THESIS
2
2002

This is to certify that the

dissertation entitled

The X-ray crystallographic structures of branching enzyme and angiostatin

presented by

Marta Cristina Abad Rivera

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Chemistry

Major professor

Date 1/14/02

MSU is an Affirmative Action/Equal Opportunity Institution

O-12771

LIBRARY
Michigan State
University

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

6/01 c:/CIRC/DateDue.p65-p.15

THE X-RAY CRYSTALLOGRAPHIC STRUCTURES OF BRANCHING ENZYME AND ANGIOSTATIN

By

Marta Cristina Abad Rivera

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Chemistry

2002

ABSTRACT

THE X-RAY CRYSTALLOGRAPHIC STRUCTURES OF BRANCHING ENZYME AND ANGIOSTATIN

By

Marta Cristina Abad Rivera

X-ray crystallographic studies have been performed for the structure determination of two proteins, branching enzyme and angiostatin. In these studies isomorphous replacement, anomalous dispersion and molecular replacement methods were used to calculate the electron density maps of the aforementioned proteins.

Branching enzyme is one of three enzymes involved in the biosynthesis of starch in plants and glycogen in animals and bacteria. It has an important role in the determination of the final structures of starch and glycogen. This enzyme catalyzes the cleavage of α-1,4 glucosidic bonds and subsequently transfers this chain into the α-1,6 position. The conversion of this linear polysaccharide into a branched network not only makes starch and glycogen more reactive to both synthesis and digestion, it also assures its solubility in the cell. Insoluble glycogen caused by mutations in the branching enzyme gene (Glycogen Storage Disease type IV) is a lethal genetic disease for which no clinical treatment is known.

Escherichia coli branching enzyme was crystallized and high-resolution data to 2.3 Å resolution was collected. Phasing information was obtained using isomorphous replacement and anomalous dispersion methods. This, in addition to four fold averaging, led to the calculation of an electron density map. The structure shows that branching

enzyme presents the central (α/β) barrel catalytic domain that is conserved among members of the α-amylase family of enzymes, to which branching enzyme belongs. In addition, a mechanism for branching enzyme has been proposed based on sugar substrate modeling and comparison of the branching enzyme structure with other members of the α-amylase family of enzymes.

Angiostatin is a protein that inhibits angiogenesis, a process in which new blood vessels form from pre-existing ones. Three decades ago Dr. Judah Folkman proposed that tumor growth and metastasis dissemination are angiogenesis-dependent processes, and the idea of angiogenesis inhibitors for cancer treatment was born. Angiostatin was catapulted to the forefront of anticancer treatment drugs when it was first discovered in the mid 1990s. The structure of human angiostatin was determined by molecular replacement at 1.75 Å resolution. The structure revealed that all three kringle lysine-binding sites contain a bound bicine molecule, while those of kringle 2 and kringle 3 are cofacial. Moreover, the separation of the kringle 2 and kringle 3 lysine binding sites is sufficient to accommodate the α-helix of the 30 residue peptide VEK-30 found in the kringle 2/VEK-30 complex. Together the three kringles produce a central cavity suggestive of a unique domain where they may function in concert.

To my Mother and Father

AKNOWLEDGMENTS

I thank my advisor Jim Geiger for his guidance, help, and support. I couldn't have asked for a better advisor. I was also honored to have the opportunity to work with Dr. Tulinsky, a renowned scientist, and pioneer in the field of protein crystallography. Dr. Tulinsky thank you for letting me become part of your team and enlightening me with your interesting tales about crystallography in the old days. I will like to show my appreciation to Dr. Raghuvir Arni for his help in the refinement of the angiostatin structure. I was also very fortunate to work with Dr. Preiss, an authority in the field of starch and glycogen biosynthesis, and who introduced me to the marvelous field of the glycogen biosynthetic enzymes. I am in gratitude to the work of Dr. Kim Binderup in providing the purified branching enzyme protein used for crystallization. I will also like to offer my appreciation to my good friend and collaborator Dr. Jorge Ríos. Dr. Ríos collected the selenium methionine data set that was crucial for the determination of the structure of branching enzyme.

We all know that graduate school is not easy but having a lab full of nice people makes a huge difference. Thanks to a group of phenomenal lab mates Xiangshu, Stacy, Erika, Sara, Aimee, Adam, Mike, Tyra, Michelle and Elena thank you to all of you who always kept me laughing even while writing this dissertation. I will like to give special thanks to Stacy Hovde who took the time to revise every single page of this thesis and pretty much everything I wrote. Thank you Stacy for all your help, for being my spell checker and over all for a wonderful friendship.

I also thank my husband for his support and encouragement. Jose, we started this together and we are both finishing in triumph. I will also like to thank my parents and brothers: Francisco, Juan and Emilio for their love, encouragement and unconditional support.

I am in gratitude for the financial support provided by the Hispanic Scholarship Fund, MSU's Equal Opportunity Fellowship and especially to the Bill Gates Foundation who supported the last two years of my graduate school.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER IV: THE THREE DIMENSIONAL STRUCTURE OF ANGIOSTATIN

LIST OF FIGURES

Images in this dissertation are presented in color.

CHAPTER I: INTRODUCTION

CHAPTER II: X-RAY STRUCTURE DETERMINATION

CHAPTER IV: THE THREE DIMENSIONAL STRUCTURE OF ANGIOSTATIN

LIST OF ABBREVIATIONS

A - alanine
ADP - Adenosine diphosphate
ADPGlc Ppase - ADP glucose pyrophosphorylase
aFGF - acidic fibroblast growth factor
APS - Advanced Photon Source
ATP - Adenosine triphosphate

BE - branching enzyme
bFGF - basic fibroblast growth factor
bicine - N,N Bis(2hydroxyethyl) glycine

C - cysteine
CGT - cyclodextrin glucanotransferase
C terminal - carboxy terminal

D - aspartic acid
DEPC - diethylpyrocarbonate

E - glutamic acid
EACA - ε-aminocaproic acid
E. coli - *Escherichia coli*
ED_{50} - half life maximum concentration to achieve 50% inhibition of bFGF-stimulated bovine capillary endothelial cell growth.
EPS - electrostatic potential surface

F - phenylalanine
Fcalc - calculated structure factors
FGF - fibroblast growth factor
F_H - structure factor contribution of the derivative
F_{hkl} - structure factor for a reflection labeled ***hkl***
Fobs - observed structure factors
F_{PH} - structure factor of the protein plus derivative

G - glycine
G(S)S - glycogen (starch) synthase
GSDIV - glycogen storage disease type IV

H - histidine
hepes - N-[2-hydroxyethyl] piperazine-N'-[ethane sulfonic acid]

I - isoleucine
I - intensity
interkringle - inter-Kg

K - lysine
Kg - kringle

L - leucine
LBS - lysine binding site

M - methionine
m - mass in mg
mBEI - *maize* endosperm BE isoform I
mBEII - *maize* endosperm BE isoform II
MPD - 2-Methyl-2,4-pentanediol

N - asparagine
N113BE - BE from *E. coli* lacking the first 112 residues at the amino terminal
N terminal - amino terminal

P – proline
PAM - group A pathological *Streptococcal* surface protein
PEG - polyethylene glycol
PP - phasing power

Q - glutamine
R - arginine
rmsd - root mean square deviation

S - serine
SAD - single wavelength anomalous dispersion
SBC - Structural Biology Center
SeMet - selenium methionine
SIR - single isomorphous replacement

T - threonine
TSP1 - thrombospodin 1
TSP2 - thrombospodin 2

UDP - Uridine diphosphate
UDPGlc Ppase - UDP glucose pyrophosphorylase

V - valine
v - volume in ml
VEGF - vascular endothelial growth factor

W - tryptophan
WT - wild type

Y - tyrosine

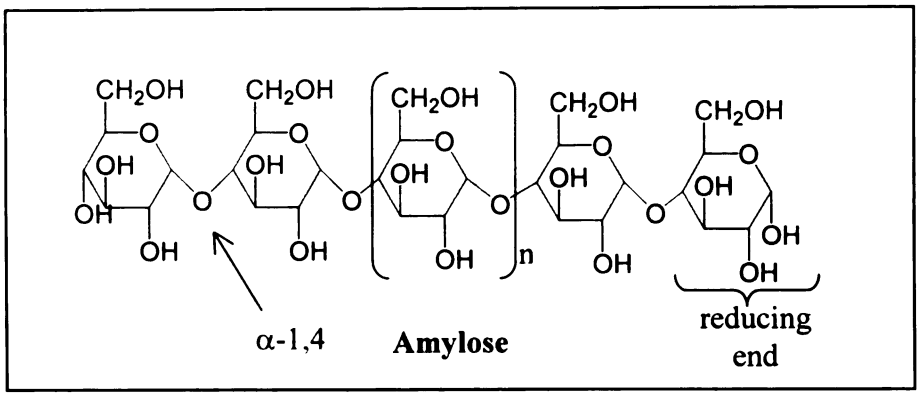
CHAPTER I: INTRODUCTION

1.1 Branching Enzyme

1.1.1 Glycogen

Glycogen, the energy storage polysaccharide in animals and bacteria cells, accumulates under environmental conditions that limit growth and offer excess in carbon supply (1-4). The major accumulation of glycogen occurs at the stationary phase of the growth cycle under nitrogen reduction conditions. Bacterial mutants with defective glycogen biosynthetic enzymes are viable under a glucose rich medium, showing that glycogen is not required for growth (4).

Starch is the form in which plants store the energy accumulated by carbon fixation via photosynthesis. Starch and glycogen are composed of polyglucose chains linked by α-1,4 and α-1-6 bonds (Figure 1.1). Amylose and amylopectin are the two molecules that form starch and glycogen (Figure 1.1). Amylose is the mainly linear polysaccharide formed by α-1,4 glucosidic bonds, although it contains some α-1,6 branches (less than 0.6% of all glucosidic linkages). Amylose chains vary in length from 840 to 22,000 glucose units (5). Amylopectin is a highly branched polymer with a molecular weight of up to one million. The α-1,6 branches occur every 8 to 21 residues in glycogen and every 24 to 30 residues in starch. Although there are many similarities between the metabolism of starch and bacterial glycogen there is a difference in its final structure. Glycogen is more branched with twice (10% of all glucosidic bonds) the amount of α-1,6 links compared to starch.



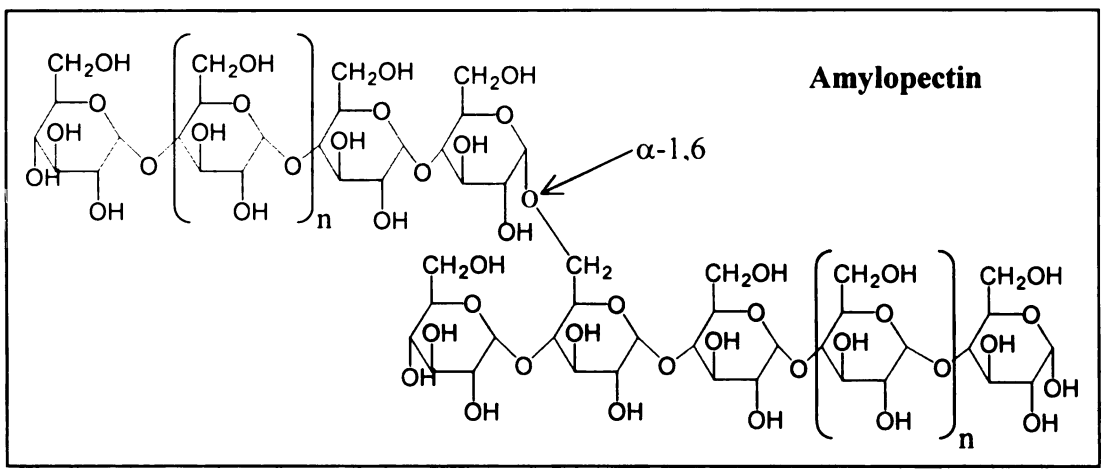


Figure 1.1 Starch and glycogen are formed by D-glucose units linked by α-1,4 and α-1,6 glucosidic bonds. Amylose and amylopectin are the two molecules that form starch and glycogen.

