distributed around the active site, (α / β) barrel region. It is clear the function of these binding sites is to hold the substrate. Whether they are involved in concerted catalysis action is unknown. But we believe that the natural substrate of BE will undergo some conformational change before the hydrolysis catalysis step.

The active site region is filled with the side chains of BE residues. Also the diameter of the double helix amylose (10.6 Å) and α-CD (13.7 Å) (*9-10*) are quite big compared to the BE active site barrel (diameter of the barrel is about 10-11 Å), so we believe that the natural substrate has no access to the active site in its double helix state. There must be an unwinding step before the catalysis step. Also acarbose was used to model the polysaccharide inside the active site. This led us to imagine that the catalysis involves both double helix state substrate and linear oligosaccharide binding.

So the next part we are going to investigate is the binding between *E.coli* BE and linear oligosaccharides.

Figure 2. 26 A composite of the crystal structures of *E.coli* BE bound to α-cyclodextrins. All 4 molecules of α-CD bound structure were overlaid and all 4 α-CDs were superimposed onto molecule C to illustrate the overall binding. An acarbose molecule is also modeled into the active site. The shortest distance between the acarbose and cyclodextrin molecules is listed.

2.5 Three dimensional structures of *E.coli* BE/linear oligosaccharides complexes

Linear oligosaccharidess used in the research include: maltose, maltotriose, maltotetraose, maltopentaose, maltohexaose and maltoheptaose.

Different protein / oligosaccharide ratios were tried in the co-crystallization experiments. Only apo crystals were grown in the maltose solution. The co-crystallization experiments with other oligosaccharides failed and no crystals have formed.

In substrate soaking experiments, all oligosaccharides were dissolved into the initial crystallization solution (100 mM Na-HEPES, pH7.2). Apo *E.coli* BE protein crystals were utilized for soaking experiments. It was found that soaking in all oligosaccharides except maltose caused the protein crystals to crack (visual inspection under microscope). In order to make sure the observation was accurate and neutral, A control experiment was performed (soaking was carried out in pure initial crystallization buffer without oligosaccharides, and even in pure water, absolutely no cracking happened.) This is an important indication that the oligosaccharides do interact with *E.coli* BE crystals. Given the behavior of maltose in the co-crystallization and soaking experiment, we believe that maltose is not binding with the *E.coli* BE protein.

At least 20 crystals were soaked with maltotriose solution (from 1mM to 45 mM) for various times(from 30 minutes to 12 hrs). The diffraction data of the crystals were examined and no electron density of maltotriose was found. But considering its ability to "crack" crystals in the soaking, which indicated some kind of interaction between maltotriose and the protein crystal, it is possible that maltotriose does interact with *E.coli*

BE, but due to binding compromise at other binding sites, the binding of maltotriose is not sufficient before the crystal lattice was destroyed: so the electron density of maltotriose is not seen.

We did find some low quality electron density maps of one or two glucose units of maltotetraose and maltopentaose in the data. But the density is not good enough to construct the corresponding oligosaccharide molecules. The binding sites they bind with are also found in the maltohexaose (M6) and maltoheptaose(M7) bound structures, so we finally decided to use the maltohexaose and maltoheptaose bound structures to illustrate the interaction.

Maltohexaose and maltoheptaose bind with *E.coli* BE at 5 binding sites. **Table 2.5** lists the residues involved

Binding site	**Bound chain**	***E.coli* BE residues involved in the binding**
III(on D chain)	M6 and M7	TRP159, LYS189, LEU201,GLN211 and GLU215
IV(on C chain)	M6 and M7	ASP542, TRP544, GLN545, PRO659 and SER689
V (on ABCD chains)	M6 and M7	ARG255, SER583, ASP585, HIS587 and GLU590
VI(on B chains)	M7	LYS546, TRP595 and HIS596
VII(on A B chain)	M7	PRO469, GLY476, TRP478, ASN518 and PHE477

Table 2. 4 Binding sites in *E.coli* BE /linear oligosacchardie complex structures

Among the 5 binding sites, III and IV are the same as the binding sites in *E.coli* BE /CD complex structures. The binding sites can only immobilize two glucose units. The rest of the linear chain is apparently disordered and can not be seen in the electron density map. This is a big difference between linear and cyclic oligosaccharides. (CDs are somewhat rigid. even though only 2 glucose units bind with the protein, the rigidity of the ring fixes the rest of the ring, So most parts of the cyclic oligosaccharides appear in the electron density maps.)

Linear oligosaccharides have 3 unique binding sites compared with cyclodextrins. Those are binding site V, VI and VII. We will discuss these sites in detail.

2.5.1 Three dimensional structure of *E.coli* BE / maltoheptaose complex

Maltoheptaose (M7) is the longest oligosaccharide we used. It binds with *E.coli* BE. And its binding mode on the surface of the protein revealed some very interesting aspects of the BE mechanism.

We discovered that the non-reducing end of maltoheptaose binds with *E.coli* BE at binding site V and the reducing end binds with binding site VI (only on molecule B). **Figure 2.27** shows the $F_o - F_c$ and $2F_o - F_c$ electron density maps calculated at 0.6 and 2.0 σ respectively. The maps depict the electron density of the oligocaccharide at the binding sites. **Figure 2.28** and **2.29** show the detailed interaction between maltoheptaose and *E.coli* BE.

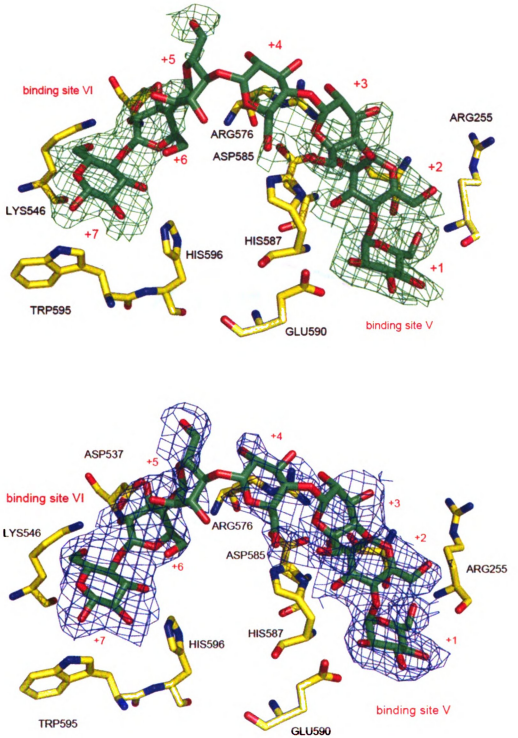


Figure 2. 27 The **top figure** shows the F_o-F_c electron density map of M7 (calculated at 2.0 σ) bound at binding site V and VI on molecule B.; The **bottom figure** shows the 0.6 σ 2FO-F_c electron density map of the M7.

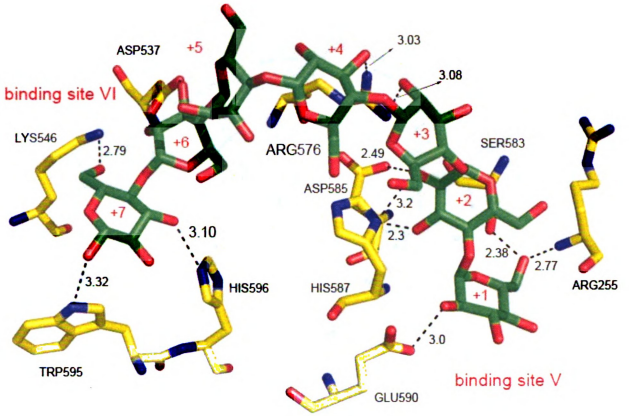


Figure 2.28 Maltoheptaose binds with *E.coli* BE at binding sites V and VI. The figure illustrates the detailed interaction between M7 and the BE B molecule. The non-reducing end of M7 binds with binding site V and the reducing end binds with binding site VI.

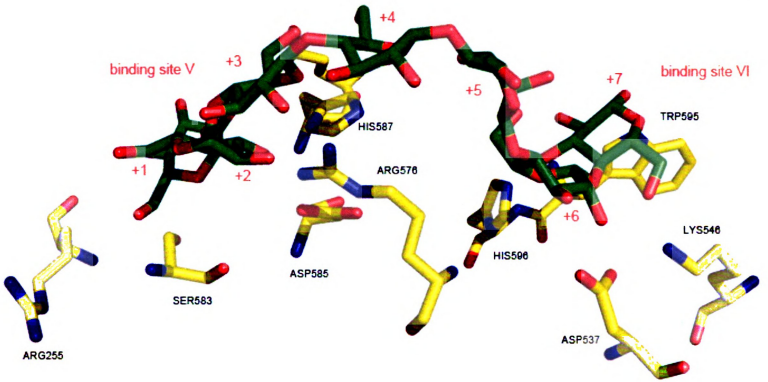


Figure 2.29 Maltoheptaose binds with *E.coli* BE at binding site V and VI. This figure shows the interaction from another angle. It can be seen ARG576 interacts with the oligosaccharide chain.

A linear oligosaccharide has one non-reducing end and a reducing end. Both ends of maltoheptaose bind with the *E.coli* BE protein on the surface. From our structure, we can see that the first two glucose units (+1, +2) bind with binding site V through extensive hydrogen bonds with residues ARG255, GLU590, HIS587 and ASP585. The reducing end interacts with ASP537, LYS546, TRP595 and HIS596.

The whole linear oligosaccharide forms a "U" shape "jump rope" with binding site V and VI as two ends, and the middle of the rope was jacked up by the side chain of ARG576.

ARG576 plays an interesting role in the binding. The side chain of ARG576 makes hydrogen bonds with hydroxyl groups of the +3 and +4 glucose units. It stabilizes the middle of the loop so that it is not very flexible. At the same time, it prevents the middle of the oligosaccharide from contacting the surface of the enzyme. Also because of this orientation, it makes the whole oligosaccharide extend to its maximal length. When the maltoheptaose binds with the surface of the protein, several glucose units rotate along the α-1,4 link. Also distortion inside the glucose ring was observed: the conformation deviates from the most energy favorable chair conformation (please see the glucose ring in previous figures). But the "U" shape jump rope binding is not usual in polysaccharide-protein binding, twisting the chain along such a short distance (7 glucose units) is likely to cause some units to adopt unusual conformations. Also we believe the whole linear chain is under extra stress caused by fully extending to its maximal length (as we mentioned before, the binding on sites V and VI with interaction of ARG576 make

maltoheptaose fully extend to its maximal extent.)

In our structure, we only found whole maltoheptaose bound to the B molecule. In the A and D molecules, only the non-reducing ends bound at binding site V were found. We can see two glucose units at binding site V in the A molecule, and four glucose units in the D molecule. The rest of the oligosaccharide chain remains flexible, so it is not detected in X-ray diffraction data. Also there is no binding in the C molecule. **Figure 2.30** shows the binding in molecule A in which two glucose units are visible.

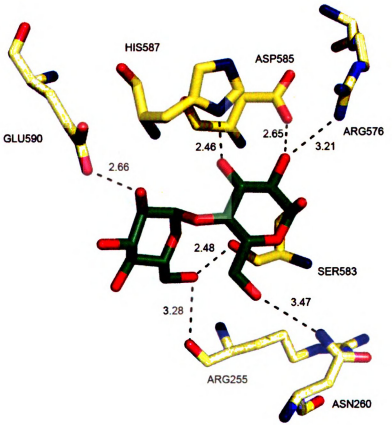


Figure 2. 30 Maltoheptaose binds with *E.coli* BE at binding site V in the A molecule. Only two glucose units are visible.

Binding site VII only exists in the A and B molecules. Because of the flexibility of maltoheptaose , only two glucose units are visible in binding(See **Figure 2.31**).

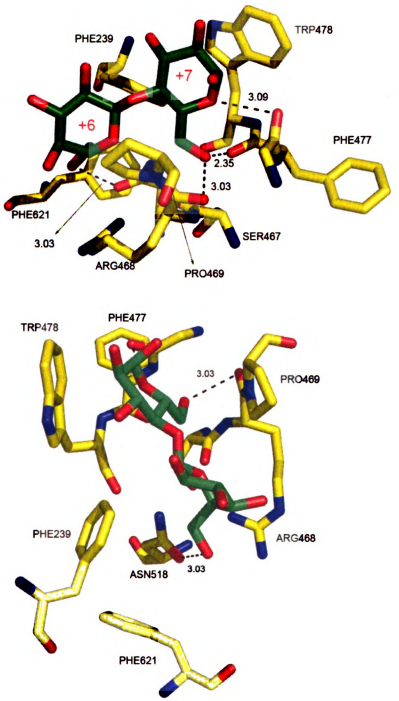


Figure 2. 31 Maltoheptaose binds with *E.coli* BE at binding site VII. Two figures show the interaction from different angles.

The reducing end of M7 binds with this binding site. TRP478 and PRO469 are parallel to each other, forming an inclusive pocket with PHE477 as well. Glucose +7 was anchored inside the pocket. The hydrophobic interaction and hydrogen bonds help to stabilize glucose units. Glucose +6 become flexible although it has a hydrogen bond with ASN518. The ideal orientation of glucose +7 is to be parallel to the side chains of TRP478 and PRO469, but because of the steric hindrance from PHE239 and PHE621, glucose +7's position was a little distorted. Also glucose +6 turns out of the pocket in a skewed way, leading the rest of the oligosaccharide chain away from the surface of BE. +1-+5 glucose units can not be seen because they are flexible. This binding site looks like a rail docking station: side chain of TRP478 and PRO469 act as the two rails, leading the glucose inside by aromatic/hydrophobic interactions.

Due to crystal packing, the same space on the C and D molecules was occupied by side chains of other symmetric molecules. So the binding is only seen at the A and B molecules.

Similar to CDs, linear oligosaccharides also bind with the *E.coli* BE at binding sites III and IV. **Figure 2.32** and **Figure 2.33** illustrate the interaction. Binding site III involves residues TRP159, GLN211, GLU215 and LYS189. This is almost the same as binding between *E.coli* BE and CD. Only two glucose units are visible in the electron density map. Which two glucose units bind with the protein are unknown. The aromatic stacking force between the indole side chain of TRP159 and glucose units, together with hydrogen bonds between the hydroxyl group of glucose and protein stabilize the oligosaccharide on

the surface of protein.

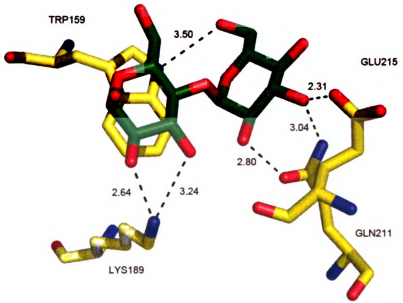


Figure 2. 32 Maltoheptaose binds with *E.coli* BE at binding site III.

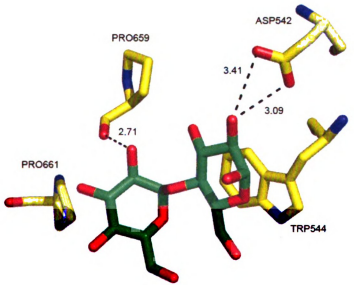


Figure 2. 33 Maltoheptaose binds with *E.coli* BE at binding site IV.

Similarly to the binding at binding site III, the combination of hydrogen bonds and aromatic stacking immobilizes two glucose units on the surface of the protein. The interaction at this binding site only anchors two glucose units. Since both sides of the binding site are quite open, these two glucose units could be any two consecutive units in the maltoheptaose chain.

From the composite view of the B molecule of the M7-bound structure, we can see the "jump rope" is pointing toward the active site cavity, and binding site VII is on the other side of the molecule near the bottom of the (α /β) barrel region (**Figure 2.34**).

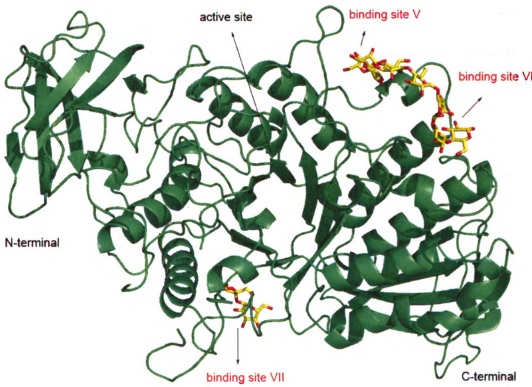


Figure 2. 34 A composite view of the B molecule of *E.coli* BE /M7 structure.

2.5.2 Three dimensional structure of *E.coli* BE / maltohexaose complex

Maltohexaose(M6) binds with *E.coli* BE protein in a pattern similar to that of maltoheptaose. The binding also happens at five binding sites.

At binding site V, usually only one glucose unit is visible in the A molecule and two glucose units were indentified in the B, C and D molecules. The binding patterns are similar to those in the M7-bound structure.

In another M6 bound structure, we found very poor electron density for the entire M6 chain bound at binding sites V and VI, similar to the M7-bound structure. We tried to build M6 at the binding sites. It indicates the binding at both site V and VI are distorted. At binding site V only one glucose has interaction with a protein residue; and at binding site VI, the only bound glucose unit inclines to slide out of the binding site. Since the density is too poor to construct an acceptable molecule, we only keep the glucose units bound at the binding sites which have good density in our refined structure. But this still revealed a quite interesting discovery. Combined with the M7-bound structure, we believe M7 is the shortest oligosaccharide chain that can simultaneously bind site V and VI at the same time. M6 is just a little short: but it still shows some binding. But oligosaccharides shorter than M6 never showed density for the entire chain in the data sets.

In the M6 bound structure, we did not see binding at binding site VII. The binding at site III and IV are similar to those in the M7-bound structure.

After examining all data sets, we believe that except binding at binding site V and VI

in the B molecule, M6 and M7 bind with *E.coli* BE in exactly the same way: part of the oligosaccharide chain (usually two glucose units) binds with BE surface residues, and the rest of the chain remains flexible outside the BE molecule. Thus in this case, the one glucose difference between M6 and M7 actually made no difference with regard to binding.

Figure 2.35 – Figure 2.37 shows the detailed interaction at different binding sites

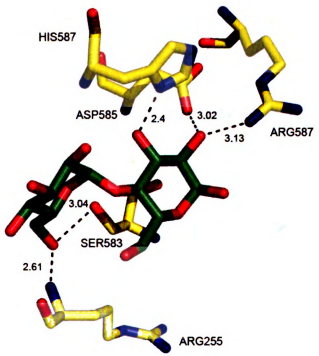


Figure 2. 35 Maltohexaose binds with *E.coli* BE at binding site V in the C molecule.

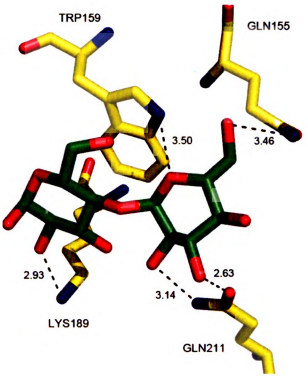


Figure 2. 36 Maltohexaose binds with *E.coli* BE at binding site III in the D molecule.

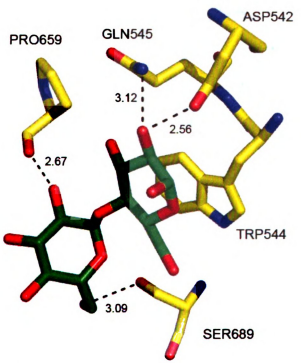


Figure 2. 37 Maltohexaose binds with *E.coli* BE at binding site IV in the C molecule.

At the binding sites, the oligosaccharide was immobilized on the surface of protein by combination of aromatic stacking and hydrogen bonds. This interaction is very similar to those in the BE/CD complex.

2.5.3 Overlay of *E.coli* BE/α-CD structure and *E.coli* BE/M7 structure

We overlaid the binding sites both BE/CD and BE/M7 share: binding site III and IV. As we indicated in the BE/CD structure section, *E.coli* BE actually only interacts with two glucose units of the cyclodextrin.

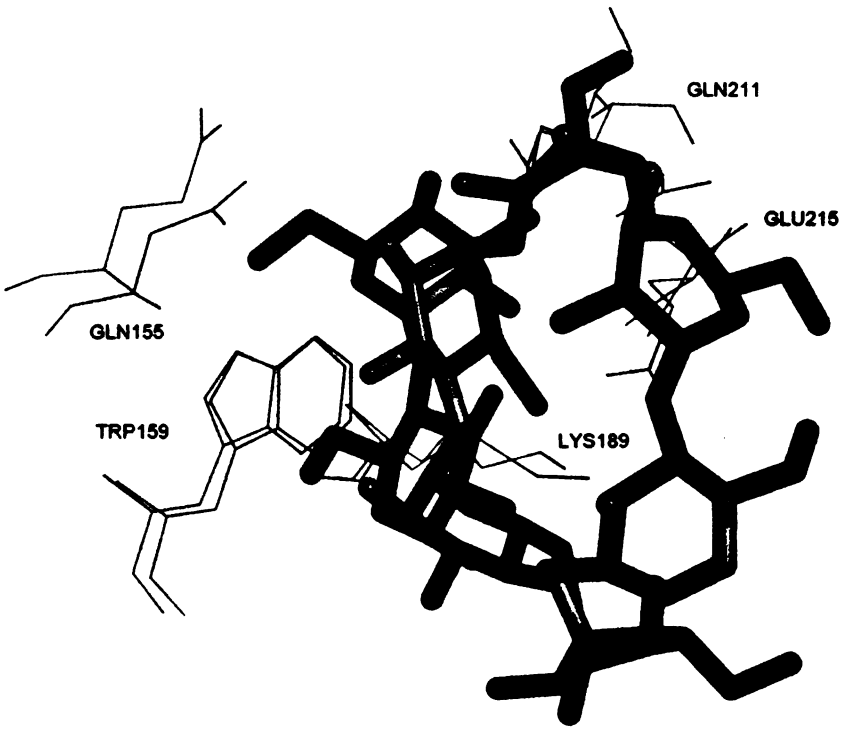


Figure 2. 38 The overlay of BE/α-CD and BE/M7 structures at binding site III. α-CD and its bound residue are represented in green , while M7 and its bound residue are in magenta.

Figure 2.38 shows the difference between α-CD and M7 bound structures at binding site III. In the figure, we can see the difference: first the C6 hydroxy group of M7 glucose is pointing left, while the same hydroxyl group in α-CD is in the opposite direction. All the CDs assume the same orientation at this site since they can adopt a more relaxed conformation at this site with more hydrogen bonds; linear oligosaccharide did not adopt the same orientation as CDs, but we believe there is no big difference in terms of energy.

Also the bound residues GLN155, GLU215 and LYS189 adopt different conformations when they bind with different substrates. It seems that this is caused by the opposite orientation of the sugar units. The hydrogen bonds between the hydroxyl groups of the sugars and protein residues cause the side chains of the residues to move. The major contributor of aromatic stacking, TRP159, overlaid perfectly. This residue keeps parallel to the sugar ring plane, interacting with the sugar through aromatic stacking.

Figure 2.39 depicts the overlay of BE/α-CD and BE/M7 structures. From the overlaid structures we can see the bound residue at the binding site remains in the same position, and the two glucose units bound to protein overlap perfectly. This indicates that the interaction between the oligosaccharide and this binding site are mainly on the two glucose units. From the figure it is clear that the glucose units and bound residues are almost identical except a minor rotamer change of one C6 hydroxyl group. Also the small side chain conformational change of ASP542 does not affect the hydrogen bond with the hydroxyl group of the oligosaccharide.

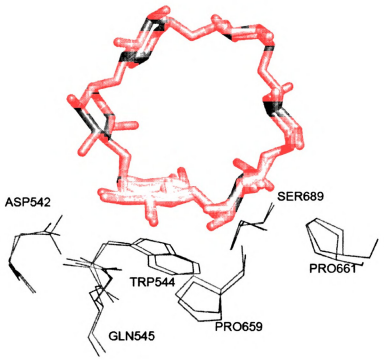


Figure 2. 39 The overlay of BE/α-CD and BE/M7 structures at binding site IV. α-CD and its bound residue are represented in green, while M7 and its bound residue are in magenta

The overlaid structures showed that binding site III and IV are ready to bind both cyclic and linear oligosaccharides without need of extra energy. These two sites are on the surface of protein. It is very likely the natural substrate for BE will keep binding at this site at different catalytic stages in which the substrate might be either naturally double helical or unwound to a linear state.

2.5.4 Overall composite of *E.coli* BE / oligosaccharide binding structure

The overall composite binding figure shows the distribution of CD and oligosaccharide binding sites.

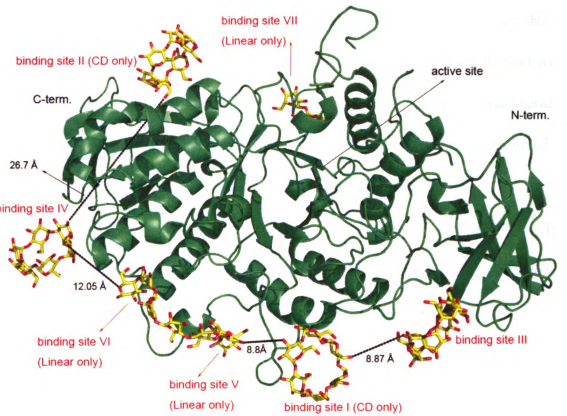


Figure 2. 40 The overall composite binding sites on the *E.coli* BE surface. All 4 molecules of α-CD and 1 molecule of M7 bound structures were overlaid. All 4 α-CDs, 1 M7 and one part of M7 (two glucose units) were superimposed onto one molecule to illustrate the overall binding. The distance between the CDs and M7 molecules are listed.

2.6 Conclusion and hypothesis

In our work, a set of binding sites on the surface of *E.coli* BE was discovered for the first time. Also the binding site's preference for substrate was tested by different substrate soaking experiments. Four binding sites were found for CDs (binding site I-IV), and they in generally bind all CDs similarly. Among the four binding sites, two sites are shared with linear oligosaccharides (binding site III and IV) although binding site III bind with the linear oligosaccharide differently than it does with CDs. Linear oligosaccharides occupy three unique binding sites (binding site V-VII) in which two of them are found to bind respectively with the non-reducing and reducing end of maltoheptaose.

All the detailed interactions between substrates and local bound residues of *E.coli* BE were depicted. By identifying the residues involved, further site directed mutagenesis becomes feasible to determine which sites are the most necessary for the activity.

2.6.1 Hypothesis

After determining the three dimensional structures of CD and linear oligosaccharide bound *E.coli* BE complexes, we realized that none of the binding took place around the active site. We know the CDs are too big to access the active site cavity, but the fact that there was no linear oligosaccharide bound to the active site surprised us. Considering that we did extensive soaking experiments under different conditions, and collected many diffraction data sets, we can conclude safely that the substrates we used so far (from Maltose to Maltoheptaose) can not bind tightly to the catalytic residues in the active site.

From the literature, we know *E.coli* BE prefers to transfer chains about 10-12 glucose units in length (*19, 30*). So we were interested in how to connect this result to our discovery of surface binding sites on the *E.coli* BE protein.

We compared structures of pseudo-maltononaose bound maltohexaose-producing amylase(*5*)(PDB database # 1WPC), maltononaose bound cyclodextrin glycosyltransferase from *Bacillus Circulans*(*6*) (PDB data base # 1CXK), acarbose and oligosaccharide bound alpha-amylase(*31-32*) (PDB database # 1HX0 and 1UA3) and maltohexaose-producing amylase from *Bacillus Circulans* (*33*)(PDB database# 2D3N). These enzymes belong to the glycosyl hydrolase GH13 super family and share similar overall structures. The structures have oligosaccharide or pseudo-oligosaccharide bound at or near the active sites that are highly conserved. Among them, there is even a 9-glucose chain bound at the active site. The active sites are very similar within this family. And from the overlaid structure, we found the substrates also overlaid perfectly. Their orientations and contour of the chains are similar (see **Figure 2.41** top figure).

Figure 2. 41 Overlay of 1WPC (green), 1HX0 (blue), 1M7X (magenta) and 1UA3 (yellow). The top figure shows the overlay of all the enzymes and substrates. The bottom figure shows the close-up of the overlay of 1UA3 (α-amylase), 1M7X (*E.coli* BE) and substrates at the active site.
1WPC: pseudo-maltononaose bound maltohexaose-producing amylase
1HX0: truncated acarbose bound alpha-amylase
1M7X: *E.coli* BE protein without any substrate
1UA3: oligosaccharide bound alpha-amylase

From the overlaid structure we can see the central α/β barrel region is conserved among the proteins. All the substrates follow the same orientation. Since there is no substrate for *E.coli* BE, a truncated acarbose was modeled into the *E.coli* BE active site(*3*). We also overlaid the acarbose model with 1M7X and other structures in **Figure 2.41**, the result proved our hypothesis: the modeled acarbose also superimposed with the other substrates seamlessly.

From all the information gathered, we believe that inside the active site the substrate for *E.coli* BE should also adopt the orientation that other substrates adopted in **Figure 2.41**. Also from the figure it is clear that the BE does not have the substrate binding loops of the other enzymes (such as α-amylase) at the active site (see **Figure 2.41** bottom figure), so the substrate for BE does not need to overcome the loop obstacle to reach into the active site. This is applied next in the modeling of a glucan chain into the active site.

We have found the binding site of linear oligosaccharide on the surface of *E.coli* BE, and we know the orientation of the substrate inside the active site. Based on this information, we modeled a glucan chain from the binding site into the active site of the *E.coli* BE protein. We tried to build the glucan chain over the surface of protein without any clashes with the surface residues. The glucan chain starts from the non-reducing end of M7 (binding site I of M7 bound structure), and ends up overlapping with the modeled acarbose in the active site (See **Figure 2.42**).

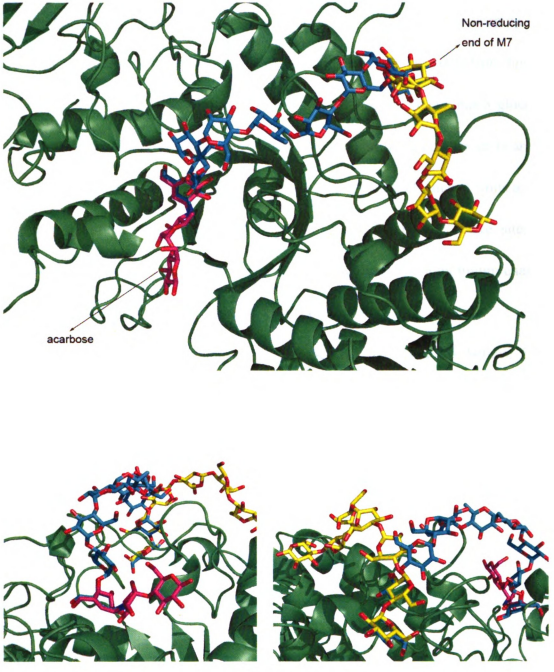


Figure 2. 42 Oligosaccharide model (blue color) binds to binding site V and the active site simultaneously. Maltoheptaose (M7) is depicted in yellow and acarbose in magenta. **Top figure** shows the overall view of the model glucan over the surface of BE. **Bottom left** shows the overlapping part of the glucan with acarbose. **Bottom right** shows the starting point of the glucan: the non-reducing end of M7 (at binding site V)

In the figure, acarbose (in magenta color) was truncated, only keeping the part overlapping with the glucan. The modeled glucan chain (in blue) starts from binding site V, overlapping with the non-reducing end of M7. This glucan chain has 8 glucose units (to the first overlapping glucose unit with acarbose), and three glucose units are needed inside the active cavity (the truncated acarbose) , so at least 11 glucose units are needed to have a simultaneous binding at both binding site V and the active site. This is consistent with the data that *E.coli* BE reacts with amylose with a minimum chain length of 12 units and mostly transfer chain of 11 glucose units (*19*).

From all the information above, we proposed our hypothsis: We believe *E.coli* BE binds with longer polysaccharide similiarly to that in Figure 2.42. The non-reducing end of the chain binds with binding site V, and the chain forms a U shape "jump rope", with the middle of the rope sinking into the active site; And the reducing end binds loosely with binding site VI. The cleavage happens inside the active site cavity, actually in the middle of the chain. This is the reason why the chain has to be at least 12 units long to be catalyzed. Binding site I fixes the chain on the surface of protein, while site V is only a "sliding regulator" for the chain. In other words, when the chain is very long, the extra part of the chain will be floppy outside binding site VI, leaving the appropriate length of chain between site V and VI to make sure the tip of the jump rope interacts with the active site properly.

Actually we observed that the binding site VI is not strictly fixed, by comparing the M6 and M7 bound structure, we realized that binding site VI is an area (like a platform,

several adjacent residues can provide hydrogen bonds), the glucose anchored on it can slide a little and still bind. Comparing this behavior with the rigid binding pattern at site V, we concluded that site VI has two functions: first it regulates the length of the oligosaccharide between site V and VI to make sure the chain can interact with surface and active sites of protein comfortably; Secondly its binding with the non-reducing end (actually maybe not "end") provides extra stabilization for the whole polysaccharide.

We proposed this hypothesis based on our binding structure, and it did explain the substrate specificity of *E.coli* BE: *E.coli* BE handles the chain transferring specificity by separating in space the substrate binding site from the active site of the enzyme, with the substrate making strong interactions with two sugars located 10-11 sugar units away from the reaction. This is the first hypothesis on a pre-catalysis structure of branching enzyme based on enzyme-substrate binding information.

2.6.2 Verification of the hypothesis

We know that short oligosaccharides such as acarbose can not bind with the *E.coli* BE, while BAYe4609 (a polysaccharide molecule containing 17 or more glucose units) is an effective inhibitor for the enzyme(*34*). This could be explained by our hypothesis: BAYe4609 is long enough to bind with both binding site V and VI simultaneously and interact with the catalytic residues in the active site while acarbose is too short.

Based on our hypothesis, if we could do substrate soaking experiments with

oligosaccharides longer than 10 glucose units (to be safe, we will try substrate with 10, 11, 12 units), very likely we will see glucose binding at binding site V, VI and inside the active site. We are currently seeking homogeneous polysaccharides that have 10, 11, 12, 13 and 14 glucose units. The polysaccharides mentioned will be synthesized by our collaborator Professor Xuefei Huang.

Also site directed mutagenesis will be carried out on critical binding residues of the enzyme. The change in the branching activity will verify the role of each residue in the branching action.

Reference cited

1. 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.

2. Binderup, K., and Preiss, J. (1998) Glutamate-459 is important for Escherichia coli branching enzyme activity, *Biochemistry 37*, 9033-9037.

3. 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.

4. Kagawa, M., Fujimoto, Z., Momma, M., Takase, K., and Mizuno, H. (2003) Crystal structure of Bacillus subtilis alpha-amylase in complex with acarbose, *Journal of Bacteriology 185*, 6981-6984.

5. 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.

6. Uitdehaag, J. C. M., Mosi, R., Kalk, K. H., van der Veen, B. A., 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.

7. Katsuya, Y., Mezaki, Y., Kubota, M., and Matsuura, Y. (1998) Three-dimensional structure of Pseudomonas isoamylase at 2.2 angstrom resolution, *Journal of Molecular Biology 281*, 885-897.

8. Sherrod, M. J. (1989) Exploration of Cyclomalto-Oligosaccharide (Cyclodextrin) Chemistry with Molecular Mechanics - Docking Calculations on the Complexation of Ferrocenes with Cyclodextrins, *Carbohyd Res 192*, 17-32.

9. Szejtli, J. (1998) Introduction and general overview of cyclodextrin chemistry,

Chem Rev 98, 1743-1753.

10. Hinrichs, W., Buttner, G., Steifa, M., Betzel, C., Zabel, V., Pfannemuller, B., and Saenger, W. (1987) An Amylose Antiparallel Double Helix at Atomic Resolution, *Science 238*, 205-208.

11. Ball, S., Guan, H. P., James, M., Myers, A., Keeling, P., Mouille, G., Buleon, A., Colonna, P., and Preiss, J. (1996) From glycogen to amylopectin: a model for the biogenesis of the plant starch granule, *Cell 86*, 349-352.

12. 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.