Energy storage in the form of a glucose polymer offers a compact and efficient way of storing the easily mobilized glucose molecule. Moreover, the formation of the α-1,6 branches increases the energetic capacity of glycogen. Glycogen and starch are synthesized and hydrolyzed by removing or adding the non-reducing glucose units (sugar without a free anomeric carbon) (Figure 1.1). With more branches, there are more non-reducing ends, this means that α-amylase and starch synthase will have more substrates to perform digestion or synthesis, increasing the molecule reactivity towards both synthesis and digestion.

1.1.2 Biosynthesis of Bacterial Glycogen

Glycogen biosynthesis occurs in bacteria during non-growing periods when carbon is in excess. During the non-growing phase energy is required for various essential processes like chemotactic response, maintenance of motility and intracellular pH, osmotic regulation, and translation among others processes. Glycogen plays an important role in preserving cell integrity in bacteria under harsh conditions (4). Studies performed on *Escherichia coli* (*E. coli*) and *Enterobacter aerogenes* show that mutants unable to synthesize glycogen degrade their RNA and protein in media that is devoid of a carbon source. When wild type bacteria are grown under the same conditions, they preserve their cellular constituents (4).

The biosynthesis of bacterial glycogen follows a pathway similar to starch synthesis in plants, although there is a higher similarity between the final structures of bacterial and mammalian glycogen. The starch and glycogen synthetic pathways consist of three steps, starting with the formation of the sugar nucleotide glucosyl donor,

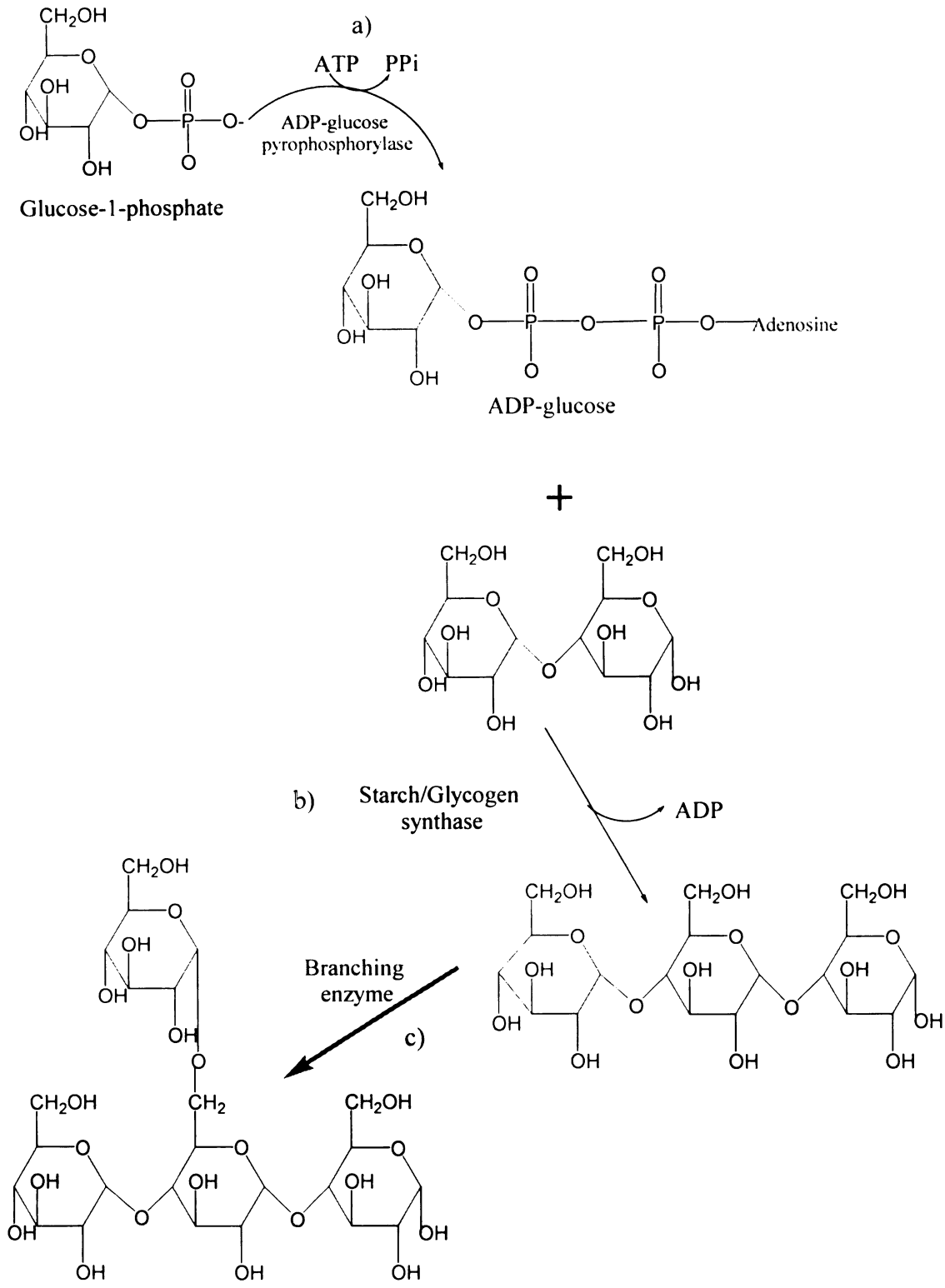


Figure 1.2 Biosynthetic pathway of starch and glycogen synthesis

followed by the elongation of the α-1,4 polyglucose chain, and finally the rearrangement of the polysaccharide (Figure 1.2).

The activation of a glucose 1-phosphate molecule into adenosine diphosphate (ADP)-glucose is catalyzed by the enzyme ADP glucose pyrophosphorylase (ADPGlc Ppase) (Figure 1a). In plants and bacteria cells, an ADP-glucose molecule is the glucose donor in the next step of the reaction. In mammalian glycogen synthesis, the glucose donor is UDP-glucose and its formation is catalyzed by UDPGlc Ppase. ADPGlc Ppase from plants consists of two isoenzymes forming a heterotetramer. For example, ADPGlc Ppase from potato tuber is formed by a small subunit involved in catalysis and a large subunit responsible for allosteric activation and inhibition (6). Bacterial and mammalian phosphorylases consist of a single unit forming a homotetramer and an homooctamer, respectively.

The elongation of the glucan chain is catalyzed by glycogen or starch synthase (G(S)S) as shown in Figure 1.2b. This enzyme forms an α-1,4 glucosidic bond between the anomeric carbon of the sugar nucleotide and a primer glucose molecule. In plant and animal cells the number of isoenzymes may vary depending on the species. In bacteria a single enzyme produces the elongation of the chain.

Tandecarz and Cardini were the first to propose that glycogen synthesis must be initiated by a self-glucosylated protein (7). This protein was later isolated and named glycogenin. Glycogenin first performs a self glucosylation that is followed by the elongation of the glucose chain to up to 8 glucose units (8). Glycogenin has been isolated from liver and muscle cells (8-10). There is no evidence of the existence of a self

glucosylated protein or if its even needed for the initiation of starch or bacterial glycogen synthesis.

In the rearrangement step, branching enzyme (BE) is responsible for the formation of the α-1,6 branch points. This is achieved by the cleavage of the α-1,4 bond and the subsequent transfer of a glucose chain to the α-1,6 position. This enzyme will be discussed in detail in the next section.

ADPGlc Ppase is the only allosterically regulated enzyme in both the bacterial glycogen and starch biosynthetic pathways. In *e. coli* it is activated by fructose 1,6 biphosphate, an indicator of carbon excess during glucogenesis, and inhibited by AMP, ADP, or inorganic phosphate, all indicators of low energy in the cell. The regulation of mammalian glycogen synthesis occurs at the GS step because GS catalyzes the first unique reaction in the pathway. UDP-glucose, the glucosyl donor on mammalian GS, is used as a precursor for synthesis of cellular constituents. Mammalian GS differs from the bacterial and plant enzymes not only in that it uses UDP-glucose as a substrate, but it also exhibits regulatory activity. Mammalian GS exists in a phosphorylated or dephosphorylated form that is either active or inactive, respectively. Because ADP-glucose formation is used solely for starch and bacterial glycogen synthesis, the regulation in the first step is energetically more efficient by conserving ATP.

1.1.3 Branching enzyme

Branching enzyme (1,4-α-glucan : 1,4-α-glucan 6-glucosyltransferase; EC 2.4.1.18) has an important role in the determination of the final structure of starch and glycogen. This enzyme catalyzes the formation of the α-1,6 branch points, transforming

a linear polysaccharide into a branched network. This is achieved by cleavage of the α-1,4-glucosidic linkage, yielding a non-reducing end polysaccharide chain, and subsequent attachment to the α-1,6 position.

The gene of the *E. coli* branching enzyme has been cloned and the nucleotide sequence determined (11). The gene consisted of 2,184 base pairs, coding 728 amino acids with a molecular weight of 84,231 Da. In bacteria there is only one BE while in plants there can be multiple isoforms. Several of these plant isoenzymes have been identified; four forms in rice seed, three in wheat endosperm and three in maize endosperm seeds (12-15).

The unique feature of branching enzyme's action lies in its specificity for the length of the glucan chain transferred. Glycogen branching enzyme from *E. coli* has a preference for transferring shorter chains between 5 and 16 glucose units from a non reducing end of at least 11 units. On the other hand, starch branching enzyme from maize transfers a wider range of chains from 6 to 30 glucose units (16). This specificity is consistent with the denser structure of glycogen due to double the number of α-1,6 links compared with starch.

The two isoenzymes from maize endosperm, mBEI and mBEII have been extensively studied (17,18). Although mBEI and mBEII are 58% identical, they differ in substrate preference, branching pattern and catalytic activity. The mBEI isoform has a preference for transferring longer chains of 11 glucose units or longer with substrate preference for amylose over amylopectin . On the other hand, mBEII transfers chains 6 glucose units or longer in length. Even though amylose is a favorable substrate, mBEII

has a higher affinity for amylopectin. Also, mBEI has a five to six fold higher V_{max} when compared to mBEII.

Branching enzyme is divided into three domains based on secondary structure prediction and its connection with the amylotic family of enzymes. These domains are; an amino (N) terminal domain, a carboxyl (C) terminal domain and a central (α/β) barrel catalytic domain. In an attempt to understand the role the three domains have in branching enzyme activity, several chimeric mutants were made (Figure 1.3) (18). From all eight mutants constructed, only three, F, G and H, showed some activity. Mutant F exhibited a higher activity than the wild type mBEI and mBEII. This mutant also has a higher specificity for amylose over amylopectin and a catalytic activity similar to mBEI. In contrast, its branch transfer pattern was similar to mBEII. Analysis showed that mutant G transferred more of the longer chains (11 glucose units or longer) similar to mBEI. From these studies it was clear that the C terminal is involved in substrate specificity and catalytic capacity, while the N terminal is involved in determining the size of the chain transferred. This study also established the importance of the amino and carboxyl termini, challenging the belief that the catalytic center of BE is limited to the (α/β) barrel domain.

As mentioned before, BE belongs to the α-amylase family of enzymes (19,20). Members of this group include α-amylases, pullulanase/isoamylase, cyclodextrin glucanotransferase (CGT) and branching enzymes. They have the common function of cleaving and/or transferring glucose chains (21). The α-amylases and pullulanase/isoamylase catalyze the hydrolysis of α-1,4 and α-1,6 glucosidic bonds, respectively. Branching enzyme and CGT are the two members that catalyze

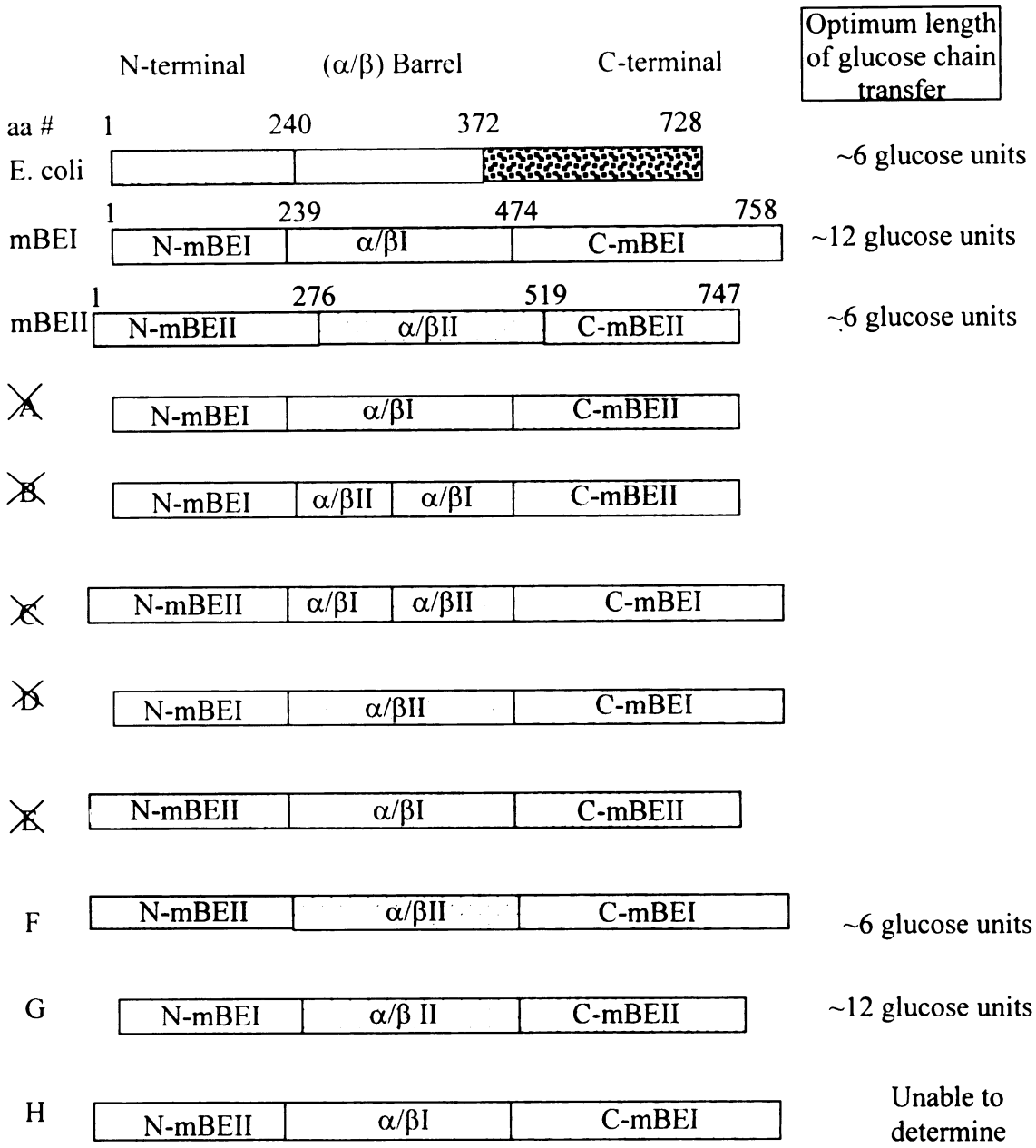


Figure 1.3 Chimeric enzymes constructed from mBEI and mBEII (18). The "X" represents the inactive mutants and aa# stands for amino acid number. N-mBEI, C-mBEI and α/βI represent the amino terminal, carboxylate terminal and α/β barrel of mBEI, respectively. The same nomenclature applies for mBEII.

transglycosylation reactions, with BE being the only one with specificity for two different glucosidic bonds. CGT catalyzes the formation of cyclodextrins by cleavage and subsequent transglycosylation of α-1,4 links. X-ray structures of α-amylases, isoamylase and CGTs show that these enzymes have a common (α/β) barrel domain that contain the enzyme's catalytic center (22-26). Based on biochemical data and X-ray structures of apo and substrate bound α-amylase and CGT, the catalytic center has been defined to be composed of seven residues; D335, H340, R403, D405, E458, H525 and D526 (*E. coli* branching enzyme numbering) (Figure 1.4)(27,28). These residues are conserved among members of this family and BE from various organisms (21,29). In an attempt to understand the catalytic relevance of these conserved residues, studies have been performed on *maize* endosperm BE by using site directed mutagenesis. When D405, E458 and D526 (386, 441 and 509 mBEII numbering) were mutated to their respective amide or alternate acid form BE activity disappeared (30).

Chemical modification experiments using the arginine specific reagent, phenylglyoxal, were performed to study the possibility of essential arginine residues present (31). Phenylglyoxal inactivated both mBEI and mBEII. In both cases the presence of amylose and amylopectin protected the enzyme from inactivation, although amylopectin was a better substrate. The sequence alignment of several branching enzymes were analyzed and two arginine residues, R312 (291 mBEII numbering) and R403 (384), conserved among BEs and located in the (α/β) barrel catalytic domain, were proposed to be the ones involved in catalysis. Site directed mutagenesis experiments performed on mBEII demonstrated that R312 is not necessary for BE activity (32). However, R403 mutations to either alanine, serine, glutamine or glutamic acid produced

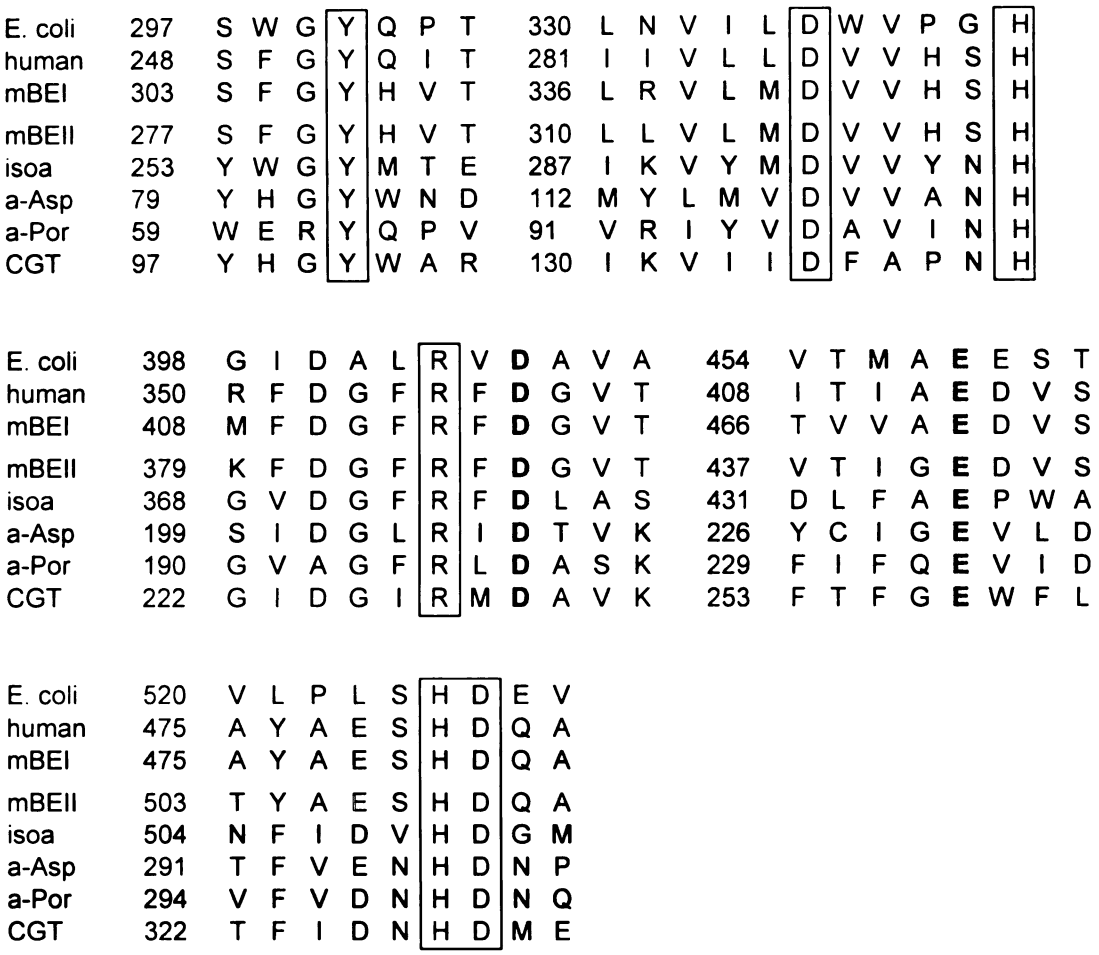


Figure 1.4 Conserved catalytic residues in the α-amylase family of enzymes. The residues enclosed by the boxes are involved in substrate binding and the shaded residues are involved in catalysis. The labels in the sequence alignment are *E. coli* for *E. coli* BE; human for *Homo Sapiens* BE; mBEI and mBEII maize endosperm for BE I and II, respectively; isoa is isoamylase from *Pseudomonas amyloderamosa,* α-Asp and α-Por are α-amylase from *Aspergillus Oryzae* and Porcine Pancreatic and CGT is cyclodextrin glucanotransferase from *Bascillus Circulans.*

an inactive enzyme. This demonstrates that R403 plays a direct role in BE catalysis (32). However, when R403 was replaced with lysine it retained approximately 5% of the wild type (WT) enzyme activity. The two enzymes' kinetic parameters, substrate specificity and size of chain transferred were compared. The R403K mutant and WT enzyme have similar substrate specificity preferring amylopectin as the substrate. The distribution of size of the chains transferred and the Km value was similar to that of WT.

Analysis of the two histidine residues, H340 (320) and H525 (508), conserved in the amylase family, was performed on mBE by site directed mutagenesis and chemical modification using diethylpyrocarbonate (DEPC) (33). DEPC reacts with histidine residues as well as with other amino acids like tyrosine, lysine and cysteine (34,35). Hydroxylamine removes the DEPC from modified hystidyl and tyrosyl residues, but not from lysyl or cysteine residues. In these studies BE was inactivated after adding DEPC and reactivated when hydroxylamine was added to the reaction mixture, ruling out lysines or cysteines as the residues responsible for the loss of activity. The presence of DEPC modified histidine increases the UV absorbance at 240 nm compared to the umodified enzyme, while the absorbance at 270 nm is decreased by the presence of DEPC modified tyrosine. These experiments showed an increase in the UV difference spectrum at 240 nm with no change observed at 270 nm. This result showed that the inactivation of mBEI and mBEII was caused by chemical modification of the histidine residues. Amylose and amylopectin protected mBEI and mBEII against DEPC inactivation, with amylopectin being the better substrate. Subsequently, mutational experiments were performed on H340 and H525. Specific activity of the H340A and H525A mutants was reduced to 0.45% and 0.15% of that of wild type activity,

respectively. These mutants have higher K_m values compare to the WT enzyme indicating that these two residues are involved in BE substrate binding.

Mutational studies performed on *E. coli* BE included Y300 which is conserved in the α-amylase family (36,37). Replacement of Y300 with A, D, L, S or W resulted in mutant enzymes with less than 1% of wild type activity (36). The Y300F mutant retained 25% of wild type activity with comparable K_m values and substrate preference. Based on the crystal structures of the members of the α-amylase family it has been proposed that the conserved tyrosine forms a hydrogen bond with the conserved E458 orienting this residue in the optimal position for catalysis (26). The inability of the Y300F mutant to form the important hydrogen bond reflects differences in heat stability and enzymatic activities at elevated temperatures. Indeed the heat stability of Y300F was found to be lower by approximately 5°C than that of WT, indicating the importance of the hydroxyl group for protein stability (36). In summary, the conserved catalytic residues Y300, H340, R403, D405, E458, H525 and D526 are indeed necessary for branching enzyme's activity.