13. Popov, D., Buleon, A., Burghammer, M., Chanzy, H., Montesanti, N., Putaux, J. L., Potocki-Veronese, G., and Riekel, C. (2009) Crystal Structure of A-amylose: A Revisit from Synchrotron Microdiffraction Analysis of Single Crystals, *Macromolecules 42*, 1167-1174.

14. Takahashi, Y., Kumano, T., and Nishikawa, S. (2004) Crystal structure of B-amylose, *Macromolecules 37*, 6827-6832.

15. Sarko, A., and Marchess.Rh. (1966) Crystal Structure of Amylose Triacetate - a Nonintegral Helix, *Science 154*, 1658-&.

16. Gessler, K., Uson, I., Takaha, T., Krauss, N., Smith, S. M., Okada, S., Sheldrick, G. M., and Saenger, W. (1999) V-amylose at atomic resolution: X-ray structure of a cycloamylose with 26 glucose residues (cyclomaltohexaicosaose), *P Natl Acad Sci USA 96*, 4246-4251.

17. Skowron, S. (2006) chemical structure of the three main types of cyclodextrin, *http://upload.wikimedia.org/wikipedia/commons/thumb/5/51/Cyclodextrin.svg/800px-Cyclodextrin.svg.png*.

18. Blennow, A., Vikso-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, *European Journal of Biochemistry 252*, 331-338.

19. Guan, H. P., Li, P., ImparlRadosevich, 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.

20. (1994) Collaborative Computational Project, No. 4. The CCP4 suite: programs for protein crystallography., *Acta Crysta. D50*, 760-763.

21. Harata, K. (1979) Structure Chemistry of Cyclodextrin Complexes, *J Jpn Soc Starch Sci 26*, 198-209.

22. Challa, R., Ahuja, A., Ali, J., and Khar, R. K. (2005) Cyclodextrins in drug delivery: An updated review, *Aaps Pharmscitech 6*, .

23. Loftsson, T., and Brewster, M. E. (1996) Pharmaceutical applications of cyclodextrins .1. Drug solubilization and stabilization, *J Pharm Sci 85*, 1017-1025.

24. Sharff, A. J., Rodseth, L. E., and Quiocho, F. A. (1993) Refined 1.8-Angstrom Structure Reveals the Mode of Binding of Beta-Cyclodextrin to the Maltodextrin Binding-Protein, *Biochemistry 32*, 10553-10559.

25. Knegtel, R. M. K., 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. Parsiegla, G., Schmidt, A. K., and Schulz, G. E. (1998) Substrate binding to a cyclodextrin glycosyltransferase and mutations increasing the gamma-cyclodextrin production, *European Journal of Biochemistry 255*, 710-717.

27. Uitdehaag, J. C. M., van Alebeek, G. J. W. M., van der Veen, B. A., 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.

28. van der Veen, B. A., Uitdehaag, J. C. M., Penninga, D., van Alebeek, G. J. W. M., Smith, L. M., Dijkstra, B. W., and Dijkhuizen, L. (2000) Rational design of cyclodextrin glycosyltransferase from Bacillus circulans strain 251 to increase alpha-cyclodextrin production, *Journal of Molecular Biology 296*, 1027-1038.

29. Boraston, A. B., Healey, M., Klassen, J., Ficko-Blean, E., van Bueren, A. L., 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, *Journal of Biological Chemistry 281*, 587-598.

30. Binderup, K., Mikkelsen, R., and Preiss, J. (2002) Truncation of the amino terminus of branching enzyme changes its chain transfer pattern, *Arch Biochem Biophys 397*, 279-285.

31. Qian, M. X., Nahoum, V., Bonicel, J., Bischoff, H., Henrissat, B., and Payan, F. (2001) Enzyme-catalyzed condensation reaction in a mammalian alpha-amylase. High-resolution structural analysis of an enzyme-inhibitor complex, *Biochemistry 40*, 7700-7709.

32. Payan, F., and Qian, M. X. (2003) Crystal structure of the pig pancreatic alpha-amylase complexed with malto-oligosaccharides, *Journal of Protein Chemistry 22*, 275-284.

33. Kanai, R., Haga, K., Akiba, T., Yamane, K., and Harata, K. (2006) Role of Trp140 at subsite-6 on the maltohexaose production of maltohexaose-producing amylase from alkalophilic Bacillus sp.707, *Protein Science 15*, 468-477.

34. Libessart, N., Binderup, K., and Preiss, J. (1999) Bay e4609, a slow-binding inhibitor of branching enzyme, *Faseb J 13*, A1350-A1350.

CHAPTER 3: OVEREXPERESSION, PURIFICATION AND LIMIT PROTEOLYSIS OF BRANCHING ENZYME II FROM MAIZE ENDOSPERM

3.1 Introduction to maize starch branching enzyme II

Starch branching enzyme (SBE) plays an important role in the bio-synthesis of amylopectin (*1-2*). Using enzymatic methods to modify starch has been emerging for decades. Research has focused on the bio-synthetic pathway enzymes (*2-3*). We discussed the application and need of modified starch in Chapter 1, and discussed the three dimensional structure of *E.coli* glycogen BE in complex with cyclic and linear oligosaccharides. So starch branching enzyme obviously becomes the next research object.

Starch branching enzyme has more isoforms (*3-4*) than glycogen branching enzyme. And the isoforms have different properties such as branching efficiency, substrate specificity and minimum chain length requirement (*3, 5*). Starch branching enzyme II from maize endosperm showed a different substrate specificity and binding pattern from *E.coli* BE (*6-7*). It is known that the α-amylase family enzymes have similar central α/β barrel domains. But the different properties mentioned above are obviously related to structural differences. The subtle difference in their structures must play an important role. Much biochemical research has been done without real structural information. Several

conserved residues have been identified to be necessary for catalysis: ASP386, GLU441, ASP509(*8*) and ARG384(*9-11*) were found necessary for the branching activity. But without structural information, we still do not have an in-depth understanding about their detailed role in the branching function. So determination of the three dimensional structure of SBEII from maize endosperm will be an important achievement in the structure-mechanism study of starch biosynthetic pathway enzymes.

SBEII from maize has 728 amino acids, with a molecular weight of 84 KDa. The coding sequence of maize SBEII was obtained(*12-13*) and inserted into T7-based vector pET-23d (*14*). The isoform was over-expressed in *E.coli* (*15*); protein is purified by DEAE-cellulose chromatography(*15-16*). It was found that by genetically introducing a silent mutation in the maize SBEII coding sequence, the expression level increased significantly(*14*).

In order to study the structure-function relationship of starch branching enzyme, SBEII from maize endosperm was selected as the target protein. Our objective in this research is to determine the three dimensional structure of SBEII by X-ray crystallography, compare the difference between *E.coli* glycogen BE and maize Starch BE to find out the reason why they have different branching specificities.

3.2 Materials and method

3.2.1 Protein over-expression and purification

A glycerol stock containing the plasmid of starch branching enzyme II from maize was obtained from our collaborator, Dr. Jack Preiss. The glycerol stock was plated on an agar plate containing ampicillin (100mg/L). a single colony was picked and incubated in 50ml L.B media with ampicillin (100mg /L) at 37 °C for overnight. Then media was transferred into 1L fresh L.B media with ampicillin (100mg/L). Growing continued at 37°C until optical density at 600nm ($O.D_{600}$) reached 0.5. IPTG was induced into media to final concentration of 0.5 mM. Then the culture was incubated for another 12 hrs at 25 °C before it was harvested (*14-15*). The protein purification was modified as below:

The cell pellet from 6 L of media was washed with 50mM Tris-Acetate, pH 7.5, 50mM NaCl solution, and then resuspended in buffer A (50mM Tris-Acetate, pH 7.5, 10mM EDTA, 5 mM DTT, 10% Glycerol), lysed by sonication and then centrifuged at 5000 rpm for 30 minutes. Then a 3.8M ammonium sulfate solution was added to the supernatant to final concentration of 0.8 M. The mixture was then centrifuged and precipitation was discarded. Next the 0.8 M supernatant was brought to 1.6 M by addition of ammonium sulfate crystals. Lots of protein precipitated, the precipitation was harvested by centrifugation. The protein pellet was then dissolved in a minimum amount

of buffer A, and dialyzed against buffer A at 4 °C overnight. The protein solution was then applied to DEAE-fractogel column. The column was washed with Buffer A/water (50:50), and then the protein was eluted with 0-400mM KCl gradient in buffer A. Protein fractions were pooled together and SDS-PAGE was used to monitor the purity of the protein. The protein eluted out mostly in the wash buffer. This was concentrated and applied to the Source Q anion exchange column. Protein was eluted with 0-200 mM KCl gradient in buffer A. The protein eluted out at about 50 mM KCl. The fractions containing protein were concentrated to a minimum volume and loaded onto the size exclusion column (SEC). Protein was eluted with buffer A. Pure protein eluted in a single sharp peak. From the SEC chromatograph it is clear that the protein is a monomer in solution. Bradford protein assay was used to measure the concentration of the final protein solution; SDS-PAGE was used to monitor the purity and Mass Spectrometry was used to verify the identity of the protein.

Figure 3.1 shows the size exclusion chromatograph and SDS-PAGE of the SBEII protein.

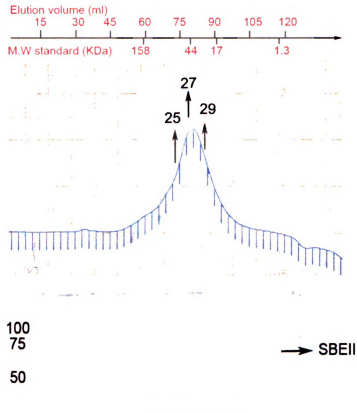


Figure 3. 1 Size exclusion chromatograph and SDS-PAGE of SBEII protein. SBEII protein eluted out of SEC column in a single peak, and SDS-PAGE showed fractions #27 -# 30 (3 ml per fraction) are pure protein.

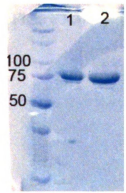


Figure 3. 2 Two final SBEII protein samples (1 and 2) used in crystallization screens.

3.2.2 Crystallization screen on maize starch branching enzyme II (SBEII)

The purified protein (see **Figure 3.2**) was used to set up a sparse matrix crystallization screen at both room temperature and 4 °C using the hanging drop vapor diffusion method. Extensive buffer conditions were attempted; Protein at different concentrations was also tried, but after 10 months no crystals were found.

We believed that the full length SBEII was not prone to crystallize. A different strategy is needed to solve this problem.

In the structural study of *E.coli* BE, a similar situation was encountered (*17*). The full length wild type protein did not crystallize, but deleting a significant segment from the N-terminal of the protein resulted in successful crystallization. Limit proteolysis was used to identify the new truncated version of the protein. We tried to use limit proteolysis to investigate the SBEII from maize and found there was a major product after proteolysis with Trypsin. On SDS-PAGE, a protein band corresponding to about 55 KDa was found after limit proteolysis (see **Figure 3.3**).

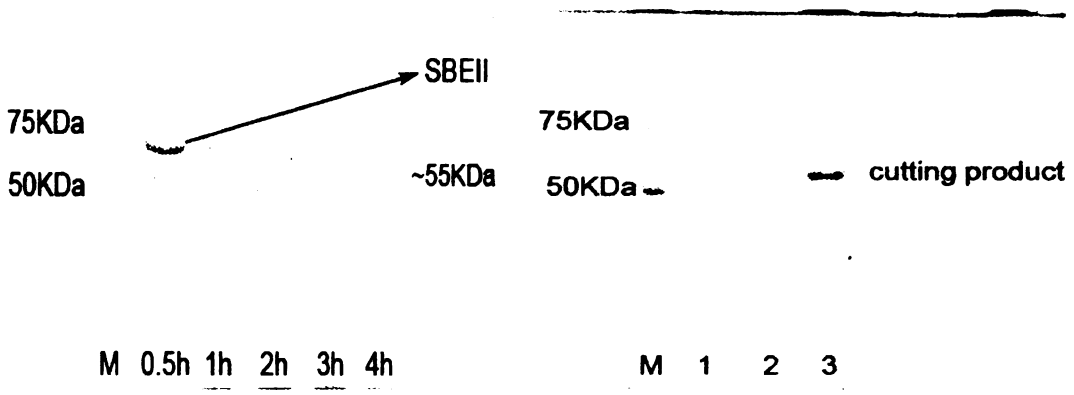


Figure 3. 3 Trypsin limit proteolysis test on SBEII at 4 °C.
Left gel sample shows incomplete cutting with Protein/ Trypsin = 1 mg / 5 units
Right gel sample shows three complete cutting sample (Protein /Trypsin = 1 mg / 7 units)

From the experiment, we found a major product with molecular weight ~55KDa. By controlling the Trypsin protein ratio and digestion time, we obtained a small amount of cut protein. The optimal proteolysis condition is using 7 units Trypsin on 1 mg protein, cutting at 4 °C for 2 hours.

After Trypsin proteolysis, PMSF was added to stop the enzymatic digestion. And the protein solution was concentrated and loaded onto the SEC column. Protein was eluted out with buffer B (50mM Tris-Acetate, pH7.5, 2.5 mM DTT, 10% Glycerol).

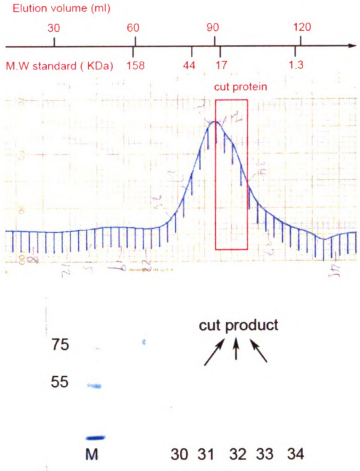


Figure 3. 4 Size exclusion chromatograph and SDS-PAGE of proteolysis product of SBEII protein. The proteolysis was eluted out at the tail of the peak. Its molecular weight is about 55 KDa.

The proteolysis yield was very low: we lost about 80% of the protein in the process. From many similar tests, the recovery yield of the truncated protein is never higher than 30%. This made the direct cutting-purification method very impractical.

We also found the proteolysis process is progressive for SBEII protein. Another low molecular weight cutting product was produced when we used more Trypsin in the

experiment. In a test 10 units Trypsin was added to a 1 mg protein solution. The proteolysis result showed another protein band (corresponding to 30KDa) appeared after 0.5hr (see **Figure 3.5**). This protein is the secondary cutting product of SBEII. From the gel sample, it is clear that the second band's intensity increases (~30KDa protein) and at the same time the first band (~55KDa) faded accordingly. This indicated that the second protein is the proteolysis product of the first protein. It is not the byproduct of the first step cutting. The secondary proteolysis product is about 30 KDa, which indicates it is a relatively small piece of the wild type protein. It is not as important as the first product because the second product is too small to contain the active site (the α/β barrel domain). Furthermore the existence of this overcutting issue makes controlled-digestion very difficult. Given this situation, we have to develop another strategy to obtain the 55 KDa truncated SBEII protein.

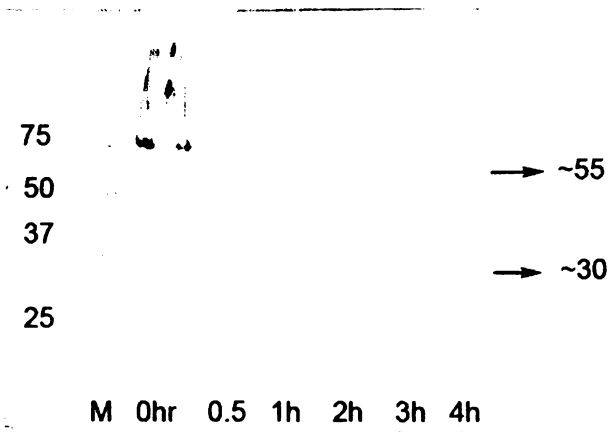


Figure 3. 5 SDS-PAGE of a proteolysis experiment. Protein / Trypsin= 1 mg / 10 units.

3.3 Further experiment plan

Since the direct cutting-purification method did not work as we expected, we decided to use protein engineering technique to sub-clone the coding sequence of the truncated protein into appropriate T7 vector. In order to do identify the Trypsin digesting site, N-terminal and C-terminal protein sequencing needs to be done. The major proteolysis product (55KDa) will be sent for protein sequencing analysis to find out the sequence of the truncated protein. Then we will subclone the coding sequence into several vectors and over-express them in an *E.coli* strain. After that the standard purification method will be also be used to purify the protein. Upon getting pure protein, a crystallization screen will be pursued. X-ray diffraction data will be collected on good protein crystals. The three dimensional structure will be determined by molecular replacement.

Although SBEII from maize is only 25% identical to *E.coli* BE in sequence. But as a member of the α-amylase family, it will share a similar overall structure. So the molecular replacement using *E.coli* BE as model is feasible. When the three dimensional structure of the apo enzyme is obtained, further investigation of its substrate specificity and substrate binding will be planned.

Reference cited

1. Hodges, H. F., Creech, R. G., and Loerch, J. D. (1969) Biosynthesis of Phytoglycogen in Maize Endosperm Branching Enzyme, *Biochimica Et Biophysica Acta 185*, 70-&.

2. Fisher, D. K., Boyer, C. D., and Hannah, L. C. (1993) Starch Branching Enzyme-Ii from Maize Endosperm, *Plant Physiology 102*, 1045-1046.

3. Guan, H. P., and Preiss, J. (1993) Differentiation of the Properties of the Branching Isozymes from Maize (Zea mays), *Plant Physiol 102*, 1269-1273.

4. Boyer, C. D., and Preiss, J. (1978) Multiple Forms of Starch Branching Enzyme of Maize - Evidence for Independent Genetic-Control, *Biochem Bioph Res Co 80*, 169-175.

5. Guan, H. P., Li, P., ImparlRadosevich, 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.

6. Fisher, D. K., Kim, K. N., Gao, M., Boyer, C. D., and Guiltinan, M. J. (1995) A Cdna-Encoding Starch Branching Enzyme-I from Maize Endosperm, *Plant Physiology 108*, 1313-1314.

7. Guan, H. P., ImparlRadosevich, J., Li, P., Zhang, L., Gao, Z., Sun, J. T., and Keeling, P. (1997) Understanding the functions and specificities of maize branching enzyme and starch synthase., *Plant Physiology 114*, 154-154.