Although members of the α-amylase family of enzymes share structural features and conserved amino acids involved in catalysis, the fact that they catalyze different reactions generates the question of which residues are responsible for the distinct catalytic properties. Analysis of the amino acid sequence alignment reveals a conserved position unique to branching enzymes. The residue E459 is either an aspartate (mostly in eukaryotes) or a glutamic acid depending on the organism. This residue is located after the E458 conserved in the α-amylase family and necessary for BE activity. Mutation of E459 to A, N or K dramatically lowered the specific activity to 30%, 12% and 6%, of

wild type respectively (37). These mutants altered the preference of the substrate from amylose to amylopectin as well as its kinetic parameters. The E459D conservative mutation increases the specific activity of BE, with kinetic properties similar to those of the WT enzyme, although this mutation has an effect in the glucose chain transfer pattern. Comparison of the pH profiles for the WT and E459D mutants rules out the possibility of E459 being involved in acid-base catalysis. All the mutational data previously presented is summarized in Table 1.1.

Glycogen branching not only increases the number of non-reducing ends, thus making glycogen more reactive to synthesis and digestion, but it is also essential for assuring glycogen solubility in the cell. Glycogen in its linear form precipitates in the cell. Accumulation of insoluble glycogen in the cell is known as glycogen storage disease type IV (GSD IV) and is caused by mutations in the gene of the ubiquitously expressed glycogen branching enzyme (38,39). These mutations result in an impaired glycogen metabolism that forbids the formation of the branch points in glycogen, producing an insoluble polymer. This genetic disease occurs in different allelic variants with various clinical presentations. GSD IV in its different forms affects the liver, muscular tissue and the central and peripheral nervous system (40). The classical form of the disease presents progressive liver cirrhosis leading to death in children at the age of five years old. It is caused by either one of the two mutations R515C, F257L (refers to 546 and 306 in *E. coli* numbering) or the C terminal truncation at residue 524 (R524Ter, 555 in *E. coli* numbering). The first two mutations leaves the enzyme with activity between 20% to 27%, while the R524ter truncation produces an inactive enzyme. There

Table 1.1 Mutational studies performed in branching enzyme from maize endosperm isoform II and *E. coli* (30-33,37,41).

Mutation	Approximate Specific Activity (%)	organism	*E. coli* numbering
D386E	2	*maize* (mBEII)	405
D386N	2	*maize* (mBEII)	405
E441D	2	*maize* (mBEII)	458
E441Q	2	*maize* (mBEII)	458
D509E	2	*maize* (mBEII)	526
D509N	2	*maize* (mBEII)	526
R384A	2	*maize* (mBEII)	403
R384S	1	*maize* (mBEII)	403
R384Q	3	*maize* (mBEII)	403
R384E	1	*maize* (mBEII)	403
R384K	12	*maize* (mBEII)	403
H320A	4	*maize* (mBEII)	340
H508A	0	*maize* (mBEII)	525
Y300A	0	*E. coli*	
Y300D	0	*E. coli*	
Y300L	0	*E. coli*	
Y300S	0	*E. coli*	
Y300W	0	*E. coli*	
Y300F	25	*E. coli*	
E459A	30	*E. coli*	
E459K	6	*E. coli*	
E459N	12	*E. coli*	
E459D	185	*E. coli*	
N113GBE	60	*E. coli*	

is a less common slow progressive form that presents liver dysfunction but no liver failure and is associated with two mutations, L224P and Y329S, leaving a totally inactive or a 50% active enzyme, respectively. The neuromuscular form is caused by a 70 residue deletion between amino acids 262 to 331 (311 to 379). It is expressed at birth, with death in the neonatal period. The Y329S mutation is also responsible for a late onset slowly progressive form of the disease that affects the central and peripheral nervous system. There is also a combined hepatic and muscular disorder caused by a single mutation of R524N (555). Some of these mutations present a mild progressive form while others are lethal in early childhood. With few exceptions, GSDIV is a progressive and lethal disease (39,42). Although the human BE gene localized in chromosome 3 has been identified, these mutations have not been tested in bacteria or plant BE (43). As it will be shown later, these mutations are likely to cause the unfolding of the protein, inactivating the enzyme.

1.1.4 Truncated branching enzyme

The determination of the three dimensional structure of branching enzyme has been a goal in Dr. Preiss's lab for many years. Previous attempts to crystallize the WT enzyme were performed by members of his group and were not successful. The technique of limited proteolysis was used to identify a more compact and stable form of the enzyme that would be successful in crystallization (44). Branching enzyme from *E. coli* was subjected to several proteases with a range of specificities. These proteases were proteinase K, protease V8, elastase, chymotrypsin, subtilisin, trypsin and carboxypeptidase Y. All proteases displayed similar digestion patterns producing a 71.6

kDa fragment, with the exception of trypsin. BE lacking the first 112 residues at the amino terminus (N113BE) was analyzed and its properties compared to the WT enzyme. Despite the fact that N113BE was only 60% active, its substrate preference and Km value were similar to the WT. N113BE presents an altered branching pattern with a higher affinity for longer chains of 12 glucose units or more.

Although glycogen synthesis has been a field of study since the 1940's and progress has been achieved in the determination of its mechanism, its chemistry is not fully understood. This is mainly due to the lack of structural models. There are no structures of any of the enzymes involved either in glycogen or starch biosynthesis. The three dimensional structure of branching enzyme is a *de novo* structure that will reveal valuable information that will aid in the understanding of this biosynthetic pathway.

1.2 Angiostatin

1.2.1 Angiogenesis

In mammals, blood vessels are formed by two processes, vasculogenesis and angiogenesis. Vasculogenesis is the new formation of blood vessels through differentiation of precursor cells into endothelial cells, forming a vascular network (45). This process occurs during embryonic development. Angiogenesis is the sprouting of blood vessels from existing capillary beds (46). In adults, new blood vessels are formed exclusively via angiogenesis, with the exception of the transient process of neovascularization that occurs in the female reproductive system. Angiogenesis is essential for development and wound healing. This process is triggered by a growth stimulus that enters the endothelial dormant cells into the cell cycle. The process begins with the degradation of the basement membrane in endothelial cells and lumen formation. Simultaneously, the endothelial cells change their morphology, proliferate, migrate, form microtubes and sprout new capillaries. This is a highly regulated process involving multiple controls that can be turned on or off throughout the process.

Pathological angiogenesis occurs in rheumatoid arthritis, diabetic retinopathy, tumor development, and in metastasis dissemination (47). In rheumatoid arthritis, capillaries invade the joints and destroy the cartilage. In diabetic retinopathy, new capillaries invade the retina causing blindness. Ocular vascularization is the most common cause of blindness, responsible for approximately 20 eye diseases. Tumor growth and metastasis dissemination are angiogenesis dependent processes as well. Avascular tumors can not grow beyond 2 to 3 mm^3, but once the tumor is vascularized, rapid growth is observed (46,47). This network of blood vessels embedding the tumor

not only promotes tumor growth, it also serves as a portal for tumor cells to enter the blood stream and to metastasize to other organs.

Angiogenesis is controlled by specific molecules; angiogenesis activators trigger the process, while angiogenesis inhibitors stop it. These molecules act in concert to maintain a balance in the vasculature with a turnover that can last for years, although during wound healing the turnover lasts days. Once enough tumor cells have switched to the angiogenic phenotype, the tumor itself stimulates the up regulation of angiogenesis activators and down regulates angiogenesis inhibitors (48-51). Some of the currently known angiogenesis activators and inhibitors and their properties are listed in Table 1.2. Among the most common angiogenesis activators found in tumor cells are the vascular endothelial growth factor (VEGF) and the family of fibroblast growth factors (FGF), which includes the basic and acid fibroblast growth factors (bFGF and aFGF). VEGF is a specific activator of endothelial cells, while FGF can act on a variety of cell types. High levels of VEGF have been detected in the majority of human tumors including bladder, breast, lung, gastrointestinal, ovary, prostate, glioblastoma, hemangioma and retinoblastoma (48). High concentrations of bFGF have also been found in cancer patients (49). Moreover, the angiogenesis inhibitor thrombospodin 1 (TSP1) has been found to be down regulated in several tumors (50,51).

Once it was determined that tumor growth and metastasis dissemination were angiogenesis dependent processes, it was proposed that blocking angiogenesis was the best strategy for cancer treatment. This presented a major breakthrough in cancer therapy, which for decades was targeted to destroy tumor cells using cytotoxic agents. The high mutational rate of tumor cells, in addition to the high toxicity that these agents

Table 1.2 Various molecules involved in angiogenesis activation or inhibition (46,52-54)

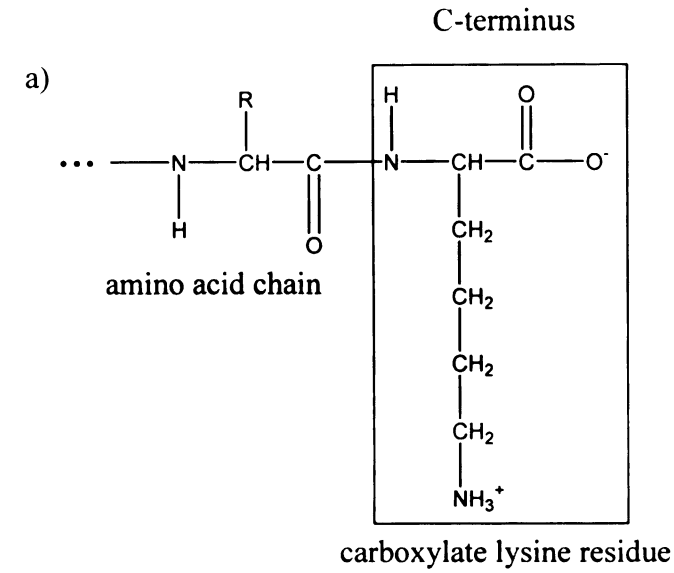
Activators	**Function**
VEGF	stimulates permeability and adhesion
bFGF and aFGF	stimulates angiogenesis
angiopoietin 1 + Tie2 endothelium receptor	stabilize vessels and inhibits permeability
angiopoietin 2 +VEGF	stimulates angiogenesis
tumor necrosis factor	induces production of bFGF
$\alpha_v\beta_5$, $\alpha_v\beta_3$ integrins	mediate cell migration and lumen formation
platelet derived GF	recruits smooth cells
VEGF receptors	integrate angiogenic and survival signals
transforming growth factor	stimulates extracellular matrix production
platelet endothelial cell adhesion protein	endothelial junctional protein

Inhibitors	Function
antithrombin III (fragment), angiostatin, endostatin and interferon-β	inhibits proliferation and/or migration of endothelial cells
TSP1 and 2	endothelial cell proliferation and migration
prolactin	inhibits VEGF and bFGF
VEG inhibitor	modulates cell growth
granulocyte macrophage colony stimulating factor	mobilization of endothelial precursor
angiopoietin 2	cause apoptosis of the vessel
P53	induces transcription of TSP1

present against normal cells, makes chemotherapy less effective and risky. Antiangiogenic therapy targets the proliferation of normal endothelial cells, which represent a uniform target, unlike the fast mutating cancerous cells. In a normal adult only 0.01% of all endothelial cells undergo division at a specific time. Antiangiogenic therapy will inhibit tumor vascularization selectively without affecting normal vasculature. Also, endothelial cells can be easily reached through the blood stream. All these reasons have directed attention to angiogenesis inhibitors as potential anti-cancer agents. Among the angiogenesis inhibitors, angiostatin has been shown to inhibit tumor growth and metastasis dissemination in animal models (55,56). Angiostatin causes no side effects, toxicity or weight loss and is currently in phase I clinical trials at the Thomas Jefferson University Hospital (Philadelphia, PA).

1.2.2 Angiostatin

The angiogenesis inhibitor angiostatin is an N-terminal fragment of plasminogen with a molecular weight of 38 kDa (57). Although angiostatin is a proteolytic fragment of plasminogen, plasminogen is inactive in the inhibition of endothelial cell growth, neovascularization and metastatic tumor growth (55,58). Plasminogen is a zymogen of plasmin activated after a single peptide bond between residues R561 and V562 (human plasminogen numbering) is cleaved. Plasmin is a 92 kDa serine protease that catalyzes the dissolution of blood clots. Plasminogen consists of a catalytic domain and five highly homologous triple disulfide binding domains, known as kringles (Kg). These kringle domains specifically bind the



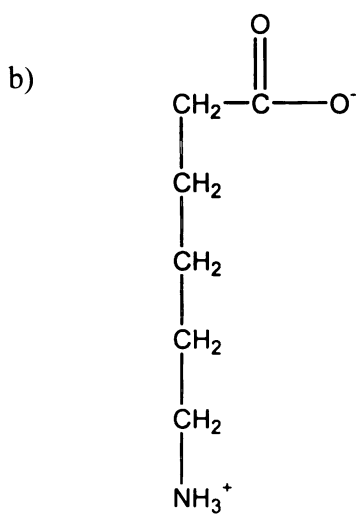


Figure 1.5 a) Carboxylate lysine residue. b) The carboxylate lysine analog, ε-aminocaproic acid (EACA).

carboxylate lysyl residues in fibrin, the skeleton of blood clots (Figure 1.5a). Upon fibrin binding, plasmin exerts its proteolytic activity, solubilizing fibrin and consequently dissolving the blood clot.

X-ray crystal structures of four of the five individual plasminogen kringle domains, Kg1, Kg2, Kg4 and Kg5, have been previously determined (59-62). Their binding modes for lysine-like ligands have also been studied structurally and by site-directed mutagenesis (61-64). The binding center of kringles is known as the lysine binding site (LBS). It consists of a cationic and an anionic center that stabilizes the carboxyl and amino group of a carboxylate lysyl residue (Figure 1.5 and 1.6). The LBS is defined by residues R115, R153, D137, and D139 in Kg1; R234, D219, and E221 in Kg2 and R290, R324, D309 and H317 in Kg3. Binding experiments using the C-terminal lysine analog ε-aminocaproic acid (EACA) showed that Kg1 has high affinity for EACA with a K_D of 15.5 μM (Figure 1.5b and 1.6) (62,65,66). Kg2 has a lower affinity for EACA (401 μM) and Kg3 shows no affinity at all. Kg2 also binds the VEK30 peptide, a 30 residue peptide encompassing residues 85 to 114 within the sequence of the group A pathological *Streptococcal* surface protein (PAM) (67). Kg2 binds VEK30 strongly (K_D = 0.46 μM), presenting a model for protein binding at the surface of bacteria. This peptide consists of a 5 turn α-helix that runs between the anionic and cationic centers of Kg2's LBS (Figure 1.7) (59). VEK30 forms a pseudo lysyl residue with amino acids R101 and E104 that fits in the Kg2's LBS.

Effects of individual or combined kringles on endothelial cell proliferation *in vitro* have been studied as well (57). This study revealed that Kg1-3 has a two fold

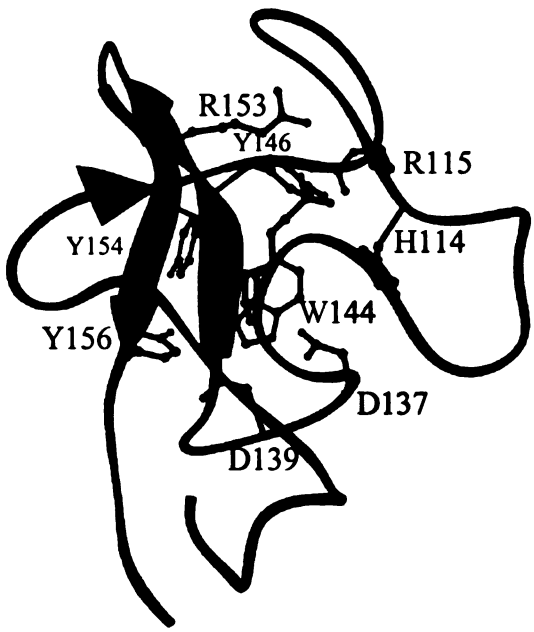


Figure 1.6 Structure of K1-EACA (61). K1 is colored green and the residues encompassing the LBS are shown in atom color (nitrogen, blue; oxygen, red and carbon, green) and EACA is colored lavender. Side chains are labeled using plasminogen numbering.

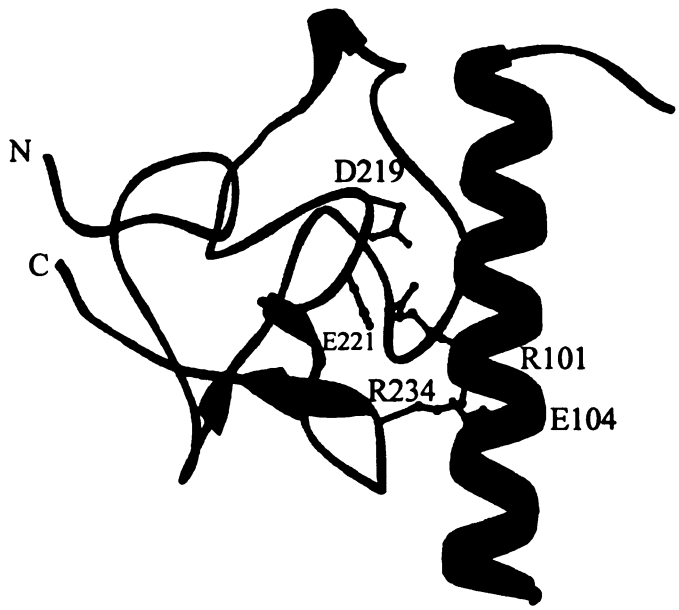


Figure 1.7 X-ray crystallographic structure of K2-VEK30 (59). K2 is colored green and the VEK30 peptide is colored cyan. Side chains are labeled using plasminogen numbering.

inhibitory activity with a half life maximum concentration (ED_{50}) of 70 nM versus 135 nM for Kg1-4. The half life maximum concentration refers to the concentration to achieve 50% inhibition of bFGF-stimulated bovine capillary endothelial cell growth. Kg1 has the highest inhibitory activity, with an ED_{50} of 320 nM. Recombinant Kg2-3 exerts inhibitory activity similar to Kg2 alone, although enhancement of inhibition is observed when individual Kg2 and Kg3 are added together. Other studies performed on animal models showed that agents containing Kg1-3, Kg1-4, Kg1-5, and Kg1-4 plus a fragment of Kg5 have potent antiangiogenic and/or anti-tumor growth activity (57,68-70). These fragments, as well as individual kringle domains, are also inhibitory toward endothelial cell migration and/or proliferation *in vitro*. Studies with recombinant angiostatin show that the maximum tumor inhibitory activity resides in a fragment of angiostatin containing Kg1-3 (29.77 kDa; residues 79 to 338) (71). This smaller fragment is the one currently used in clinical trials.

It has been observed that once a primary tumor is removed, dormant distant metastasis may grow (72-76). A number of animal experiments showed that primary tumors sometimes inhibits the growth of their metastasis (77-79). Although some hypotheses have been proposed to explain this phenomenon, none of them has provided a molecular explanation for this mechanism. The best explanation for this phenomenon occurred after angiostatin was isolated from serum and urine of mice bearing Lewis lung carcinoma (58). These experiments demonstrated that in fact, tumors promote the release of angiostatin into the bloodstream to inhibit the development of metastasis and therefore prevent competition.

Three molecules that mediate cleavage of plasminogen into angiostatin *in vivo* have been identified. In Lewis lung carcinoma, the proteolytic activity of elastase catalyses this transformation (57,58). It was also found that tumor cells up regulate the production of elastase. In human prostate carcinoma, plasmin works as an enzyme and substrate to generate angiostatin in the presence of free sulfhydral donors (70). The antiangiogenic properties of pharmacological sulfhydril donors such as D-penicillamine and captopril have been previously reported (80-83). Another proposed mechanism for the generation of angiostatin involves the enzyme phosphoglycerate kinase (84). This enzyme reduces the disulfide links in Kg5, a process that triggers the cleavage of plasminogen into angiostatin. High levels of phosphoglycerate kinase have been observed in plasma of mice bearing fibrosarcoma tumors. Moreover, administration of recombinant phosphoglycerate kinase increases the plasma levels of angiostatin by 86% and inhibits tumor growth between 50% to 70%, depending on the type of tumor. Also, reduced fibrinolysis and hypercoagulation is often observed in cancer patients (85).

Human angiostatin potently inhibits the growth of transplanted tumors in mice (55,56). The growth of Lewis lung carcinoma, T241 fibrosarcoma, and reticulum cell sarcoma tumors were inhibited by an average of 84% at doses of 100 mg/kg/day. These tumors had poor response to other therapies. Angiostatin also inhibits the growth of human breast carcinoma by 95%, colon carcinoma by 97% and prostate carcinoma by almost 100% (55). Angiostatin also reduced the size of brain tumors in mice by more than 71% (69). Brain tumors are known to be problematic to treat due to the difficulty of crossing the brain-blood barrier, which hinders drug delivery. Angiostatin treatment does not result in weight loss or other toxicity in mice.

The mechanisms by which angiostatin inhibits angiogenesis are still unclear. It has been observed that in angiostatin treated tumors, the rate of apoptosis (process by which a cell actively commits suicide) in tumor cells increases to five times that of the control (58,86) (55). Although the mechanism by which angiostatin therapy leads to an increase in death rate of tumor cells is not known, it is believed that angiostatin may bring the tumor to a dormant stage between cell proliferation and apoptosis. It has also been observed that angiostatin interacts with several proteins involved in endothelial cell lysis, migration or proliferation. Angiostatin binds the α/β subunit of ATP synthase, angiomotin and the $\alpha_v\beta_3$ integrin receptor (87,88). The α/β subunit of ATP synthase is found on the surface of several tumor cell lines. ATP synthase catalyzes the transport of H^+ across the membrane, resulting in tumor cell lysis. Furthermore, previous studies demonstrated that the addition of ATP synthase to cultures of tumor cell lines induces lysis of the cell (89-92). In this case angiostatin binding may mediate the mechanism for inhibition of endothelial cell growth and tumor apoptosis. It has been reported recently that angiostatin also binds angiomotin, a protein involved in endothelial cell migration (88). Moreover, angiostatin inhibited migration in angiomotin expressing cells but not in the absence of angiomotin. This indicates that angiostatin inhibits cell migration by interfering with angiomotin activity in endothelial cells. An important specific binding is its interaction with the $\alpha_v\beta_3$ integrin receptor (93). $\alpha_v\beta_3$ is an endothelial cell surface receptor implicated in the activation of angiogenesis. The angiostatin interaction can be blocked by EACA at a concentration high enough to occupy the LBS in Kg2 (much higher than that needed for Kg1). This indicates that Kg2 is more important than Kg1 for integrin binding. Although the mechanism of action for angiostatin is still unknown. It is

known that angiostatin is one of the most potent inhibitors for angiogenesis and it has proved to be successful in decreasing tumor size and metastasis in animal models. To better understand the structure and function of angiostatin, we have determined its three dimensional structure to a resolution of 1.75 Å. We hope that this structure will facilitate the development of more effective anti-cancer therapies.

1.3 Literature Cited

1. Preiss, J., and Romeo, T. (1989) *Adv Microb Physiol* **30,** 183-238.

2. Dawes, E. A., and Senior, P. J. (1973) *Adv Microb Physiol* **10,** 135-266.

3. Preiss, J. (1984) *Annu Rev Microbiol* **38,** 419-458.

4. Preiss, J. (1989) in *Bacteria in Nature* (Poindexter, J. S., and Leadbetter, E., eds) Vol. 3, pp. 189-258, Plenum Publishing Corp., New York

5. Preiss, J. (1996) in *Cellular and Molecular Biology* Vol. 1, pp. 1015-1024, Amer. Soc. Microbiol., Washington, DC

6. Ballicora, M. A., Laughlin, M. J., Fu, Y., Okita, T. W., Barry, G. F., and Preiss, J. (1995) *Plant Physiol* **109**(1), 245-251.