8. Kuriki, T., Guan, H. P., Sivak, M., and Preiss, J. (1996) Analysis of the active center of branching enzyme II from maize endosperm, *Journal of Protein Chemistry 15*, 305-313.

9. Libessart, N., and Preiss, J. (1998) Arginine residue 384 at the catalytic center is important for branching enzyme II from maize endosperm, *Arch Biochem Biophys 360*, 135-141.

10. Cao, H., and Preiss, J. (1996) Evidence for essential arginine residues at the active sites of maize branching enzymes, *J Protein Chem 15*, 291-304.

11. Cao, H. P., and Preiss, J. (1999) Site-directed mutagenesis evidence for arginine-384 residue at the active site of maize branching enzyme II, *Journal of Protein Chemistry 18*, 379-386.

12. 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 a Sbe2a cDNA, *Plant Physiology 114*, 69-78.

13. 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, *J Biol Chem 272*, 28999-29004.

14. Libessart, N., and Preiss, J. (1998) High-level expression of branching enzyme II from maize endosperm in Escherichia coli, *Protein Expres Purif 14*, 1-7.

15. Guan, H. P., Baba, T., and Preiss, J. (1994) Expression of Branching Enzyme-Ii of Maize Endosperm in Escherichia-Coli, *Cell Mol Biol 40*, 981-988.

16. 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.

17. Binderup, K., Mikkelsen, R., and Preiss, J. (2000) Limited proteolysis of branching enzyme from Escherichia coli, *Archives of Biochemistry and Biophysics 377*, 366-371.

CHAPTER 4 X-RAY CRYSTALLOGRAPHIC STUDY OF RNA POLYMERASE III TRANSCRIPTION FACOTOR TFIIIB COMPLEX

4.1 Our Research Objective

In order to investigate transcription initiation, the first stage of transcription, the three-dimensional structure of the TFIIIB/DNA complex is essential. Although the binary TBP/DNA and tertiary TBP/Brf/DNA complexes (*1-2*) have been crystallized, the structure of the complete TFIIIB/DNA complex is still not available. Toward this goal, recombinant mutants of individual subunits of TFIIIB, TBP, Brf and B" containing the essential functional core domains were designed, expressed in *E. coli* cells and purified. Various oligonucleotides were designed based on yeast U6 promoter. TBP/Brf/B"/DNA quaternary complexes were made and screened for crystallization.

4.2 Materials and Methods

Our experimental design is to obtain the recombinant TBP, Brf, B" proteins, purified oligonucleotides, and use them to make the TFIIIB/DNA quaternary complexes *in vitro*. The complexes were concentrated and screened for crystallization.

4.2.1 Over-expression and purification of the *S. cerevisiae* TBP protein

The *S. cerevisiae* TBP (61-240) was used in the project. This N-terminal deletion mutant was used to crystallize the TBP/TATA box complex and the TBP/Brf/TATA box complex successfully. The contact between TBP, promoter and Brf is mainly at TBP's C-terminus. So the N-terminal deletion construct (61-240) of the *S. cerevisiae* TBP is used in our crystallization studies. From the previous TBP/TATA, TFIIA/TBP/DNA and TFIIA/TBP/DNA complexes (*1-3*) (see **Figure 4.1**) and recent work on the architecture of TFIIIB complex (*4-5*), it is suggested that this construct is sufficient for the pre-initiation complex assembly. In order to facilitate the purification of the protein, 6x poly histidine tag with a Thrombin cleavage site at the N-terminus was attached to the construct.

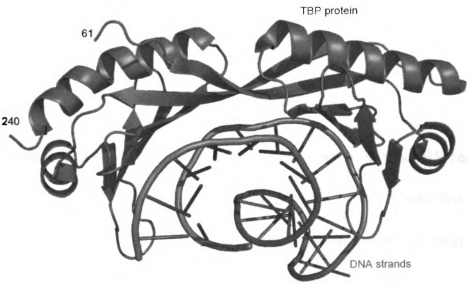


Figure 4. 1 Crystal structure of the TBP-TATA box promoter complex (*1*)

4.2.1.1 Over-expression of the TBP constructs

The pET14 based plasmid encoding the *S. cerevisiae* TBP was transformed into BL21 (DE3) stain of *E.coli*. The transformed *E.coli* sample was plated on an agar media plate with ampicillin (50mg /ml) over night at 37°C. Colonies were used to inoculate in 50 ml Terrific Broth (TB) media containing ampicillin (100mg/ml). After overnight growth at 37°C, the media was transferred into 1L fresh TB media containing ampicillin (100mg/ml). Growth continued until the optical density (600nm) ($O.D_{600}$) reaches 0.8-1.0. The media was cooled to 20°C and the inducer IPTG (0.1mM) induced

producing the protein. The media was grown at 20°C for 10-14 hours before the cells were harvested (*3, 6*).

4.2.1.2 Ni-NTA affinity chromatography purification of the TBP constructs

Cells were harvested and re-suspended in lysis buffer A (25 mM Tris-HCl,pH 8.0, 10% glycerol,500mM NaCl, 50mM Ammonium Acetate) with the protease inhibitor cocktail (EDTA free) from Roche, and sonicated for 5x1 minuts on an ice bath. Cell debris was removed by centrifugation at 6000 rpm for 30 minutes. The clear supernatant containing TBP was loaded onto a Ni-NTA agarose (from Qiagen) column. The His-tagged TBP was immobilized on the Ni-NTA resin. The column was washed with the 50 ml lysis buffer and 100 ml wash buffer A and buffer B (25 mM Tris-HCl,pH 8.0, 10% glycerol,500mM NaCl, 50mM Ammonium Acetate, 20mM imidazole), 30 ml Wash buffer C (25 mM Tris-HCl,pH 8.0, 10% glycerol,500mM NaCl, 50mM Ammonium Acetate, 100mM imidazole), 30 ml wash buffer D (25 mM Tris-HCl,pH 8.0, 10% glycerol,500mM NaCl, 50mM Ammonium Acetate, 250mM imidazole) and 30 ml final elution buffer E (25 mM Tris-HCl,pH 8.0, 10% glycerol,500mM NaCl, 50mM Ammonium Acetate, 400mM imidazole). UV-vis absorption was used to monitor the protein concentration of each fraction. Fractions containing the TBP protein were evaluated by the Bradford Assay and SDS-PAGE gel (see **Figure 4.2**) for concentration and purity. The purified protein fractions were pooled together for histidine tag removal.

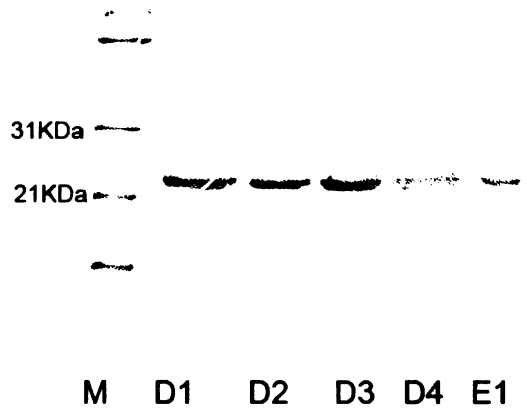


Figure 4. 2 A SDS-PAGE sample of the TBP protein after Ni-NTA column purification. Buffer D and E fractions were shown.

4.2.1.3 Histidine tag removal

Since the histidine tag may affect crystallization, deleting the poly histidine tag was tried before making complexes with the TBP protein. In order to remove the 6x histidine tag (61-240), Thrombin was used to cleave the poly histidine tag from the TBP construct. After trial experiments, the optimal proteolysis condition was determined to be 15 unit Thrombin / 1 mg protein for 2 hours at 4 °C. Under this condition, the 6x histidine tag was removed completely with minimum amount of protein degradation. The result was reproducible. **Figure 4.3** shows a TBP protein proteolysis SDS-PAGE sample.

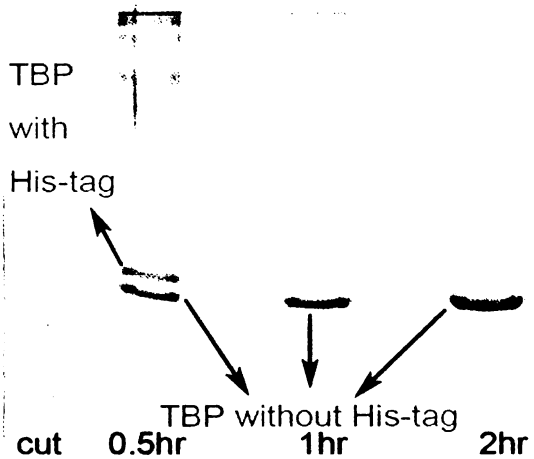


Figure 4. 3 Histidine tag removal in the TBP purification. The first sample was taken after 0.5 hr of cutting and the second and third samples were taken after 1 and 2 hr cutting respectively.

4.2.1.4 Chromatographic purification by heparin column

Heparine resin was used in the last purification step. The TBP protein after the proteolysis step was loaded onto a heparin sepharose (Pharmacia) column. The positive-charged TBP protein bound with the polyanionic heparin sepharose resin. And the protein was eluted against a NaCl gradient (250 mM-1000mM) in the buffer A (20 mM Tris-HCl, pH 8.0, 10% glycerol, 250mM NaCl, 50mM Ammonium Acetate). The protein was eluted at 600-800mM NaCl. The protein fractions were pooled together and concentrated to about 1.3 mg /ml for making complexes.

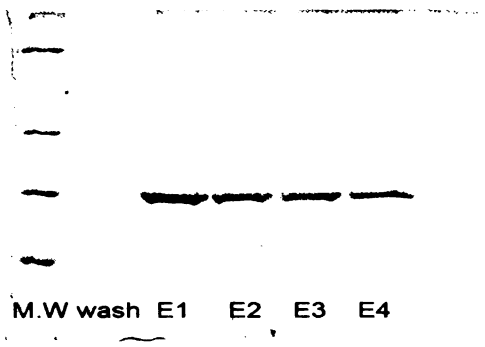


Figure 4. 4 SDS-PAGE samples of TBP protein (without Hisidine tag) after the Heparin column purification. E1-E4 are elution fractions containing most of the pure protein.

4.2.2 Over-expression and purification of the Brf protein

From previous unpublished result (*6*), It is difficult to obtain crystals of the Brf protein with the Zinc ribbon domain. No full length Brf protein from any source has been reported to yield good crystals. Also the N-terminus of the Brf protein has only weak interaction with TBP and DNA, so we decided to exclude it from the protein constructs. Only the C-terminal domain of Brf was used in our investigation (please see **Figure 4.5)**. All of our Brf constructs have 6x poly histidine tag on either the C or N terminus. The *S. cerevisiae* Brf constructs had Thrombin cleavable 6x histidine tag while the 6x histidine

tags on *K .lactis* constructs were non-cleavable. Our collaborator Dr. Steven Hahn kindly provided all the plasmids.

***S.cerevisiae* Brf Construct**		***K. lactis* Brf Construct**	
Mutant number	Construct (cleavable His-tag)	Mutant number	Construct (non-cleavable His-tag)
PSH 582	420-551-$(His)_6$	PSH 552	302-501-$(His)_6$
PSH 583	420-531-$(His)_6$	PSH 553	395-501-$(His)_6$
PSH 584	435-531-$(His)_6$	PSH 554	395-556-$(His)_6$
PSH 585	$(His)_6$-435-531	PSH 555	302-552-$(His)_6$
PSH 586	435-551-$(His)_6$		

Table 4. 1 List of Brf protein constructs

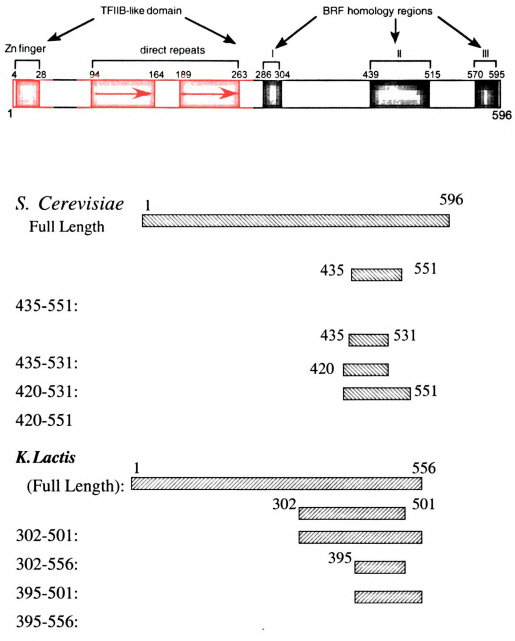


Figure 4. 5 A schematic representation of the *S.Cerevisiae* and *K. lactis* Brf constructs used in our project(*6*). The top figure shows the overall architecture of Brf1 protein from *S.Cerevisiae*

4.2.2.1 Over-expression of the Brf constructs

The Brf protein plasmid was transformed into *E.coli* BL21 (DE3) cells and plated on an Argar plate containing 50 mg/L ampicillin. The plate was incubated overnight at 37°C. A single colony was transferred into 50ml fresh LB media containing 100 mg /ml ampicillin. Cells were grown overnight at 37°C before being transferred into 1L fresh LB media containing 100 mg/ml ampicillin. Growth continued until $O.D._{600}$ reach 0.6-0.8. The inducer IPTG (0.4 mM) was added and shaking was maintained for three hours before the cells were harvested(*6*).

4.2.2.2 Ni-NTA affinity chromatography purification of the Brf proteins

Cells were re-suspended in the lysis buffer (6 M Guanidine-HCl, 0.1 M NaH_2PO_4 , 0.01 M Tris-HCl, PH 8.0) and sonicated on an ice bath. The lysate was centrifuged at 5000 rpm to separate clear supernatant solution. The supernatant was loaded onto a Ni-NTA agarose (Qiagen) column

Ni-NTA agarose (Qiagen) resin was used for 6x Histidine tagged protein purification. The histidine-tagged protein was immobilized onto the Ni-NTA resin. Then the Brf protein was washed with buffer C and D. Buffer E was used to elute the protein fractions. The fractions were monitored by UV-VIS absorption. The protein fractions in buffer D and E were pooled together. The protein was purified under denaturing condition. SDS-PAGE was used to verify the purity of the protein.

The buffers used in the purification include:

lysis buffer :

Buffer A: 6 M Guanidine-HCl, 0.1 M NaH_2PO_4 , 0.01 M Tris-HCl, PH 8.0

Buffer B: 6 M urea, 0.1 M NaH_2PO_4, 0.01 M Tris-HCl, PH 8.0

Wash buffer;

Buffer C: 6 M urea, 0.1 M NaH_2PO_4, 0.01 M Tris-HCl, PH 6.8

Buffer D: 6 M urea, 0.1 M NaH_2PO_4, 0.01 M Tris-HCl, PH 4.5

Elution buffer:

Buffer E: 6 M Guanidine-HCl, 0.2 M Acetic acid

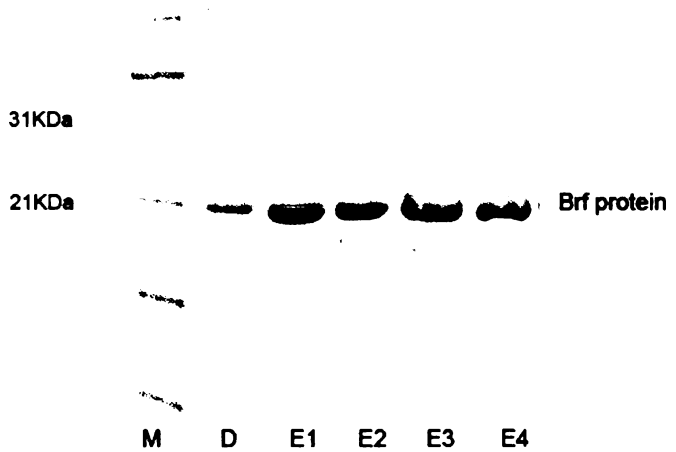


Figure 4. 6 A SDS-PAGE of the *K.lactis* Brf (302-501) after Ni-NTA purification.

4.2.2.3 Protein refolding

The Brf protein was insoluble, so it was purified in denatured state(*6*) and then refolded to its native state. This method was shown to work previously (*7*). Fractions after the Ni-NTA column purification were denatured (*8*).

The protein fractions from the Ni-NTA column were diluted in dilution buffer (6 M Urea, 10% glycerol, 20 mM Tris-HCl, 500 mM KCl, 5 mM DTT) to about 0.3-0.5 mg/ml. Then they were dialyzed against the refolding buffer (10% Glycerol, 20 MM Tris-HCl, 500 MM KCl, 2 mM EDTA, 5 mM DTT, 1 mM PMSF) for 4 x 6 hours. Then the protein was dialyzed against the refolding buffer without PMSF for 2 x 3 hours. The refolding was carried out at 4 °C. Then the protein was concentrated to ~1.5 mg/ml. The *S. cerevisiae* Brf protein was prepared for removal of histidine tag (This step is skipped for *K. lactis* constructs since they had non-cleavable histidine tags.)

4.2.2.4 Removal of cleavable histidine tag from the *S. cerevisiae* Brf protein

Thrombin was used to cut off the poly histidine tag from the *S. cerevisiae* Brf constructs. 2.5 unit Thrombin /mg protein was the best ratio at 4 °C. No degradation was found until 18 hours. **Figure 4.7** shows the cleavage test result of one Brf construct.

After the cleavage of the histidine tag, PMSF was added to the protein solution to a final concentration of 1mM to inhibit the Thrombin to prevent further degradation of the protein.

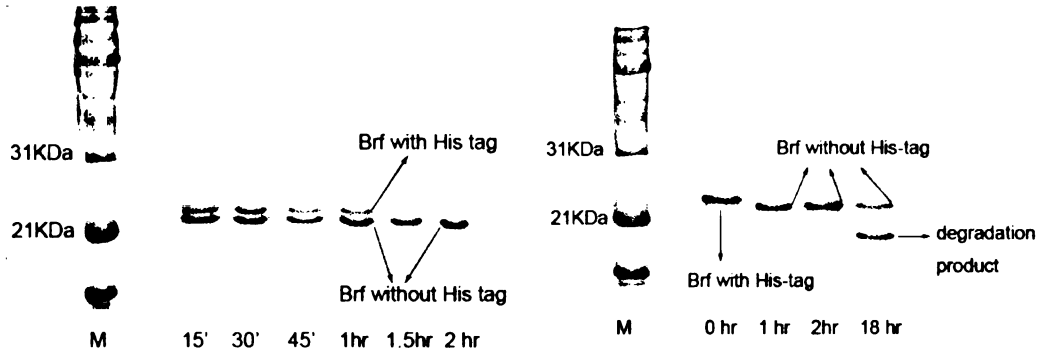


Figure 4. 7 The cleavage of His-tag on Brf (420-531): left figure shows the cleavage test, samples were taken at different time. And at 2 hr the cutting is complete. The right figure shows the overcutting product appearing after 18 hrs. The test was carried out at 4 °C, Thrombin / protein ratio is 2.5 unit Thrombin / 1mg protein.

4.2.2.5 FPLC ion exchange chromatography purification

A Pharmacia FPLC system was used to purify the refolded protein without histidine tag at 4°C. The Source 15 Q matrixes (anion exchange resin with an average particle size of 15μm) were used in the column. The protein to be purified was diluted with FPLC dilution buffer (20mM Tris-HCl, 5mM DTT, 10% Glycerol, PH 8.5), then loaded onto the Source-Q column. Buffer A (20mM Tris-HCl, 50mM KCl, 5mM DTT, 10% Glycerol, PH 8.5) and buffer B (20mM Tris-HCl, 1 M KCl, 5mM DTT, 10% Glycerol, PH 8.5) were used to elute the protein. The gradient was from 50 mM KCl to 1 M KCl. The Brf proteins eluted out at the salt concentration of ~200mM KCl. The fractions were evaluated by Bradford Assay and SDS-PAGE gel for their concentration and purity. After the ion exchange purification, all the Brf proteins were pure enough for making complexes. **Figure 4.8** shows the SDS-PAGE sample of final *K.lactis* Brf proteins.

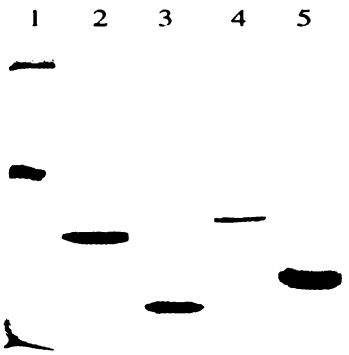


Figure 4. 8 A SDS-PAGE sample of purified *K. lactis* Brf proteins. 1: M.W marker; 2: *K. lactis* Brf (302-501); 3: *K. lactis* Brf (395-501); 4: *K. lactis* Brf (302-556); 5: *K. lactis* Brf (395-556).

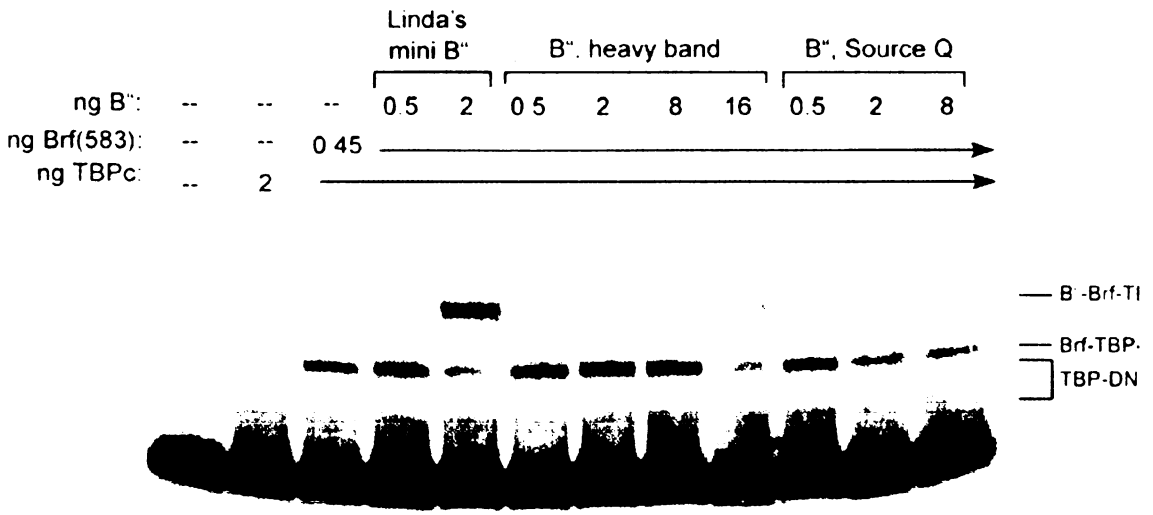


Figure 4. 9 A Gel mobility shift assay using U6 promoter probe, and TBP_c and indicated amount of Brf and B''(*9*).

The *S. Cerevisiae* Brf (420-531) constructs were sent to Dr. Steven Hahn for binding assays. **Figure 4.9** shows the Gel mobility shift assay of PSH583 (420-581) on U6 promoter probe. The assay confirmed binding of our Brf, TBP and B" constructs on the U6 promoter.

4.2.2.6 New *S.cerevisiae* Brf construct design and purification attempt

Besides the current Brf constructs, we also designed a longer Brf construct (69-531) and five internal deletion mutants of the *S.cerevisiae* Brf construct. Based on the reported TBP-Brf-DNA structures and secondary structure predictions on the Brf protein, we believe that internal deletion constructs (69-531Δ266-407), (69-531Δ280-420), (69-531Δ323-407), (69-531Δ335-420) and (69-531Δ266-430) will bind with the TBP, B" proteins and DNA, and do not have the existing floppy loop that might be a detrimental factor in the complex crystallization.

The constructs were cloned and sequences were verified. These constructs of the Brf protein were also insoluble. After purification in denatured condition, protein refolding was carried out. Unfortunately, all of the constructs precipitated during the refolding step in all attempts. After extensive trials at different conditions, we believed that those constructs can not be refolded correctly to their natural state. Thus those constructs are not suitable for the crystallization trials.

4.2.3 Over-expression and purification of the B" protein

Different B" constructs were tested extensively for their abilities of binding the TBP, Brf protein and TATA containing promoters(*9*). It was determined that the construct

containing the central SANT domain can bind with the other components of TFIIIB-DNA complex. We obtained the plasmid of B" (240-520) with uncleavable poly Histidine tag on its C-terminus from our collaborator, Dr. Steven Hahn. The construct encompasses the SANT domain and has been tested to have full activity(*9*). We sub-cloned the coding sequence of B" (240-540) and B" (265-540) into a pET vector with a SUMO domain and a cleavable poly His-tag (made by Geiger lab). The construct design was based on secondary structure predictions: complete secondary structure elements (alpha helix or beta sheet) were kept intact while at the same time the whole SANT domain was included in the construct. The B" constructs were also shown to be active(*9*).

4.2.3.1 Over-expression of the B" (240-520)

The plasmid of the B" mutant was transformed into *E. coli* BL21 (DE3) cells. Cells were plated on an agar plate containing 50mg/L ampicillin and 30mg/L kanamycin. After inoculated overnight at 37°C, a single colony was used to inoculate 50 mL of LB media containing 100mg/ml ampicillin and 60 mg/ml kanamycin. After overnight growth at 37°C, the media was transferred into 1L fresh LB containing ampicillin and kanamycin. Growth continued at 37°C until $O.D._{600}$ of the media reach 0.5-0.6. The media was cooled to 30°C and IPTG (0.5mM) was added. After 3 hours, cells were harvested by centrifugation.

4.2.3.2 Ni-NTA purification of the B" protein

Ni-NTA agarose resin (from Qiagen) was used to purify the His-tagged B" mutants. The purification was under native condition. The buffers used included:

Lysis buffer: 20mM HEPES, 300mM NaCl, 10% Glycerol, PH 8.0, 5 mM 2-mercaptoethanol (BME) and protease inhibitor added before use)

Wash buffer: lysis buffer plus 20mM imidazole, PH8.0, BME and PMSF added before use.

Elution buffer:lysis buffer plus 100mM imidazole, PH8.0, BME and PMSF added before use.

The cells were re-suspended with a protease inhibitor tablet in the lysis buffer and sonicated on an ice bath. The supernatant after centrifugation was loaded onto a Ni-NTA column. The column was washed with the lysis buffer, wash buffer and elution buffer. Fractions containing the B" protein were evaluated by Bradford Assay and SDS-PAGE gel. The protein fractions were pooled together for further purification.

4.2.3.3 FPLC ion exchange chromatography purification

A Pharmacia FPLC system was used to purify the the B" mutants at 4°C. The Source 15 Q matrix was used. The protein solution was loaded onto the Source-Q column. Buffer A (20mM HEPES, 75mM NaCl, 5mM BME,1mM PMSF, 10% Glycerol, PH 8.0) and buffer B (20mM HEPES, 500mM NaCl, 5mM BME,1mM PMSF, 10% Glycerol, PH 8.0) were used to elute the protein. The gradient was from 75mM NaCl to 500 mM NaCl.

The B" protein fractions eluted over a wide salt concentration (100 ~ 200 mM NaCl). The fractions were evaluated by Bradford Assay and SDS-PAGE gel.

4.2.3.4 Cleavage of poly His tag from the B" (240-540) and B" (265-540)

The B" (240-540) and B" (265-540) have cleavable poly Hisidine tags. SUMO protease was used to cut off the poly hitidine tags. The purified protein with Histidine tag was mixed with SUMO protease (prepared by our lab) at a ratio 1mg protein / 0.001 mg SUMO protease. The cutting was complete at 4 °C after 2 hours. After the cutting, the mixture was loaded onto a Ni-NTA column containing 1 ml Ni-NTA resin. The solution containing the cleaved B" was collected. The protein was concentrated to about 2 mg /ml for making complexes (**Figure 4.10).**

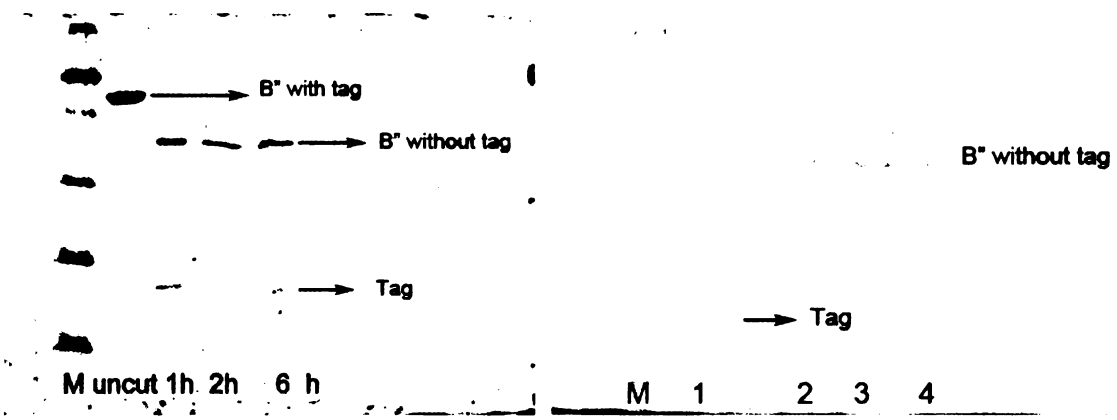


Figure 4. 10 A SDS-PAGE sample of B"(240-540) His-tag cleavage. Left shows time course reaction of SUMO protease cleavage. It can be seen that cutting is complete after 1hr. Right figure: 1 is cutting mixture containing B" protein and the cut tag; 2, 3 and 4 are flow through of Ni-NTA column showing SUMO His-tag removal

.

4.2.4 Purification of oligonucleotides

In the RNA pol III pre-initiation complex, the promoter anchors all the three subunits of TFIIIB. The interaction between the proteins and DNA requires certain base pairs available both upstream and downstream of the TATA box. It was determined that full length proteins have interaction with DNA from 15 b.p upstream to 10 b.p downstream (*10-11*). Considering we were using the truncated Brf, B" and TBP proteins, which might have less interaction with the promoter, the formation of the complex might require less nucleotide, so we decided to use a series of DNAs with different lengths for the complex making.

DNA	Sequence of DNA
iiib44I	CGTCCACTATTTTCGGCTACTATAAAAGAATGTTTTTTTCGCAA
iiib44II	CACTATTTTCGGCTACTATAAAAGAATGTTTTTTTCGCAACTAT
iiib45II	TCGTCCACTATTTTCGGCTACTATAAAAGAATGTTTTTTTCGCAA
iiib46I	TCCACTATTTTCGGCTACTATAAAAGAATGTTTTTTTCGCAACTAT
iiib47II	TCGTCCACTATTTTCGGCTACTATAAAAGAATGTTTTTTTCGCAAACT
iiib47III	CGTCCACTATTTTCGGCTACTATAAAAGAATGTTTTTTTCGCAACTA
DNA1	CTATTTTCGGCTACTATAAAAGAATGTTTTTT
DNA2	CTATTTTCGGCTACTATAAAAGAATGTTTTTTTC
DNA3	CACTATTTTCGGCTACTATAAAAGAATGTTTTTT
DNA4	CACTATTTTCGGCTACTATAAAAGAATGTTTTTTTC
DNA5	TCCACTATTTTCGGCTACTATAAAAGAATGTTTTTT
DNA6	TCCACTATTTTCGGCTACTATAAAAGAATGTTTTTTTC
DNA19	CTATAAAAAAATGTTTTTT
DNA25	CGGCTACTATAAATAAATGTTTTTT
DNA26	TCGGCTACTATAAATAAATGTTTTTT
DNA27	TTCGGCTACTATAAATAAATGTTTTTT
DNA29	TTTTCGGCTACTATAAATAAATGTTTTTT
DNA30	ATTTTCGGCTACTATAAATAAATGTTTTTT

Table 4. 2 DNA used in our project

The designed oligonucleotides (single strands of DNA) were ordered from Keck facility, Yale University. The purification and annealing were done in our lab.

Oligonucleotides ordered were purified with a Perkin Elmer HPLC system equipped with a Source-Q column. The single strands of DNA were loaded onto the column. Buffer A (10mM NaOH, 0.2M NaCl) and buffer B (10mM NaOH, 1M NaCl) were used in gradient. The oligonucleotide fractions collected were neutralized by Tris buffer and diluted to pH 7.5. Then they were loaded onto a DEAE column. DNAs eluted out with buffer C (10mM Tris, 1M NaCl, PH 7.5). Then they were concentrated and annealed in equivalent amount(*6*).

4.2.4 Making quaternary complexes and crystallization screening

One TBP construct, 8 Brf constructs, 3 B" constructs and 18 DNAs were used to form complexes (in our project, about 130 complexes were made). Quaternary complexes were formed by mixing TBP, Brf, B" and DNA at a ratio of 1:1:1:1.2 on an ice bath for about 30 minutes. Then the complexes were concentrated to about 200μL. The final concentrations of complexes for crystallization are ~6-10 mg/ mL. SDS-PAGE was used to verify complex formation. **Figure 4.11** shows the Comassie Blue stained gel sample of one quaternary complex. **Figure 4.12** is one sample UV scan of the TFIIIB only, oligonucleotide and the final pure quaternary complex.

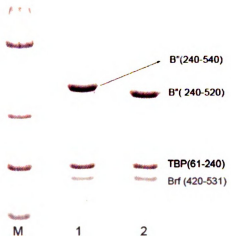


Figure 4. 11 SDS-PAGE of two quaternary complexes. Protein bands were shown on Commasie blue gel. DNA is not stained in Commasie blue dye.

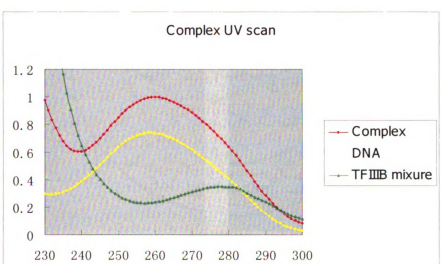


Figure 4. 12 UV spectrum of the TFIIIB protein, DNA and pure quaternary complex (after size exclusion purification). TFIIIB has λ_{max} at 280 nm while DNA at 260 nm. The complex has maximal absorption at about 260 nm, which confirmed the binding between TFIIIB proteins and the oligonucleotide

Sparse Matrix Crystallization screening yielded some tiny crystals, but could not be improved. The sizes of crystals were smaller than 0.005 mm x 0.005mm x 0.005 mm. **Figure 4.13** shows four pictures of putative quaternary complex crystals.

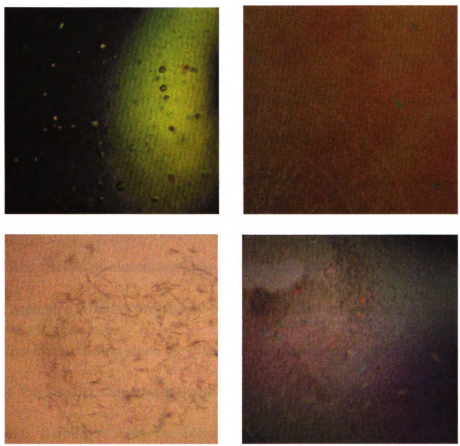


Figure 4. 13 Photos of putative quaternary complex crystals. The four photos were taken from different complex crystal screening boxes. All the crystals were too small, and showed no diffraction. Crystals in the top two photos are from (3.4 M 1,6 hexanediol, 0.1 M Tris, pH 8.5, 0.2 M $MgCl_2$); crystals from bottom left photo are from (30% MPD, 0.1 M Sodium Cacodylate, 0.2 M NaOAc, pH 6.5); crystals from bottom right photo are from (30%PEG4000, 0.1M Tris, pH 8.5, 0.2 M NaOAc).

4.2.5 SEC chromatographic purification of quaternary complexes

We used proteins and oligonucleotides to make complexes, and then used the complexes for crystallization screening. After extensive screening failed to yield good crystals, we realized that forming complex might not be perfect. If the four components of the pre-initiation complex were not mixed in a perfect ratio, there could be a mixture of one or two (TBP/DNA), or even three (TBP/Brf/DNA) components. In this case, homogeneity of the solution will be compromised and crystallization will be difficult. Since it is difficult to get a perfect ratio because there is always some error in determining the concentration of the protein and DNA, we sought another way to guarantee the correct stoichiometry.