7. Tandecarz, J. S., and Cardini, C. E. (1978) *Biochim Biophys Acta* **543**(4), 423-429.

8. Krisman, C. R., and Barengo, R. (1975) *Eur J Biochem* **52**(1), 117-123.

9. Lomako, J., Lomako, W. M., and Whelan, W. J. (1988) *Faseb J* **2**(15), 3097-3103.

10. Pitcher, J., Smythe, C., Campbell, D. G., and Cohen, P. (1987) *Eur J Biochem* **169**(3), 497-502.

11. Baecker, P. A., Greenberg, E., and Preiss, J. (1986) *J Biol Chem* **261**(19), 8738-8743.

12. Mizuno, K., Kimura, K., Arai, Y., Kawasaki, T., Shimada, H., and Baba, T. (1992) *J Biochem (Tokyo)* **112**(5), 643-651.

13. Morell, M. K., Blennow, A., Kosar-Hashemi, B., and Samuel, M. S. (1997) *Plant Physiol* **113**(1), 201-208.

14. Fisher, D. K., Gao, M., Kim, K. N., Boyer, C. D., and Guiltinan, M. J. (1996) *Plant Mol Biol* **30**(1), 97-108.

15. Gao, M., Fisher, D. K., Kim, K. N., Shannon, J. C., and Guiltinan, M. J. (1997) *Plant Physiol* **114**(1), 69-78.

16. Guan, H., Li, P., Imparl-Radosevich, J., Preiss, J., and Keeling, P. (1997) *Arch. Biochem. Biophys.* **342,** 92-100.

17. Hong, S., and Preiss, J. (2000) *Arch Biochem Biophys* **378**(2), 349-355.

18. Kuriki, T., Stewart, D. C., and Preiss, J. (1997) *J Biol Chem* **272**(46), 28999-29004.

19. Baba, T., Kimura, K., Mizuno, K., Etoh, H., Ishida, Y., Shida, O., and Arai, Y. (1991) *Biochem Biophys Res Commun* **181**(1), 87-94.

20. Romeo, T., Kumar, A., and Preiss, J. (1988) *Gene* **70**(2), 363-376.

21. Svensson, B. (1994) *Plant Mol Biol* **25**(2), 141-157.

22. Matsuura, Y., Kusunoki, M., Harada, W., and Kakudo, M. (1984) *J Biochem (Tokyo)* **95**(3), 697-702.

23. Buisson, G., Duee, E., Haser, R., and Payan, F. (1987) *Embo J* **6**(13), 3909-3916.

24. Boel, E., Brady, L., Brzozowski, A. M., Derewenda, Z., Dodson, G. G., Jensen, V. J., Petersen, S. B., Swift, H., Thim, L., and Woldike, H. F. (1990) *Biochemistry* **29**(26), 6244-6249.

25. Klein, C., and Schulz, G. E. (1991) *J Mol Biol* **217**(4), 737-750.

26. Katsuya, Y., Mezaki, Y., Kubota, M., and Matsuura, Y. (1998) *J Mol Biol* **281**(5), 885-897.

27. Uitdehaag, J. C., van Alebeek, G. J., van Der Veen, B. A., Dijkhuizen, L., and Dijkstra, B. W. (2000) *Biochemistry* **39**(26), 7772-7780.

28. Brzozowski, A. M., and Davies, G. J. (1997) *Biochemistry* **36**(36), 10837-10845.

29. Jespersen, H. M., MacGregor, E. A., Sierks, M. R., and Svensson, B. (1991) *Biochem J* **280**, 51-55.

30. Kuriki, T., Guan, H., Sivak, M., and Preiss, J. (1996) *J Protein Chem* **15**(3), 305-313.

31. Cao, H., and Preiss, J. (1996) *J Protein Chem* **15**(3), 291-304.

32. Libessart, N., and Preiss, J. (1998) *Arch Biochem Biophys* **360**(1), 135-141.

33. Funane, K., Libessart, N., Stewart, D., Michishita, T., and Preiss, J. (1998) *J Protein Chem* **17**(7), 579-590.

34. Miles, E. W. (1977) *Meth. Enzymol.* **47**, 431-442

35. Lundblad, R. L. (1991) *Chemical Reagents for Protein Modification*, 2nd Ed.
36. Mikkelsen, R., Binderup, K., and Preiss, J. (2001) *Arch Biochem Biophys* **385**(2), 372-377.

37. Binderup, K., and Preiss, J. (1998) *Biochemistry* **37**(25), 9033-9037.

38. DiMauro, S., and Tsujino, S. (1994) in *Myology* (A.G., E., and C., F.-A., eds) Vol. 2, pp. 1554-1576.

39. Chen, Y. T., and Burchell, A. (1995) in *The metabolic and molecular basis of inherited diseases* (C.R., S., A.L., B., W.S., S., and D., V., eds) Vol. 1, pp. 935-965.

40. Bao, Y., Kishnani, P., Wu, J. Y., and Chen, Y. T. (1996) *J Clin Invest* **97**(4), 941-948.

41. Binderup, K., Mikkelsen, R., and Preiss, J. (2001) *Arch. Biochem. Biophys.* , in press

42. Schroder, J. M., May, R., Shin, Y. S., Sigmund, M., and Nase-Huppmeier, S. (1993) *Acta Neuropathol* **85**(4), 419-430.

43. Thon, V. J., Khalil, M., and Cannon, J. F. (1993) *J Biol Chem* **268**(10), 7509-7513.

44. Binderup, K., Mikkelsen, R., and Preiss, J. (2000) *Arch Biochem Biophys* **377**(2), 366-371.

45. Risau, W. (1995) *Faseb J* **9**(10), 926-933.

46. Folkman, J., and Shing, Y. (1992) *J Biol Chem* **267**(16), 10931-10934.

47. Folkman, J. (1995) *Nat Med* **1**(1), 27-31.

48. Ferrara, N., and Davis-Smyth, T. (1997) *Endocr Rev* **18**(1), 4-25.

49. Nguyen, M., Watanabe, H., Budson, A. E., Richie, J. P., Hayes, D. F., and Folkman, J. (1994) *J Natl Cancer Inst* **86**(5), 356-361.

50. Grossfeld, G. D., Ginsberg, D. A., Stein, J. P., Bochner, B. H., Esrig, D., Groshen, S., Dunn, M., Nichols, P. W., Taylor, C. R., Skinner, D. G., and Cote, R. J. (1997) *J Natl Cancer Inst* **89**(3), 219-227.

51. Dameron, K. M., Volpert, O. V., Tainsky, M. A., and Bouck, N. (1994) *Science* **265**(5178), 1582-1584.

52. Yancopoulos, G. D., Davis, S., Gale, N. W., Rudge, J. S., Wiegand, S. J., and Holash, J. (2000) *Nature* **407**(6801), 242-248.

53. Carmeliet, P., and Jain, R. K. (2000) *Nature* **407**(6801), 249-257.

54. Zhang, L., Chen, Q. R., and Mixson, A. J. (2000) *Curr. Genomics* **1**(2), 117-133.

55. O'Reilly, M. S., Holmgren, L., Chen, C., and Folkman, J. (1996) *Nat Med* **2**(6), 689-692.

56. Sim, B. K., O'Reilly, M. S., Liang, H., Fortier, A. H., He, W., Madsen, J. W., Lapcevich, R., and Nacy, C. A. (1997) *Cancer Res* **57**(7), 1329-1334.

57. Cao, Y., Ji, R. W., Davidson, D., Schaller, J., Marti, D., Sohndel, S., McCance, S. G., O'Reilly, M. S., Llinas, M., and Folkman, J. (1996) *J Biol Chem* **271**(46), 29461-29467.

58. O'Reilly, M. S., Holmgren, L., Shing, Y., Chen, C., Rosenthal, R. A., Moses, M., Lane, W. S., Cao, Y., Sage, E. H., and Folkman, J. (1994) *Cell* **79**(2), 315-328.

59. Rios-Steiner, J. L., Schenone, M., Mochalkin, I., Tulinsky, A., and Castellino, F. J. (2001) *J Mol Biol* **308**(4), 705-719.

60. Mulichak, A. M., Tulinsky, A., and Ravichandran, K. G. (1991) *Biochemistry* **30**(43), 10576-10588.

61. Mathews, I., Vanderhoff-Hanaver, P., Castellino, F. J., and Tulinsky, A. (1996) *Biochemistry* **35**(8), 2567-2576.

62. Chang, Y., Mochalkin, I., McCance, S. G., Cheng, B., Tulinsky, A., and Castellino, F. J. (1998) *Biochemistry* **37**(10), 3258-3271.

63. Wu, T. P., Padmanabhan, K., Tulinsky, A., and Mulichak, A. M. (1991) *Biochemistry* **30**(43), 10589-10594.

64. McCance, S. G., Menhart, N., and Castellino, F. J. (1994) *J Biol Chem* **269**(51), 32405-32410.

65. Marti, D., Schaller, J., Ochensberger, B., and Rickli, E. E. (1994) *Eur J Biochem* **219**(1-2), 455-462.

66. Burgin, J., and Schaller, J. (1999) *Cell Mol Life Sci* **55**(1), 135-141.

67. Nilsen, S. L., Prorok, M., and Castellino, F. J. (1999) *J Biol Chem* **274**(32), 22380-22386.

68. Cao, R. H., Wu, H. L., Veitonmaki, N., Linden, P., Farnebo, J., Shi, G. Y., and Cao, Y. H. (1999) *Proc Nat Acad Sci USA* **96**(10), 5728-5733.

69. Joe, Y. A., Hong, Y. K., Chung, D. S., Yang, Y. J., Kang, J. K., Lee, Y. S., Chang, S. I., You, W. K., Lee, H., and Chung, S. I. (1999) *Int J Cancer* **82**(5), 694-699.

70. Gately, S., Twardowski, P., Stack, M. S., Cundiff, D. L., Grella, D., Castellino, F. J., Enghild, J., Kwaan, H. C., Lee, F., Kramer, R. A., Volpert, O., Bouck, N., and Soff, G. A. (1997) *Proc Natl Acad Sci U S A* **94**(20), 10868-10872.

71. MacDonald, N. J., Murad, A. C., Fogler, W. E., Lu, Y. Y., and Sim, B. K. L. (1999) *Biochem Biophys Res Comm* **264**(2), 469-477.

72. Southam, C., and Brunschwig, A. (1961) *Cancer* **14**, 971-978.

73. Warren, B. A., Khan, S., and Chauvin, W. J. (1977) *Bibl Anat* **(16**(Pt 2)), 396-402.

74. Woodruff, M. (1980) *The interactions of cancer and host*, Grune and Stratton, New York

75. Fidler, I. J., and Balch, C. M. (1987) *Curr Probl Surg* **24**(3), 129-209.

76. Clark, W. H., Jr., Elder, D. E., Guerry, D. t., Braitman, L. E., Trock, B. J., Schultz, D., Synnestvedt, M., and Halpern, A. C. (1989) *J Natl Cancer Inst* **81**(24), 1893-1904.

77. Gorelik, E., Segal, S., and Feldman, M. (1978) *Int J Cancer* **21**(5), 617-625.

78. Gorelik, E., Segal, S., and Fredman, M. (1980) *J. Natl. Cancer Inst.* **89**, 219-227.

79. Bonfil, R. D., Ruggiero, R. A., Bustuoabad, O. D., Meiss, R. P., and Pasqualini, C. D. (1988) *Int J Cancer* **41**(3), 415-422.

80. Volpert, O. V., Ward, W. F., Lingen, M. W., Chesler, L., Solt, D. B., Johnson, M. D., Molteni, A., Polverini, P. J., and Bouck, N. P. (1996) *J Clin Invest* **98**(3), 671-679.

81. Brem, S. S., Zagzag, D., Tsanaclis, A. M., Gately, S., Elkouby, M. P., and Brien, S. E. (1990) *Am J Pathol* **137**(5), 1121-1142.

82. Matsubara, T., and Ziff, M. (1987) *J Clin Invest* **79**(5), 1440-1446.

83. Matsubara, T., Saura, R., Hirohata, K., and Ziff, M. (1989) *J Clin Invest* **83**(1), 158-167.

84. Lay, A. J., Jiang, X. M., Kisker, O., Flynn, E., Underwood, A., Condron, R., and Hogg, P. J. (2000) *Nature* **408**(6814), 869-873.

85. Green, K. B., and Silverstein, R. L. (1996) *Hematol Oncol Clin North Am* **10**(2), 499-530.

86. Holmgren, L., O'Reilly, M. S., and Folkman, J. (1995) *Nat Med* **1**(2), 149-153.

87. Moser, T. L., Stack, M. S., Asplin, I., Enghild, J. J., Hojrup, P., Everitt, L., Hubchak, S., Schnaper, H. W., and Pizzo, S. V. (1999) *Proc Nat Acad Sci USA* **96**(6), 2811-2816.

88. Troyanovsky, B., Levchenko, T., Mansson, G., Matvijenko, O., and Holmgren, L. (2001) *J Cell Biol* **152**(6), 1247-1254.

89. Rozengurt, E., Heppel, L. A., and Friedberg, I. (1977) *J Biol Chem* **252**(13), 4584-4590.

90. Rozengurt, E., and Heppel, L. A. (1979) *J Biol Chem* **254**(3), 708-714.

91. Chahwala, S. B., and Cantley, L. C. (1984) *J Biol Chem* **259**(22), 13717-13722.

92. Saribas, A. S., Lustig, K. D., Zhang, X., and Weisman, G. A. (1993) *Anal Biochem* **209**(1), 45-52.

93. Tarui, T., Miles, L. a., and Takada, Y. (2001) *J. Mol. Biol.* **276**(43), 39562-39568.

CHAPTER II: X-RAY CRYSTAL STRUCTURE DETERMINATION

2.1 Branching enzyme

The first step in every structure determination is the production of single and well diffracting crystals. For this purpose it is essential to have access to large quantities of highly pure material. The protein is set up for crystallization using sparse matrices of precipitating solutions that have been proven successful in crystallizing other proteins. The most common method for the crystallization of macromolecules is the hanging drop vapor diffusion method (Figure 2.1). In this method, a drop containing a mixture of protein and precipitating solution is equilibrated against a reservoir containing the precipitant. The protein is slowly precipitated and the molecules adopt identical orientations forming an orderly three dimensional array of molecules held together by non-covalent interactions. The crystallization process not only involves setting up thousands of drops, but it also involves the constant monitoring of these drops. It is important to look for precipitation behavior, relative solubility and the appearance of crystals. Based on the observations from the initial sparse screens, new optimized screens can be made. This turns into an iterative process that can produce crystals suitable for X-ray diffraction data collection. Such a strategy was followed in the crystallization of BE and angiostatin.

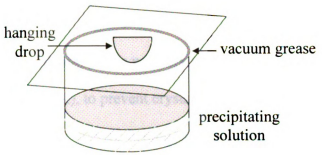


Figure 2.1 Crystallization using the hanging drop vapor diffusion method

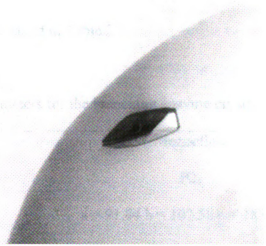


Figure 2.2 A monoclinic crystal of glycogen branching enzyme. The crystal has dimensions of 0.3 x 0.1 x 0.1 mm^3.

2.1.1 Crystallization

The recombinant native and selenium methionine (SeMet) substituted BE lacking the first 112 residues were overexpressed and purified by Dr. Binderup, a member of Dr. Preiss's laboratory (1,2).

The native (Figure 2.2) and SeMet BEs were crystallized and X-ray data was collected at cryogenic temperatures (123K), to prevent crystal damage.

The branching enzyme crystals belong to the $P2_1$ space group with unit cell parameters a = 91.44 Å, b = 102.58 Å, c = 185.41 Å, and $\beta = 91.38^{\circ}$. Assuming four molecules of branching enzyme (71.6 kDa) per asymmetric unit, the crystal volume per protein mass is 3.1 $Å^3\ Da^{-1}$ corresponding to approximately 56.5% solvent in the crystal. This value is within the range observed for protein crystals (3). The crystal parameters for the branching enzyme crystal are listed in Table 2.1. Data was 99.6% complete for 152,002 unique reflections derived from a total of 499,161 reflections. Detailed data collection statistics are found in Table 2.2.

Table 2.1 Crystal parameters for the branching enzyme crystal.

Crystal form	Monoclinic
Space group	$P2_1$
Unit cell	a = 91.44 b = 102.58 c = 185.41 Å
	$\alpha = \gamma = 90^{\circ}$ and $\beta = 91.38^{\circ}$
Solvent content	56%
Molecules per asymmetric unit	4

Table 2.2 Statistics for the branching enzyme X-ray diffraction data collection ¶

	Native [1]	Se-Met [2]	Hg soak [3]
Wavelength (Å)0.97794		0.97938	1.54180
Resolution range (Å)	35.0 – 2.3	20 – 2.5	40 – 3.5
(last resolution shell)	(2.38 – 2.30)	(2.59 – 2.50)	(3.63 – 3.50)
Cell parameters (Å, deg)			
(a, b, c and β)	91.48, 102.56,	91.65, 102.48,	91.57, 102.79,
	185.10 and 91.45	195.92 and 91.68	185.58 and 91.68
Completeness (%)	99.6 (98.6)	94.2 (77.5)	91.5 (86.3)
R_{merge} (I) ¥ (%)	8.6 (30.3)	10.1 (29.2)	22.8 (50.0)
$\langle I \rangle / \langle \sigma_I \rangle$	10.4 (2.6)	9.3 (1.5)	5.1 (2.3)

1. Data collected at the Advanced Photon Source, Structural Biology Center ID19 beamline

2. Data collected at the Advanced Photon Source, IMCA beamline ID17

3. Data collected at Michigan State University Macromolecular X-ray Facility home source

¶ Values in parentheses refer to the last resolution shell

¥ $R_{merge} = \Sigma |I - \langle I \rangle| / \Sigma I$, where I is an individual intensity measurement and $\langle I \rangle$ is the average intensity for this reflection, with summation over all data

2.1.2 Structure determination

Once high resolution data have been obtained, an electron density map is calculated. It is from this electron density map that the three dimensional structure of the protein is deduced. This map not only has the positions of the main and side chain atoms but also includes solvent molecules, ions, substrates and any other molecules that form the lattice. The electron density is represented by the following equation,

$$\rho(x,y,z)=\frac{1}{V}\sum_{h}\sum_{k}\sum_{l}F_{hkl}e^{-2\pi(hx+ky+lz)}$$

The structure factor (F_{hkl}) for each reflection labeled *hkl*, is a complete description of all the atoms that contribute to that reflection. F_{hkl} is a wave function with frequency, amplitude and phase. The frequency is the same as the X-ray source and the amplitude is proportional to the square root of the measured intensity of the reflection. The only information that is unknown is the phase; this is known in crystallography as the phase problem.

The phase problem can be solved by any of the following methods; molecular replacement, isomorphous replacement or multiple wavelength anomalous dispersion. The molecular replacement method is used when the protein is homologous to another protein for which the structure has been solved.

In the case of a *de novo* structure determination, anomalous dispersion, isomorphous replacement, or a combination of both methods can be used. BE was a *de novo* structure determination and its phases were a hard and challenging problem to solve. This was because BE is not only a large enzyme (616 residues for the truncated form), but there are four molecules of BE in the asymmetric unit. The asymmetric unit is

the simplest volume of the unit cell that can generate the complete lattice by symmetry operations. The solution of an X-ray crystal structure involves the determination of the contents of the asymmetric unit. The asymmetric unit can consist of one, or more than one, molecule and each of the molecules within the asymmetric unit will be related to each other by non-crystallographic symmetry operators. The structure of BE was solved by a combination of the isomorphous replacement and anomalous dispersion methods.

2.1.2.1 Heavy atom isomorphous replacement

In this technique, differences in the diffraction pattern are measured after the introduction of a heavy atom. In order to be able to record these differences reliably, the atom or group of atoms must have many electrons. If hard ions (lanthanides and actinides) are use as derivatives, the interaction between the protein and the heavy atom is usually ionic. On the other hand, soft ions (mercury, gold, platinum, etc.) tend to react with sulfurs on cysteines, deprotonated nitrogen on histidines or even with sulfur from methionine (4). These interactions are covalent, so they tend to bind more specifically. For this technique to work, it is important that the protein molecules within the crystal be identically bound by the heavy atom compound, with essentially no change in the structure of the native crystal lattice.

The coordinates and the diffraction pattern of the heavy atom alone can be determined by calculating the difference between the diffraction patterns of the native crystal and the heavy atom replaced crystal. This is because the contributions from every atom to a reflection are combined in an additive way. With the diffraction pattern of just the heavy atom, we have a small number of atoms and an easier structure to be

determined. For example, BE has four Hg atoms in its asymmetric unit, a rather simpler pattern than that of the more than 20,000 non-hydrogen atoms in the protein structure.

A systematic and exhaustive search for isomorphous heavy atom derivatives was performed; all 37 heavy atoms tried are listed in Table 2.3. Two different methods were used to obtain the heavy atom derivative. In the first method, pre-grown crystals were soaked with the heavy atom reagents at various concentrations through different time spans. Another method involved the co-crystallization of the crystal in the presence of the heavy atom at various concentrations. However, crystals could not be grown in the presence of any of the heavy atoms, not even at heavy atom concentrations as low as 1 μM.

The search for heavy atoms was done by soaking the crystals in a 1mM heavy atom solution overnight. Visible signs of reaction or crystal degradation such as cracking or color change were looked for. Based on these results, repeated soaks were done at lower concentrations. Compounds that show no evidence of degradation were tried at a higher concentration. The soaking time was varied in both cases as well. Data sets were compared with the native for intensity differences and isomorphism. From all the heavy atoms tried, only mercuric chloride, mercuric acetate, mercurochrome, uranyl nitrate, p-chloro mercuri benzoic acid and p-chloro mercuri benzene sulphonate were isomorphous and presented significant intensity differences.

The automated heavy atom Patterson search routines from the program SOLVE were used to find the heavy atom positions (5). The Patterson (P_{uvw}) search is performed by the calculation of a vector map known as the Patterson function. This Patterson map is calculated based on the product of the electron density at two points separated by a

distance u,

$$P_{uvw} = \int_V \rho(x,y,z)\rho(x+u,y+v,z+w)dV$$

$$P_{uvw} = \frac{1}{V}\sum_h\sum_k\sum_l \left|F^2{}_{hkl}\right| e^{-2\pi i(hu+kv+lw)}$$

The Patterson map is calculated using only the contribution from the heavy atom.

$$P_{uvw} = \frac{1}{V}\sum_h\sum_k\sum_l \left|F^2{}_{hkl}\right|_H e^{-2\pi i(hu+kv+lw)}$$

where $|F_{hkl}|_H$ is the structure factor contribution of the derivative. A peak will only be observed when there is an electron density contribution from the heavy atom at both positions (xyz and x + u, y + v, x + w). The result obtained from the Patterson function, using only the contribution of the heavy atom, are peaks located at positions defined by interatomic vectors between heavy atoms. Symmetry within the unit cell is used to transform these interatomic vectors into crystallographic coordinates.

This process, although simplified from the original protein structure, can be cumbersome and time consuming. Today we have the advantage of fast and automated programs that will perform these calculations.

Although the mercury positions were determined, this experiment didn't provide enough phasing information to solve the structure. The poor diffraction of the derivatives (3.5 to 5.0 Å) resulted in phases that were useful only to low resolution.