Size exclusion chromatography (SEC) was used to further purify the complexes. SEC separates different proteins or complexes based on their sizes. Bigger complexes or proteins elute out first since they can not be retained inside the porous stationary matrix, while the smaller components will be retained inside the matrix much longer. In our complex forming process, the excess protein, binary or tertiary complex will be eluted out later then the correct quaternary complex since the quaternary complex is the biggest in size or molecular weight.

We mixed the proteins and oligonucleotide together to form a complex, then purified the complex by size exclusion chromatography (SEC). **Figure 4.14** shows the SEC chromatograph of a complex TBP(61-240)/ Brf(435-551) /B"(240-520)/DNA1. #30 and 31 fractions (2 ml each fraction) contain the quaternary complex. In some case, we found

we still had binary and tertiary complexes which resulted in the second and third peaks on the SEC chromatograph. For example, **Figure 4.15** shows the SDS-PAGE of the complex fractions after the SEC purification. From the result, it confirmed that the quaternary complex formed *in vitro*, and was stable enough to survive the mild SEC separation. But it was also clear that other complexes formed when we mixed the proteins and oligonucleotide. Without SEC purification, there will be several kinds of complexes in the crystallization solution. The homogeneity of the solution will be significantly impaired, which is fatal to the crystallization screening.

We used this method to purified 10 complexes, and collected the right fractions containing only the quaternary complex. The fractions were concentrated and used for crystallization screening. But still no good crystals were found.

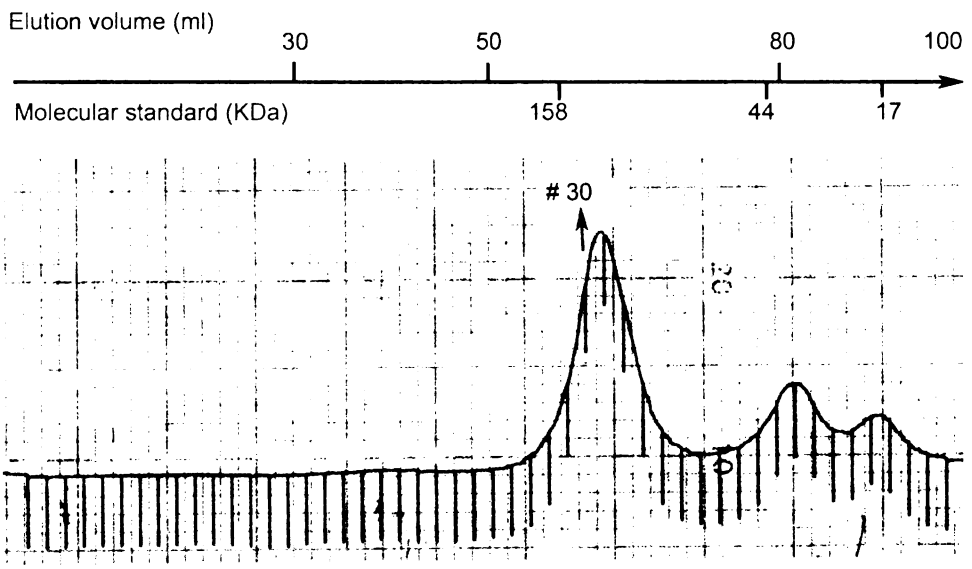


Figure 4. 14 SEC chromatograph of quaternary complex TBP(61-240)/ Brf(435-551) B"(240-520)/DNA1. The complex is in #30, #31 fractions.

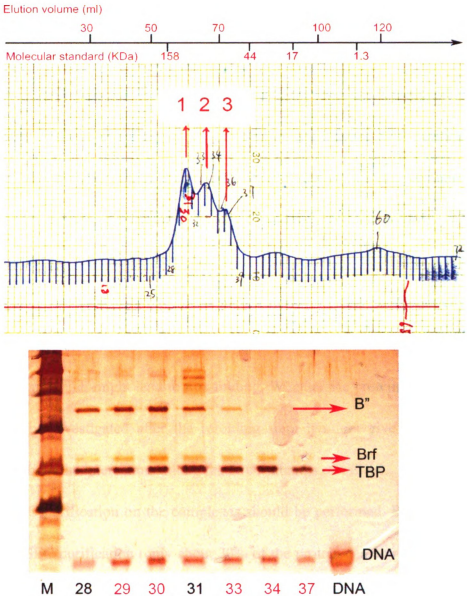


Figure 4. 15 The SEC chromatograph and SDS-PAGE of a quaternary complex TBP(61-240) /Brf (407-531) /B"(265-540)/DNA1 (see Table 4.2). In top figure, peak 1 is the TBP-Brf-B" /DNA quaternary complex (#28, 29 and 30 in the gel sample); peak 2 is the TBP-Brf-DNA tertiary complex (#33 and 34 in gel sample); peak 3 represent the TBP/DNA binary complex (#37 in gel sample). The lower figure is the silver stained SDS-PAGE sample of the fractions purified by SEC.

4.3 Conclusion and future plan on this project

In the TFIIIB-DNA pre-initiation complex project, we purified three subunits of the TFIIIB factor, and formed the complete pre-initiation complexes *in vitro*. Extensive crystallization screenings were attempted on about 130 complexes. Despite all the effort, unfortunately we only got some tiny crystals that do not diffract. SEC purification was adopted for 10 complexes, which theoretically should improve the complex homogeneity and increase the chance of crystallization, but crystals were still not yielded.

We purified the truncated versions of the three subunits of the TFIIIB and made the complexes *in vitro*. We should have done more extensive assays on the activity of each protein, especially the Brf and B" protein. The Brf protein was denatured and refolded in the purification, which might lead to mis-folding. Whether the protein was still active *in vitro* should be investigated after the refolding step. It might give us more accurate information about the interaction among the three subunits of TFIIIB.

More SEC purification on the complexes should be performed. But considering the low yield of SEC purification (only about 30% of the protein can be recovered) we need much more protein and oligonucleotide to finish the project.

We believe an alternative route for making the pre-initiation complex could be the co-expression of the three subunits and purification of the TFIIIB together. This should avoid the risky refolding step for the Brf protein. Also the binding between the three subunits should be close to the real natural binding *in vivo,* which is definitely preferred. This should give us a higher chance of forming good crystals.

Reference Cited

1. Kim, Y. C., Geiger, J. H., Hahn, S., and Sigler, P. B. (1993) Crystal-Structure of a Yeast Tbp Tata-Box Complex, *Nature 365*, 512-520.

2. Juo, Z. S., Kassavetis, G. A., Wang, J. M., Geiduschek, E. P., and Sigler, P. B. (2003) Crystal structure of a transcription factor IIIB core interface ternary complex, *Nature 422*, 534-539.

3. Geiger, J. H., Hahn, S., Lee, S., and Sigler, P. B. (1996) Crystal structure of the yeast TFIIA/TBP/DNA complex, *Science 272*, 830-836.

4. Vilalta, A., Trivedi, A., and Johnson, D. L. (1996) Mutation in the tata-binding protein (TBP) selectively abolishes both RNA polymerase III transcription and the interaction of TBP with a TFIIIB subunit., *Mol Biol Cell 7*, 2704-2704.

5. Colbert, T., Lee, S., Schimmack, G., and Hahn, S. (1998) Architecture of protein and DNA contacts within the TFIIIB-DNA complex, *Mol Cell Biol 18*, 1682-1691.

6. Jin, X. (2002) X-ray crystallographic studies of RNA Polymerase III transcription factor TFIIIB and 1L-*MYO*-Inositol 1-phosphate synthase, In *Department of Chemistry*, Michigan State University, East Lansing.

7. Librizzi, M. D., Moir, R. D., Brenowitz, M., and Willis, I. M. (1996) Expression and purification of the RNA polymerase III transcription specificity factor IIIB70 from Saccharomyces cerevisiae and its cooperative binding with TATA-binding protein, *J Biol Chem 271*, 32695-32701.

8. Sadana, A. (1995) Protein Refolding and Inactivation during Bioseparation - Bioprocessing Implications, *Biotechnol Bioeng 48*, 481-489.

9. Hahn, S. Gel mobility shift assay on Brf and B" protein, Fred Hutchinson Cancer Research Center, seatle.

10. Kassavetis, G. A., and Geiduschek, E. P. (2006) Transcription factor TFIIIB and transcription by RNA polymerase III, *Biochem Soc T 34*, 1082-1087.

11. Huang, Y., and Maraia, R. J. (2001) Comparison of the RNA polymerase III transcription machinery in Schizosaccharomyces pombe, Saccharomyces cerevisiae and human, *Nucleic Acids Res 29*, 2675-2690.

APPENDIXES

Appendix 1. X-ray data collection statistics of the ligand-bound *E.coli* BE protein

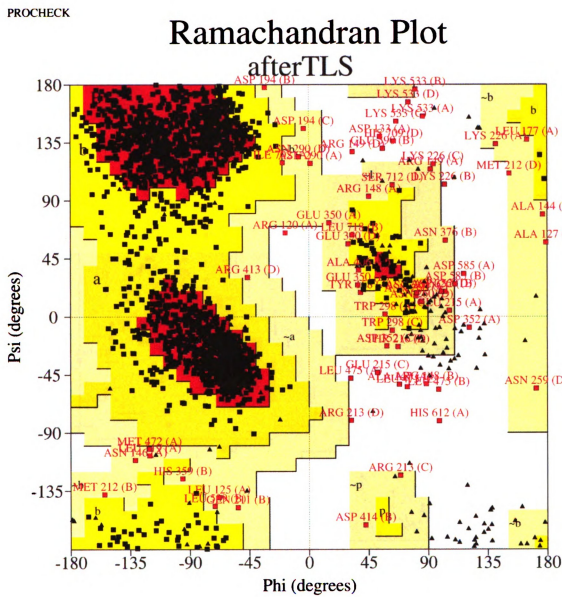
Substrates	α- CD	β-CD	γ-CD	M6	M7
Spacegroup	P 2_1	P 2_1	P 2_1	P 2_1	P 2_1
Z (mol/asymm. unit)	4	4	4	4	4
Unit cell dimensions					
a (Å)	93.7	93.7	93.7	93.7	93.7
b (Å)	104.1	104.1	104.1	104.1	104.1
c (Å)	186.7	186.7	186.7	186.7	186.7
α (°)	90.0	90.0	90.0	90.0	90.0
β (°)	91.8	91.8	91.8	91.8	91.8
γ (°)	90.0	90.0	90.0	90.0	90.0
Wavelength (Å)	0.9785	0.9785	0.9785	0.9785	0.9785
Resolution range (Å)	50-2.60	50-1.88	50-2.35	50-2.55	50-2.40
	(2.69-2.60)	(1.95-1.88)	(2.43-2.35)	(2.64-2.55)	(2.49-2.40)
Redundancy	3.6(3.5)	3.8(2.5)	3.7(3.7)	3.7(3.7)	3.2 (3.2)
Average I/σ	19.8 (2.4)	18.5 (2.1)	27.7 (4.3)	29.4 (4.2)	19.1(2.7)
Total reflections used	2519768	2776967	2913803	3544223	3565445
Unique reflections	108134	281824	147874	116155	136725
Completeness	95.4 (93.3)	99.7 (99.2)	95.2 (95.7)	96.3 (97.1)	89.3 (85.4)
R_{merge} (%)	8.2 (45.9)	9.5(52.6)	8.7(46.8)	8.0(43.9)	8.3(48.6)

Values in parentheses refer to the last resolution shell.

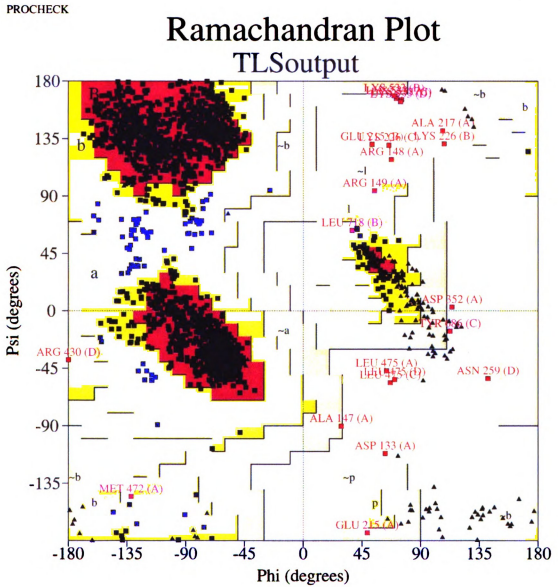
Appendix 2.Structure refinement statistics of the ligand-bound *E.coli* BE protein.

Substrate	α- CD	β-CD	γ-CD	M6	M7
Average B-Factor (Å)	45.4	28.2	40.5	42.4	44.1
R_{work} (%)	17.8	19.6	16.4	18.2	17.9
R_{free} (%)	25.2	23.3	21.7	24.2	24.2
RMSD from ideality					
Bond length (Å)	0.017	0.011	0.016	0.014	0.012
Bond angle (°)	1.874	1.325	1.325	1.617	1.387
Ramachandran plot					
Most favored (%)	80.1	87.7	85.9	82.5	83.2
Allowed (%)	16.7	11.2	12.4	15.4	14.1
Generously allowed (%)	2.1	0.3	1.0	1.4	1.3
Disallowed (%)	1.0	0.7	0.7	0.7	1.4

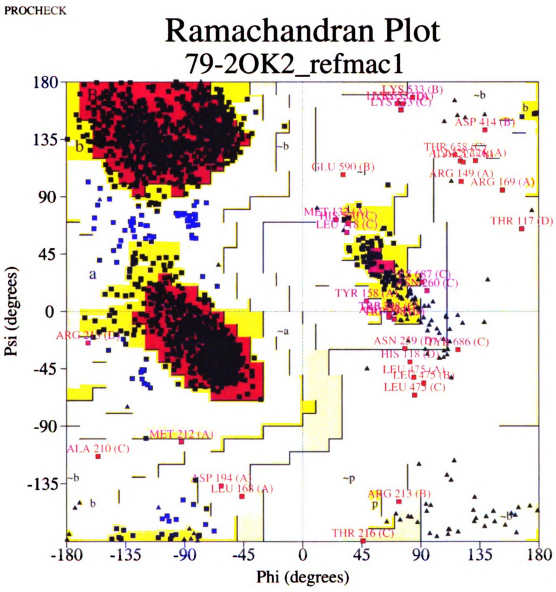
Appendix 3. Ramachandran plot of the BE/alphaCD structure



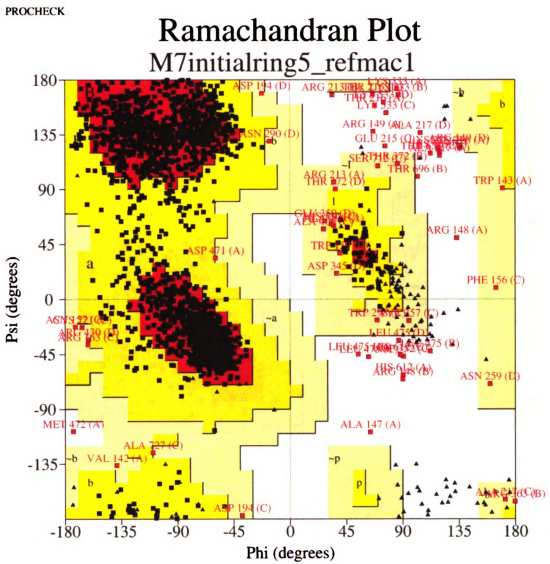
Appendix 4. Ramachandran plot of the BE/betaCD structure



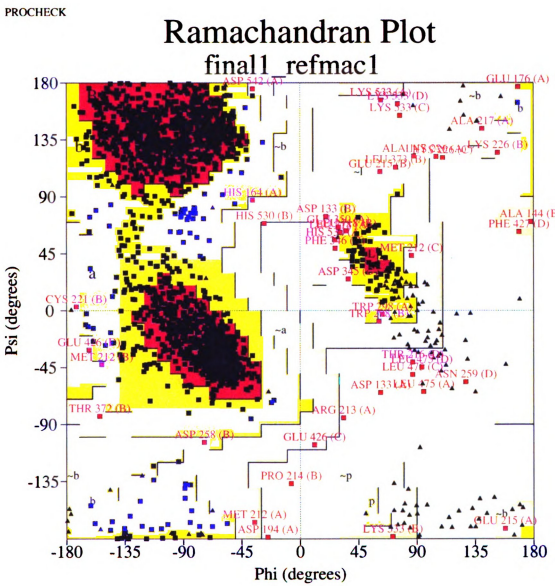
Appendix 5. Ramachandran plot of the BE/gamma CD structure



Appendix 6. Ramachandran plot of the BE/M7 structure



Appendix 7. Ramachandran plot of the BE/M6 structure



Appendix 8. Protein-ligand close contacts between BE and alpha CD. The interactions listed are within a 3.8 Å cutoff.

At binding site I

#	Chain	Residue	Atom	#	Chain	Ligand	Atom	Distance(Å)
255	D	ARG	NH1	1	G	ACX	O2A	3.41
255	D	ARG	NH1	1	G	ACX	O3B	3.78
259	D	ASN	OD1	1	G	ACX	O3D	3.67
259	D	ASN	OD1	1	G	ACX	O2C	3.47
259	D	ASN	C	1	G	ACX	O3C	3.74
259	D	ASN	O	1	G	ACX	C3C	3.60
259	D	ASN	O	1	G	ACX	O3C	2.83
259	D	ASN	O	1	G	ACX	O2B	3.13
260	D	ASN	CB	1	G	ACX	O3B	3.49
260	D	ASN	CB	1	G	ACX	O2B	3.28
260	D	ASN	CG	1	G	ACX	O3B	3.31
260	D	ASN	ND2	1	G	ACX	C3B	3.65
260	D	ASN	ND2	1	G	ACX	O3B	2.53
260	D	ASN	ND2	1	G	ACX	C2B	3.79
260	D	ASN	ND2	1	G	ACX	O2B	3.55
260	D	ASN	O	1	G	ACX	O2A	3.33
260	D	ASN	O	1	G	ACX	O3B	3.28
261	D	PHE	CA	1	G	ACX	O3A	3.30
261	D	PHE	CA	1	G	ACX	O2A	3.77
261	D	PHE	CB	1	G	ACX	O3A	3.43
261	D	PHE	CB	1	G	ACX	O2F	3.72
261	D	PHE	CG	1	G	ACX	C3A	3.75
261	D	PHE	CG	1	G	ACX	O3A	3.72
261	D	PHE	CE2	1	G	ACX	O1B	3.71
261	D	PHE	CE2	1	G	ACX	O1A	3.53
261	D	PHE	CD2	1	G	ACX	C3A	3.25
261	D	PHE	CD2	1	G	ACX	O3A	3.68
261	D	PHE	CD2	1	G	ACX	O1A	3.70
261	D	PHE	C	1	G	ACX	O3A	3.53
262	D	TRP	N	1	G	ACX	O3A	2.81
262	D	TRP	CA	1	G	ACX	O3A	3.74
262	D	TRP	CB	1	G	ACX	O3A	3.50
262	D	TRP	CG	1	G	ACX	O3A	3.61
262	D	TRP	CD2	1	G	ACX	O3A	3.53
262	D	TRP	CE3	1	G	ACX	O3A	3.41
262	D	TRP	O	1	G	ACX	O2F	3.76

At binding site II

#	Chain	Residue	Atom	#	Chain	Ligand	Atom	Distance(Å)
505	D	ASP	CA	1	J	ACX	O6D	3.47
505	D	ASP	CB	1	J	ACX	O6D	2.96
505	D	ASP	C	1	J	ACX	O6D	3.72
505	D	ASP	O	1	J	ACX	O6D	3.15
508	D	THR	CB	1	J	ACX	O6E	3.75
508	D	THR	CB	1	J	ACX	O5D	3.35
508	D	THR	CB	1	J	ACX	O6D	3.72
508	D	THR	OG1	1	J	ACX	O5D	3.18
508	D	THR	OG1	1	J	ACX	O6D	2.79
508	D	THR	C	1	J	ACX	O6E	3.61
508	D	THR	C	1	J	ACX	O6D	3.68
508	D	THR	O	1	J	ACX	C6E	3.06
508	D	THR	O	1	J	ACX	O6E	2.71
509	D	PHE	N	1	J	ACX	O6D	3.21
509	D	PHE	CA	1	J	ACX	C6D	3.59
509	D	PHE	CA	1	J	ACX	O6D	3.42
509	D	PHE	CB	1	J	ACX	O6D	3.76
509	D	PHE	CE2	1	J	ACX	O6C	3.37
509	D	PHE	CD2	1	J	ACX	C6D	3.71
511	D	ILE	CB	1	J	ACX	O6E	3.70
511	D	ILE	CG2	1	J	ACX	O6F	3.55
512	D	LEU	CD1	1	J	ACX	C6C	3.73
512	D	LEU	CD1	1	J	ACX	O6B	3.71
512	D	LEU	CD2	1	J	ACX	O6F	3.72
512	D	LEU	CD2	1	J	ACX	C5E	3.40
512	D	LEU	CD2	1	J	ACX	C6E	3.26
512	D	LEU	O	1	J	ACX	O6A	3.62
628	D	TRP	CD1	1	J	ACX	C1E	3.74
628	D	TRP	CD1	1	J	ACX	O5E	3.48
628	D	TRP	CD1	1	J	ACX	O6E	3.79
628	D	TRP	CD1	1	J	ACX	C2E	3.57
628	D	TRP	NE1	1	J	ACX	O5E	3.32
628	D	TRP	NE1	1	J	ACX	C6E	3.74
628	D	TRP	NE1	1	J	ACX	O6E	2.66
628	D	TRP	CE2	1	J	ACX	O6E	3.32
628	D	TRP	CZ2	1	J	ACX	O6E	3.43

At binding site III

#	Chain	Residue	Atom	#	Chain	Ligand	Atom	Distance(Å)
155	D	GLN	NE2	1	E	ACX	O2B	3.80
189	D	LYS	CD	1	E	ACX	O6B	3.58
189	D	LYS	NZ	1	E	ACX	C6A	3.73
189	D	LYS	NZ	1	E	ACX	O6B	3.55
201	D	LEU	CD1	1	E	ACX	C6A	3.67
201	D	LEU	CD1	1	E	ACX	O6A	3.25
211	D	GLN	CD	1	E	ACX	C1B	3.74
211	D	GLN	CD	1	E	ACX	O5B	3.23
211	D	GLN	OE1	1	E	ACX	C1B	3.69
211	D	GLN	OE1	1	E	ACX	O5B	2.74
211	D	GLN	OE1	1	E	ACX	C5B	3.63
211	D	GLN	OE1	1	E	ACX	C6B	3.44
211	D	GLN	NE2	1	E	ACX	C5C	3.04
211	D	GLN	NE2	1	E	ACX	C6C	3.23
211	D	GLN	NE2	1	E	ACX	O6C	3.70
211	D	GLN	NE2	1	E	ACX	C4C	3.31
211	D	GLN	NE2	1	E	ACX	O1B	3.28
211	D	GLN	NE2	1	E	ACX	C1B	2.98
211	D	GLN	NE2	1	E	ACX	O5B	3.00
212	D	MET	CB	1	E	ACX	O6C	3.72
215	D	GLU	CB	1	E	ACX	O5C	3.53
215	D	GLU	CB	1	E	ACX	C5C	3.66
215	D	GLU	CB	1	E	ACX	O1B	3.52
215	D	GLU	CG	1	E	ACX	O1C	3.56
215	D	GLU	CG	1	E	ACX	O5C	3.40
215	D	GLU	OE1	1	E	ACX	O1B	3.74
215	D	GLU	O	1	E	ACX	C5B	3.70
215	D	GLU	O	1	E	ACX	C6B	3.09
215	D	GLU	O	1	E	ACX	O6B	2.92
217	D	ALA	CB	1	E	ACX	C6B	3.49

At binding site IV

#	Chain	Residue	Atom	#	Chain	Ligand	Atom	Distance(Å)
542	C	ASP	CG	1	F	ACX	O2E	3.54
542	C	ASP	OD1	1	F	ACX	O3F	3.49
542	C	ASP	OD1	1	F	ACX	C2E	3.32
542	C	ASP	OD1	1	F	ACX	O2E	2.44
544	C	TRP	CG	1	F	ACX	O5E	3.57
544	C	TRP	NE1	1	F	ACX	C6E	3.75
544	C	TRP	CE2	1	F	ACX	C6E	3.62
544	C	TRP	CD2	1	F	ACX	O5E	3.69
545	C	GLN	CG	1	F	ACX	O3E	3.71
545	C	GLN	CD	1	F	ACX	O3E	3.54
545	C	GLN	OE1	1	F	ACX	O3E	3.38
545	C	GLN	NE2	1	F	ACX	O2E	3.63
659	C	PRO	CB	1	F	ACX	O2D	3.63
659	C	PRO	C	1	F	ACX	O2D	3.58
659	C	PRO	O	1	F	ACX	O3D	3.49
659	C	PRO	O	1	F	ACX	C2D	3.12
659	C	PRO	O	1	F	ACX	O2D	2.62
661	C	PRO	CG	1	F	ACX	O3D	3.51
661	C	PRO	CD	1	F	ACX	O3D	3.22
689	C	SER	CB	1	F	ACX	O5D	3.72
689	C	SER	CB	1	F	ACX	C6D	3.78
689	C	SER	CB	1	F	ACX	O6D	3.24
689	C	SER	OG	1	F	ACX	O6D	2.93
717	C	PRO	CB	1	F	ACX	C2D	3.79
717	C	PRO	CG	1	F	ACX	C2D	3.69

Appendix 9. Protein-ligand close contacts between BE and beta CD. The interactions listed are within a 3.8 Å cutoff.

At binding site I

#	Chain	Residue	Atom	#	Chain	Ligand	Atom	Distance(Å)
258	A	ASP	C	1	I	BCD	O33	3.77
258	A	ASP	O	1	I	BCD	O24	3.30
258	A	ASP	O	1	I	BCD	C33	3.29
258	A	ASP	O	1	I	BCD	O33	2.61
258	A	ASP	O	1	I	BCD	O23	3.44
259	A	ASN	OD1	1	I	BCD	C32	3.44
259	A	ASN	OD1	1	I	BCD	O32	3.06
259	A	ASN	OD1	1	I	BCD	O22	3.64
259	A	ASN	C	1	I	BCD	O34	3.79
259	A	ASN	O	1	I	BCD	C35	3.74
259	A	ASN	O	1	I	BCD	O44	3.62
259	A	ASN	O	1	I	BCD	C34	3.38
259	A	ASN	O	1	I	BCD	O34	3.29
260	A	ASN	CB	1	I	BCD	O25	3.42
260	A	ASN	OD1	1	I	BCD	O35	3.76
260	A	ASN	OD1	1	I	BCD	O25	3.40
260	A	ASN	C	1	I	BCD	O35	3.80
260	A	ASN	O	1	I	BCD	O26	3.69
260	A	ASN	O	1	I	BCD	O35	3.16
261	A	PHE	CA	1	I	BCD	C36	3.78
261	A	PHE	CA	1	I	BCD	O36	3.48
261	A	PHE	CA	1	I	BCD	O26	3.73
261	A	PHE	CB	1	I	BCD	O36	3.49
261	A	PHE	CE1	1	I	BCD	O47	3.73
261	A	PHE	CE2	1	I	BCD	O45	3.55
261	A	PHE	CD2	1	I	BCD	C36	3.40
261	A	PHE	CD2	1	I	BCD	O45	3.39
261	A	PHE	C	1	I	BCD	O36	3.66
262	A	TRP	N	1	I	BCD	O36	2.91
262	A	TRP	N	1	I	BCD	O26	3.54
262	A	TRP	CB	1	I	BCD	O36	3.58
262	A	TRP	CD2	1	I	BCD	O26	3.79
262	A	TRP	CE3	1	I	BCD	O36	3.43
262	A	TRP	CE3	1	I	BCD	C26	3.67
502	D	TYR	CA	1	I	BCD	O31	3.58

502	D	TYR	CB	1	I	BCD	O31	3.43
502	D	TYR	C	1	I	BCD	O31	3.48
502	D	TYR	O	1	I	BCD	C41	3.64
502	D	TYR	O	1	I	BCD	C31	3.50
502	D	TYR	O	1	I	BCD	O31	2.64
503	D	HIS	CE1	1	I	BCD	C22	3.74
503	D	HIS	CE1	1	I	BCD	C12	3.73

At binding site II

#	Chain	Residue	Atom	#	Chain	Ligand	Atom	Distance(Å)
505	B	ASP	O	1	F	BCD	O62	3.61
508	B	THR	CB	1	F	BCD	O62	3.65
508	B	THR	CB	1	F	BCD	O52	3.39
508	B	THR	CB	1	F	BCD	O61	3.69
508	B	THR	OG1	1	F	BCD	O62	2.61
508	B	THR	OG1	1	F	BCD	O52	3.00
508	B	THR	OG1	1	F	BCD	C12	3.75
508	B	THR	C	1	F	BCD	O62	3.52
508	B	THR	C	1	F	BCD	O61	3.66
508	B	THR	O	1	F	BCD	C61	3.35
508	B	THR	O	1	F	BCD	O61	2.83
509	B	PHE	N	1	F	BCD	O62	3.12
509	B	PHE	CA	1	F	BCD	C62	3.78
509	B	PHE	CA	1	F	BCD	O62	3.31
509	B	PHE	CB	1	F	BCD	O62	3.72
509	B	PHE	CE2	1	F	BCD	C63	3.72
511	B	ILE	CG1	1	F	BCD	O67	3.41
511	B	ILE	CG2	1	F	BCD	C61	2.89
511	B	ILE	CG2	1	F	BCD	O61	2.54
512	B	LEU	CD1	1	F	BCD	C52	3.67
512	B	LEU	CD1	1	F	BCD	C62	3.44
512	B	LEU	CD1	1	F	BCD	C51	3.67
512	B	LEU	CD1	1	F	BCD	C61	2.67
512	B	LEU	CD1	1	F	BCD	O61	3.76
512	B	LEU	CD2	1	F	BCD	C63	3.71
512	B	LEU	CD2	1	F	BCD	C62	3.61
628	B	TRP	CD1	1	F	BCD	O51	3.48
628	B	TRP	CD1	1	F	BCD	C21	3.72

628	B	TRP	NE1	1	F	BCD	O61	2.79
628	B	TRP	NE1	1	F	BCD	O51	3.36
628	B	TRP	CE2	1	F	BCD	O61	3.48
628	B	TRP	CZ2	1	F	BCD	O61	3.64
631	B	VAL	CG1	1	F	BCD	O31	3.77

At binding site IV

#	Chain	Residue	Atom	#	Chain	Ligand	Atom	Distance(Å)
542	C	ASP	CG	1	J	BCD	O27	3.31
542	C	ASP	OD1	1	J	BCD	C27	3.50
542	C	ASP	OD1	1	J	BCD	O27	2.62
542	C	ASP	OD1	1	J	BCD	C17	3.67
542	C	ASP	OD1	1	J	BCD	O36	3.15
542	C	ASP	OD2	1	J	BCD	C27	3.72
542	C	ASP	OD2	1	J	BCD	O27	3.27
544	C	TRP	CB	1	J	BCD	O57	3.73
544	C	TRP	CG	1	J	BCD	O57	3.73
544	C	TRP	CE3	1	J	BCD	O37	3.75
545	C	GLN	CG	1	J	BCD	O37	3.60
545	C	GLN	CD	1	J	BCD	O37	3.69
545	C	GLN	OE1	1	J	BCD	O37	2.98
545	C	GLN	OE1	1	J	BCD	O27	3.45
659	C	PRO	CB	1	J	BCD	O21	3.50
659	C	PRO	C	1	J	BCD	O21	3.67
659	C	PRO	O	1	J	BCD	O31	3.27
659	C	PRO	O	1	J	BCD	C21	3.18
659	C	PRO	O	1	J	BCD	O21	2.71
661	C	PRO	CG	1	J	BCD	O22	3.67
661	C	PRO	CG	1	J	BCD	C41	3.72
661	C	PRO	CG	1	J	BCD	O31	3.67
661	C	PRO	CD	1	J	BCD	O22	3.60
661	C	PRO	CD	1	J	BCD	O31	3.23
689	C	SER	CB	1	J	BCD	C61	3.64
689	C	SER	CB	1	J	BCD	O61	3.38
689	C	SER	CB	1	J	BCD	O51	3.77
689	C	SER	OG	1	J	BCD	C61	3.13
689	C	SER	OG	1	J	BCD	O61	2.96
717	C	PRO	CG	1	J	BCD	C61	3.59

Appendix 10. Protein-ligand close contacts between BE and gamma CD. The interactions listed are within a 3.8 Å cutoff.

At binding site II

#	Chain	Residue	Atom	#	Chain	Ligand	Atom	Distance(Å)
505	A	ASP	OD1B	1	E	RCD	C6D	3.75
505	A	ASP	OD2B	1	E	RCD	C1C	3.70
508	A	THR	CB	1	E	RCD	O5D	3.45
508	A	THR	OG1	1	E	RCD	C1D	3.79
508	A	THR	OG1	1	E	RCD	O5D	2.95
508	A	THR	OG1	1	E	RCD	C5D	3.55
508	A	THR	OG1	1	E	RCD	C6D	2.85
508	A	THR	OG1	1	E	RCD	O6D	3.73
508	A	THR	C	1	E	RCD	O6E	3.66
508	A	THR	C	1	E	RCD	O6D	3.75
508	A	THR	O	1	E	RCD	C6E	3.54
508	A	THR	O	1	E	RCD	O6E	2.72
509	A	PHE	N	1	E	RCD	C6D	3.72
509	A	PHE	N	1	E	RCD	O6D	3.23
509	A	PHE	CA	1	E	RCD	O6D	2.98
509	A	PHE	CB	1	E	RCD	O6D	3.26
509	A	PHE	CG	1	E	RCD	O6D	3.79
509	A	PHE	CD2	1	E	RCD	O6D	3.41
511	A	ILE	CB	1	E	RCD	O6E	3.65
511	A	ILE	CG2	1	E	RCD	O6F	3.71
512	A	LEU	CD1	1	E	RCD	C6E	3.72
512	A	LEU	CD2	1	E	RCD	C6E	2.81
512	A	LEU	CD2	1	E	RCD	O6E	3.62
512	A	LEU	CD2	1	E	RCD	O6D	3.76
628	A	TRP	CD1	1	E	RCD	O5E	3.49
628	A	TRP	CD1	1	E	RCD	C2E	3.78
628	A	TRP	NE1	1	E	RCD	O5E	3.42
628	A	TRP	NE1	1	E	RCD	O6E	3.08

At binding site III

#	Chain	Residue	Atom	#	Chain	Ligand	Atom	Distance(Å)
159	D	TRP	CG	1	H	RCD	C6H	3.50
159	D	TRP	CD1	1	H	RCD	C6H	3.66
159	D	TRP	CD2	1	H	RCD	C6H	3.65
159	D	TRP	CZ2	1	H	RCD	C1G	3.68
189	D	LYS	CD	1	H	RCD	O2H	3.58
189	D	LYS	CE	1	H	RCD	O3H	3.21
189	D	LYS	CE	1	H	RCD	O2H	3.35
189	D	LYS	NZ	1	H	RCD	O3H	3.28
189	D	LYS	NZ	1	H	RCD	C2H	3.72
189	D	LYS	NZ	1	H	RCD	O2H	2.67
201	D	LEU	CD1	1	H	RCD	C4A	3.79
201	D	LEU	CD1	1	H	RCD	C6A	3.59
201	D	LEU	CD1	1	H	RCD	O2H	3.80
211	D	GLN	CD	1	H	RCD	O3G	3.67
211	D	GLN	CD	1	H	RCD	O2G	3.32
211	D	GLN	OE1	1	H	RCD	O3G	2.79
211	D	GLN	OE1	1	H	RCD	C2G	3.74
211	D	GLN	OE1	1	H	RCD	O2G	3.48
211	D	GLN	NE2	1	H	RCD	C2G	3.35
211	D	GLN	NE2	1	H	RCD	O2G	2.38
215	D	GLU	CB	1	H	RCD	O3G	3.62
215	D	GLU	CG	1	H	RCD	O3H	3.64
215	D	GLU	CG	1	H	RCD	C3G	3.58
215	D	GLU	CG	1	H	RCD	O3G	3.50
215	D	GLU	CG	1	H	RCD	C2G	3.61
215	D	GLU	CG	1	H	RCD	O2G	2.61
215	D	GLU	CD	1	H	RCD	C3H	3.73
215	D	GLU	CD	1	H	RCD	O3H	3.23
215	D	GLU	CD	1	H	RCD	O2G	3.43
215	D	GLU	OE1	1	H	RCD	C3H	3.55
215	D	GLU	OE1	1	H	RCD	O3H	2.83
217	D	ALA	CB	1	H	RCD	O2G	3.73

At binding site IV

#	*Chain*	*Residue*	*Atom*	#	*Chain*	*Ligand*	*Atom*	Distance(Å)
542	C	ASP	CG	1	I	RCD	O2F	3.51
542	C	ASP	OD1	1	I	RCD	O2F	2.80
542	C	ASP	OD2	1	I	RCD	O2F	3.48
544	C	TRP	CG	1	I	RCD	O5F	3.59
544	C	TRP	CE3	1	I	RCD	C4F	3.79
545	C	GLN	OE1	1	I	RCD	O3F	3.44
659	C	PRO	CB	1	I	RCD	O2E	3.62
659	C	PRO	C	1	I	RCD	O2E	3.61
659	C	PRO	O	1	I	RCD	O3E	3.42
659	C	PRO	O	1	I	RCD	C2E	2.99
659	C	PRO	O	1	I	RCD	O2E	2.59
661	C	PRO	CG	1	I	RCD	O3E	3.77
661	C	PRO	CD	1	I	RCD	O3E	3.39
661	C	PRO	CD	1	I	RCD	O2D	3.77
689	C	SER	CB	1	I	RCD	C6E	3.73
689	C	SER	CB	1	I	RCD	O6E	3.48
689	C	SER	OG	1	I	RCD	C6E	3.56
689	C	SER	OG	1	I	RCD	O6E	2.77
717	C	PRO	CG	1	I	RCD	O5E	3.58

Appendix 11. Protein-ligand close contacts between BE and maltoheptaose (M7). The interactions listed are within a 3.8 Å cutoff.

At binding site III

#	Chain	Residue	Atom	#	Chain	Ligand	Atom	Distance(Å)
159	D	TRP	NE1	1	K	MAL	O62	3.52
159	D	TRP	NE1	1	K	MAL	O52	3.71
159	D	TRP	CE2	1	K	MAL	O52	3.71
159	D	TRP	CZ2	1	K	MAL	O52	3.45
159	D	TRP	CZ2	1	K	MAL	C12	3.56
189	D	LYS	CD	1	K	MAL	O21	3.68
189	D	LYS	CE	1	K	MAL	O21	3.68
189	D	LYS	NZ	1	K	MAL	O31	3.21
189	D	LYS	NZ	1	K	MAL	C21	3.61
189	D	LYS	NZ	1	K	MAL	O21	2.60
211	D	GLN	CD	1	K	MAL	O32	3.79
211	D	GLN	CD	1	K	MAL	C22	3.63
211	D	GLN	OE1	1	K	MAL	O32	3.72
211	D	GLN	OE1	1	K	MAL	C22	3.02
211	D	GLN	OE1	1	K	MAL	O22	2.86
211	D	GLN	NE2	1	K	MAL	C42	3.78
211	D	GLN	NE2	1	K	MAL	C32	3.58
211	D	GLN	NE2	1	K	MAL	O32	3.06
211	D	GLN	NE2	1	K	MAL	C22	3.48
215	D	GLU	CB	1	K	MAL	O32	3.22
215	D	GLU	CB	1	K	MAL	O22	3.67
215	D	GLU	CG	1	K	MAL	O32	3.30
215	D	GLU	CD	1	K	MAL	O32	3.11
215	D	GLU	OE1	1	K	MAL	C32	3.52
215	D	GLU	OE1	1	K	MAL	O32	2.24
215	D	GLU	O	1	K	MAL	O22	3.57
216	D	THR	O	1	K	MAL	O22	3.64

At binding site IV

#	Chain	Residue	Atom	#	Chain	Ligand	Atom	Distance(Å)
542	C	ASP	CG	1	L	MAL	O21	3.56
542	C	ASP	OD1	1	L	MAL	O21	3.44
542	C	ASP	OD2	1	L	MAL	C21	3.73
542	C	ASP	OD2	1	L	MAL	O21	3.03
544	C	TRP	CB	1	L	MAL	O51	3.78
544	C	TRP	CG	1	L	MAL	O51	3.57
545	C	GLN	CG	1	L	MAL	O31	3.66
659	C	PRO	CB	1	L	MAL	O22	3.58
659	C	PRO	C	1	L	MAL	O22	3.75
659	C	PRO	O	1	L	MAL	O32	3.35
659	C	PRO	O	1	L	MAL	C22	3.10
659	C	PRO	O	1	L	MAL	O22	2.68
661	C	PRO	CG	1	L	MAL	O32	3.39
661	C	PRO	CD	1	L	MAL	O32	3.00
689	C	SER	CB	1	L	MAL	O62	3.63
717	C	PRO	CG	1	L	MAL	O52	3.67

At binding site V and VI

#	*Chain*	*Residue*	*Atom*	#	*Chain*	*Ligand*	*Atom*	Distance(Å)
674	A	LYS	NZ	1	G	M7	O37	3.17
254	B	ARG	CA	1	G	M7	O67	3.72
254	B	ARG	C	1	G	M7	O67	3.70
255	B	ARG	N	1	G	M7	O67	2.77
255	B	ARG	CA	1	G	M7	O67	3.56
255	B	ARG	CB	1	G	M7	O67	3.77
255	B	ARG	C	1	G	M7	O67	3.64
255	B	ARG	O	1	G	M7	O67	2.92
255	B	ARG	O	1	G	M7	C67	3.48
257	B	THR	OG1	1	G	M7	O47	3.59
260	B	ASN	OD1	1	G	M7	O66	3.57
537	B	ASP	CG	1	G	M7	O22	3.31
537	B	ASP	OD1	1	G	M7	C32	3.69
537	B	ASP	OD1	1	G	M7	O32	3.00
537	B	ASP	OD1	1	G	M7	O22	3.53
537	B	ASP	OD2	1	G	M7	C22	3.60
537	B	ASP	OD2	1	G	M7	O22	2.33
543	B	ALA	CB	1	G	M7	O51	3.68
546	B	LYS	CD	1	G	M7	O61	3.58
546	B	LYS	CE	1	G	M7	O61	3.67
546	B	LYS	NZ	1	G	M7	O22	2.99
546	B	LYS	NZ	1	G	M7	O61	2.79
576	B	ARG	CZ	1	G	M7	O25	3.13
576	B	ARG	NH1	1	G	M7	C25	3.73
576	B	ARG	NH1	1	G	M7	O25	2.39
576	B	ARG	NH1	1	G	M7	C34	3.19
576	B	ARG	NH1	1	G	M7	O34	2.87
576	B	ARG	NH2	1	G	M7	O26	3.05
576	B	ARG	NH2	1	G	M7	C35	3.57
576	B	ARG	NH2	1	G	M7	O25	3.08
583	B	SER	CB	1	G	M7	O67	3.28
583	B	SER	CB	1	G	M7	C66	3.78
583	B	SER	OG	1	G	M7	O67	2.38
583	B	SER	OG	1	G	M7	C67	3.25
583	B	SER	OG	1	G	M7	O57	3.44
584	B	LEU	O	1	G	M7	O36	3.62
585	B	ASP	CA	1	G	M7	O36	3.29
585	B	ASP	CG	1	G	M7	O26	3.39

585	B	ASP	OD2	1	G	M7	C26	3.44
585	B	ASP	OD2	1	G	M7	O26	2.49
585	B	ASP	C	1	G	M7	O36	3.39
586	B	TRP	N	1	G	M7	O36	3.28
586	B	TRP	CB	1	G	M7	O57	3.59
586	B	TRP	CB	1	G	M7	C17	3.72
586	B	TRP	CD1	1	G	M7	O57	3.52
587	B	HIS	N	1	G	M7	C27	3.77
587	B	HIS	N	1	G	M7	O27	3.46
587	B	HIS	N	1	G	M7	C17	3.61
587	B	HIS	N	1	G	M7	O36	3.22
587	B	HIS	CA	1	G	M7	O27	3.39
587	B	HIS	CB	1	G	M7	O27	3.47
587	B	HIS	CB	1	G	M7	O36	3.45
587	B	HIS	CG	1	G	M7	O36	3.23
587	B	HIS	ND1	1	G	M7	C36	3.13
587	B	HIS	ND1	1	G	M7	O36	2.30
587	B	HIS	ND1	1	G	M7	C26	3.77
587	B	HIS	ND1	1	G	M7	O26	3.20
587	B	HIS	ND1	1	G	M7	C55	3.70
587	B	HIS	CE1	1	G	M7	O36	3.31
587	B	HIS	CE1	1	G	M7	O26	3.16
587	B	HIS	CE1	1	G	M7	C55	3.65
587	B	HIS	CE1	1	G	M7	O55	3.77
590	B	GLU	OE1	1	G	M7	O27	3.00
595	B	TRP	CD1	1	G	M7	O11	3.64
595	B	TRP	NE1	1	G	M7	O11	3.32
596	B	HIS	CE1	1	G	M7	O31	3.62
596	B	HIS	NE2	1	G	M7	O31	3.10

At binding site VII

#	*Chain*	*Residue*	*Atom*	*#*	*Chain*	*Ligand*	*Atom*	Distance(Å)
467	A	SER	CA	1	I	MAL	O61	3.61
467	A	SER	C	1	I	MAL	C61	3.40
467	A	SER	C	1	I	MAL	O61	2.95
467	A	SER	O	1	I	MAL	C61	3.15
467	A	SER	O	1	I	MAL	O61	3.26

468	A	ARG	N	1	I	MAL	C61	3.55
468	A	ARG	N	1	I	MAL	O61	2.90
468	A	ARG	CA	1	I	MAL	C61	3.51
468	A	ARG	CA	1	I	MAL	O61	3.25
468	A	ARG	C	1	I	MAL	C61	3.43
468	A	ARG	C	1	I	MAL	O61	3.02
468	A	ARG	O	1	I	MAL	O61	3.09
469	A	PRO	N	1	I	MAL	C61	3.51
469	A	PRO	N	1	I	MAL	O61	3.54
469	A	PRO	CD	1	I	MAL	C61	3.77
470	A	GLN	CD	1	I	MAL	O11	3.67
470	A	GLN	OE1	1	I	MAL	O11	3.09
470	A	GLN	NE2	1	I	MAL	O11	3.77
476	A	GLY	CA	1	I	MAL	O61	3.18
476	A	GLY	C	1	I	MAL	O61	3.13
476	A	GLY	C	1	I	MAL	O51	3.41
476	A	GLY	O	1	I	MAL	O51	3.23
476	A	GLY	O	1	I	MAL	C11	3.41
477	A	PHE	N	1	I	MAL	O61	3.12
477	A	PHE	C	1	I	MAL	O61	3.27
477	A	PHE	C	1	I	MAL	O51	3.69
477	A	PHE	O	1	I	MAL	C51	3.77
477	A	PHE	O	1	I	MAL	C61	2.93
477	A	PHE	O	1	I	MAL	O61	2.33
477	A	PHE	O	1	I	MAL	O51	3.36
478	A	TRP	CE2	1	I	MAL	C21	3.58
478	A	TRP	CZ2	1	I	MAL	C21	3.51
478	A	TRP	CZ2	1	I	MAL	O21	3.60
518	A	ASN	OD1	1	I	MAL	O62	3.07
518	A	ASN	OD1	1	I	MAL	C62	3.50
518	A	ASN	ND2	1	I	MAL	C62	3.33

Appendix 12. Protein-ligand close contacts between BE and maltohexaose (M6). The interactions listed are within a 3.8 Å cutoff.

At binding site III

#	Chain	Residue	Atom	#	Chain	Ligand	Atom	Distance(Å)
155	D	GLN	CG	1	F	MAL	O62	3.61
155	D	GLN	NE2	1	F	MAL	O62	3.46
155	D	GLN	NE2	1	F	MAL	C62	3.30
159	D	TRP	CG	1	F	MAL	O61	3.72
159	D	TRP	CD1	1	F	MAL	O61	3.36
159	D	TRP	NE1	1	F	MAL	O62	3.55
159	D	TRP	NE1	1	F	MAL	O52	3.50
159	D	TRP	NE1	1	F	MAL	O61	3.57
159	D	TRP	CZ2	1	F	MAL	C12	3.63
189	D	LYS	CD	1	F	MAL	O21	3.47
189	D	LYS	CE	1	F	MAL	O31	3.66
189	D	LYS	CE	1	F	MAL	O21	3.17
189	D	LYS	NZ	1	F	MAL	O21	2.92
201	D	LEU	CD1	1	F	MAL	O21	3.39
201	D	LEU	CD1	1	F	MAL	C11	3.51
211	D	GLN	CD	1	F	MAL	O32	3.72
211	D	GLN	CD	1	F	MAL	O22	3.65
211	D	GLN	OE1	1	F	MAL	C42	3.60
211	D	GLN	OE1	1	F	MAL	C32	3.28
211	D	GLN	OE1	1	F	MAL	O32	2.63
211	D	GLN	OE1	1	F	MAL	C22	3.28
211	D	GLN	OE1	1	F	MAL	O22	3.34
211	D	GLN	NE2	1	F	MAL	C22	3.79
211	D	GLN	NE2	1	F	MAL	O22	3.14
215	D	GLU	CG	1	F	MAL	O22	3.26
215	D	GLU	CG	1	F	MAL	O31	2.93

At binding site IV

#	Chain	Residue	Atom	#	Chain	Ligand	Atom	Distance(Å)
542	C	ASP	CG	1	E	MAL	O21	3.42
542	C	ASP	OD1	1	E	MAL	C21	3.28
542	C	ASP	OD1	1	E	MAL	O21	2.56
542	C	ASP	OD1	1	E	MAL	C11	3.52
542	C	ASP	OD2	1	E	MAL	O21	3.66
544	C	TRP	CG	1	E	MAL	O51	3.65
544	C	TRP	NE1	1	E	MAL	C61	3.79
544	C	TRP	CE2	1	E	MAL	C61	3.72
545	C	GLN	CG	1	E	MAL	O31	3.75
545	C	GLN	CD	1	E	MAL	O31	3.59
545	C	GLN	NE2	1	E	MAL	O31	3.60
545	C	GLN	NE2	1	E	MAL	O21	3.12
659	C	PRO	O	1	E	MAL	C22	3.71
659	C	PRO	O	1	E	MAL	O22	2.66
661	C	PRO	CG	1	E	MAL	O32	2.97
661	C	PRO	CD	1	E	MAL	O32	3.03
689	C	SER	CB	1	E	MAL	O62	3.41
689	C	SER	CB	1	E	MAL	C62	3.52
689	C	SER	CB	1	E	MAL	O52	3.77
689	C	SER	OG	1	E	MAL	O62	3.09
689	C	SER	OG	1	E	MAL	C62	3.73

At binding site V

#	Chain	Residue	Atom	#	Chain	Ligand	Atom	Distance(Å)
674	C	LYS	NZ	1	G	MAL	O32	2.78
254	D	ARG	CA	1	G	MAL	O62	3.21
254	D	ARG	CB	1	G	MAL	O62	3.61
254	D	ARG	CB	1	G	MAL	C62	3.76
254	D	ARG	CG	1	G	MAL	C62	3.76
254	D	ARG	C	1	G	MAL	O62	3.30
255	D	ARG	N	1	G	MAL	O62	2.57
255	D	ARG	N	1	G	MAL	C62	3.64

255	D	ARG	CA	1	G	MAL	O62	3.61
255	D	ARG	C	1	G	MAL	O62	3.76
255	D	ARG	O	1	G	MAL	O62	3.06
255	D	ARG	O	1	G	MAL	C62	3.38
257	D	THR	CG2	1	G	MAL	C32	3.80
257	D	THR	CG2	1	G	MAL	C61	3.76
260	D	ASN	OD1	1	G	MAL	O61	3.60
576	D	ARG	NH2	1	G	MAL	O21	3.59
583	D	SER	CB	1	G	MAL	O62	3.17
583	D	SER	CB	1	G	MAL	O51	3.50
583	D	SER	OG	1	G	MAL	O62	2.54
583	D	SER	OG	1	G	MAL	C62	3.57
583	D	SER	OG	1	G	MAL	O52	3.58
584	D	LEU	O	1	G	MAL	O31	3.48
584	D	LEU	O	1	G	MAL	C21	3.38
585	D	ASP	CA	1	G	MAL	O31	3.49
585	D	ASP	OD2	1	G	MAL	C21	3.62
585	D	ASP	OD2	1	G	MAL	O21	2.91
585	D	ASP	C	1	G	MAL	O31	3.50
586	D	TRP	N	1	G	MAL	O31	3.21
586	D	TRP	CB	1	G	MAL	O52	3.68
586	D	TRP	CD1	1	G	MAL	O52	3.73
587	D	HIS	N	1	G	MAL	O31	3.06
587	D	HIS	CB	1	G	MAL	O31	3.61
587	D	HIS	CG	1	G	MAL	O31	3.62
587	D	HIS	ND1	1	G	MAL	C31	3.53
587	D	HIS	ND1	1	G	MAL	O31	2.82
587	D	HIS	ND1	1	G	MAL	O21	3.29
587	D	HIS	CE1	1	G	MAL	O21	3.44
590	D	GLU	CD	1	G	MAL	O22	3.65
590	D	GLU	OE1	1	G	MAL	O32	3.76
590	D	GLU	OE1	1	G	MAL	C22	3.22
590	D	GLU	OE1	1	G	MAL	O22	2.64