Table 2.3 List of all the heavy atom compounds tried

1	$HgCl_2$	Mercuric chloride
2	$Hg(C_2H_3O_2)_2$	Mercuric acetate
3	$C_{13}H_{17}HgNO_6$	2-(*N*-(3-hydroxymercuri-2-methoxypropyl)carbamoyl) phenoxyacetic acid
4	$C_9H_9HgO_2S$	Thimerosal (EMTS)
5	$C_{20}H_8Br_2HgO_6Na_2$	Mercurochrome
6	$C_6H_4HgClSO_3$	p-Chloromercuri benzene sulphonate
7	CH_3ClHg	Methyl mercury chloride
8	HgCl	Mercurous chloride
9	$Hg(SCN)_2$	Mercuric thiocyanate
10	HgO	Mercuric oxide
11	$C_7H_5ClHgO_2$	p-Chloromercuri benzoic acid
12	K_2PtCl_4	Platinum (II) potassium chloride
13	K_2PtCl_6	Platinum (IV) potassium chloride
14	$K_2Pt(CN)_4$	Potassium tetracyano platinate
15	$KAuCl_4$	Potassium tetracyano aureate III
16	$KAu(CN)_2$	Potassium dicyano aureate
17	$(CH_3)_3Pb(C_2H_3O_2)$	Trimethyl lead acetate
18	$C_6H_9LaO_6$	Lanthanum acetate
19	$LaCl_3$	Lanthanum chloride
20	$La(NO_3)_3$	Lanthanum nitrate
21	La_2O_3	Lanthanum oxide
22	$UO_2(CH_3CO_2)_2$	Uranyl acetate
23	$UO_2(NO_3)_2$	Uranyl nitrate
24	$Sm\ (CH_3CO_2)_3$	Samarium acetate
25	Sm_2O_3	Samarium oxide
26	$Tl(C_2H_3O_2)$	Thallium acetate
27	$Pr(C_2H_3O_2)_3$	Praseodymium acetate
28	$Ce(C_2H_3O_2)_3$	Cerium acetate
29	Na_2WO_4	Sodium tungstate
30	$YbCl_3$	Ytterbium chloride
31	$NbCl_3$	Niobium chloride
32	K_2OsCl_6	Potassium hexachloroosmiate
33	$ErCl_3$	Erbium chloride
34	KI	Potassium iodine
35	$EuCl_3$	Europium chloride
36	HgI_2	Mercury iodine
37	$LuCl_3$	Lutetium chloride

2.1.2.2 Anomalous dispersion

Like the isomorphous replacement method, the strategy in anomalous dispersion is to solve a smaller and simpler structure. In this case, anomalous scatterers are introduced in the protein molecule. In BE, we substituted the sulfurs in the methionines with selenium by expressing the protein in the presence of SeMet. Advantage is taken of the anomalous differences by irradiating the crystal with X-ray radiation at the absorption edge of the scatterer (0.98 Å for Se).

A single wavelength anomalous dispersion (SAD) experiment was performed at the selenium absorption edge to a resolution of 2.5 Å. BE has a total of 64 Se (16 Se/monomer), so this substructure was somewhat complex to solve. In this case, the 64 Se will provide enough phasing information, but the difficulty resides in finding the positions of all 64 seleniums. To simplify the problem, we combined the SAD data with the single isomorphous replacement (SIR) experimental data, previously collected. Search routines were used with the SAD data and different combinations of the six SIR data sets (5). Of all the SIR data sets, the p-chloro mercuri benzoic acid soak was the most successful in locating Se sites. Four mercury sites and 27 of the total 64 selenium sites were eventually identified (5). These sites were refined and an initial set of phases was calculated using the program SHARP (6). This information was used to identify a total of 61 selenium sites (6).

Among all the programs available, SHARP presents the most powerful software by using the maximum likelihood method and taking into account variables previously ignored (6). The maximum likelihood method follows the principle of least squares with the difference that the calculated and observed values are actually distributions and not

set values. The parameters are then varied in such a way that the calculated values approach the observed ones. The first programs that were use to determine the phases employed a straight implementation of the least square method. Those programs took the phase information generated by the derivative in the refinement and in the recalculation of the structure factors of the native data, introducing bias into the observable. Also, previously developed programs assumed that the native amplitude measurement is error free causing the measurement errors to be combined with the error caused by the lack of isomorphism of the derivative, underestimating the phasing power of the derivative. SHARP assigns a probability distribution not only to the phase but also to the structure factor. This new generation of programs made the determination of the structure of BE possible, something that under the same circunstances would not have been possible ten years ago.

The contribution of the derivative to the phase determination is known as the phasing power (PP). The PP is defined by the following equation,

$$\frac{\sum_{hkl} \left| F_{hkl} \right|_H^2}{\sum_{hkl} \left(\left| F_{hkl,obs} \right|_{PH} - \left| F_{hkl,calc} \right|_{PH} \right)^2}$$

where $|F_{hkl}|_H$ is the structure factor contribution of the derivative, $|F_{hkl,obs}|_{PH}$ is the observed structure factor of the protein plus derivative and $|F_{hkl,calc}|_{PH}$ is the structure factor of the protein plus derivative calculated based on the phases determined by the derivative. A table of the phasing power versus resolution for the mercury derivative, and the selenium SAD data set is presented in Table 2.4.

With this information, an electron density map was calculated; and although the quality of this map was poor, it was good enough to determine the boundaries of each molecule in the asymmetric unit (Figure 2.3a). The non-crystallographic symmetry elements were determined and the quality of the initial electron density map was improved by applying non-crystallographic four fold symmetry averaging (Figure 2.3b). Using this map we were able to build all four molecules in the asymmetric unit of BE (Figure 2.3b). Figure 2.3 shows the electron density map calculated before averaging (Figure 2.3a), after averaging (Figure 2.3b) and the final map after refinement (Figure 2.3c).

Table 2.4 Phasing power of the mercury and selenium methionine derivatives

Resolution, Å	Hg Phasing power	Se Phasing power isomorphous	Se Phasing power anomalous
10.76	1.97	2.08	2.24
6.85	1.71	2.09	1.92
5.34	1.26	1.45	1.47
4.52	0.965	1.09	1.23
4.00	0.800	0.948	1.00
3.61	0.705	0.815	0.834
3.33		0.812	0.769
2.63		0.788	0.712

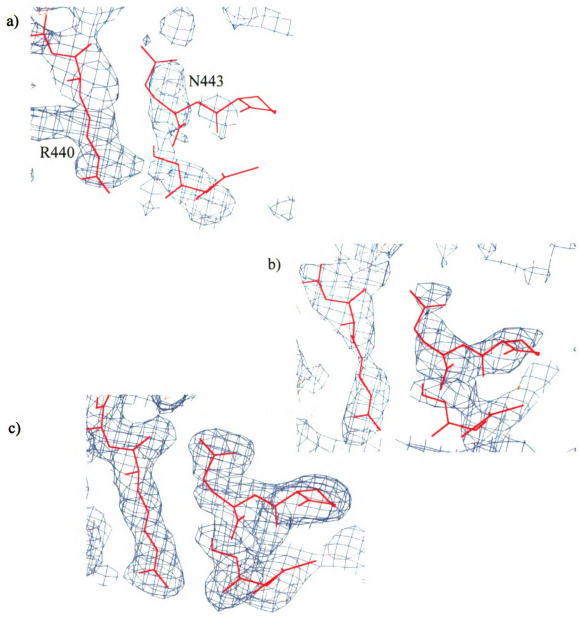


Figure 2.3 A section of the electron density map. a) Initial experimental electron density map b) Electron density map after four fold averaging c) Final electron density map after refinement.

2.1.3 Structure refinement

Once the BE structure was traced, it was subjected to multiple rounds of structure refinement. Structure refinement involves the adjustment of the model to find a closer agreement between the calculated and the observed structure factors. The agreement index between the calculated and observed structure factors is represented by the R_{factor} described below:

$$R_{factor} = \frac{\sum \left| |F_{obs}| - |F_{calc}| \right|}{\sum |F_{obs}|} \times 100$$

The BE structure was refined using the simulated annealing method (7,8). This method uses molecular dynamics to simulate the various parameters in the conformational space of the molecule. In the annealing process the molecule is heated until all particles arrange themselves in the liquid phase. This is followed by slowly cooling, so that all particles will arrange in the lowest energy state. The target function consists of an empirical potential energy which is described by the stereochemistry and nonbonding interactions in the macromolecule.

Refinement was followed by the addition of water molecules and resolution extension to 2.3Å. The final refinement parameters are listed in Table 2.5. The final model consists of 19,323 non hydrogen atoms with two disordered regions between residues 361 to 373 and 414 to 429 in all four molecules. The structure also includes 1,142 water molecules that represent ordered water molecules that form part of the lattice. An example of the final $2F_o$ - F_c electron density map is shown in Figure 2.4.

Figure 2.5 presents the Ramachandran plot of the BE structure. A Ramachandran diagram is a plot of ϕ (angle between N and Cα) versus φ (angle between C and Cα). According to geometry and steric restrictions the ϕ and φ angles must be within certain values, which are marked by the different shadows of gray in Figure 2.5. All residues, with the exception of glycines and prolines, that don't have a side chain, must lie within these allowed regions. In the BE structure only 5 residues out of the 505 non glycine residues lie in disallowed regions. This corresponds to only 1% of the structure a very small fraction for a structure of this magnitude. Among these residues M472 and L475 lie in a sequence rich in glycines which could explain their troubled geometry.

Table 2.5 Refinement statistics of BE

R_{factor}	20.00%
R_{free}	26.53%
Resolution	2.3 Å
rmsd bonds	0.0075 Å
rmsd angles	1.45°

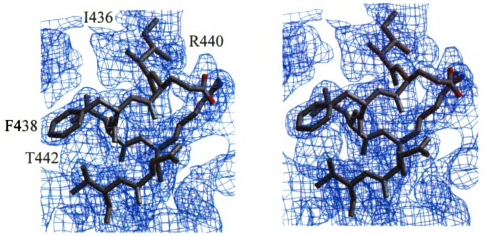


Figure 2.4 Stereoview example of the final $2F_o - F_c$ electron density map.

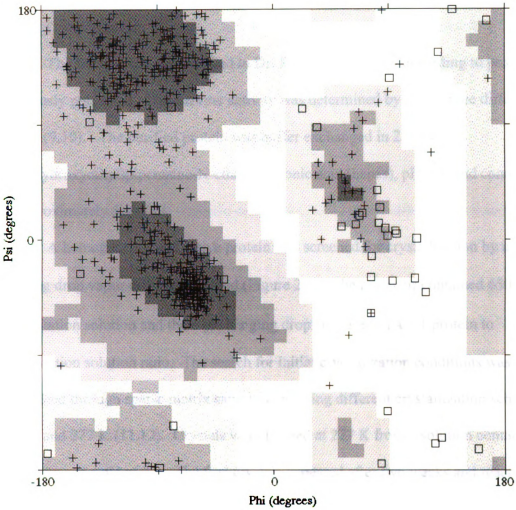


Figure 2.5 Ramachandran plot of BE. Showing one of the four molecules for clarity.

2.1.4 Materials and methods

2.1.4.1 Crystallization

The BE enzyme was purified in Dr. Preiss's laboratory according to protocols previously published. The enzyme activity was determined by using three different assays (9,10). The purified protein was buffer exchanged in 25mM N-(2-hydroxyethyl)piperazine-N-ethane sulfonic acid (hepes), pH 7.5, and concentrated to approximately 5 mg/ml.

A homogeneous and active protein was screened for crystallization by using the hanging drop vapor diffusion method (Figure 2.1). The reservoir contained 650 μL of the precipitation solution and the 4 μL hanging drop consisted of a 1:1 protein to precipitation solution ratio. The search for initial crystallization conditions was performed through sparse matrix sampling by using different crystallization screens at 298 K and 277 K (11,12). Crystals were formed at 277 K from a solution containing 100 mM Hepes at a pH of 7.2. The first crystals appeared after two weeks and reached a maximum size of 0.3 x 0.1 x 0.1 mm^3 in four weeks (Figure 2.2).

2.1.4.2 Native data collection

The crystals were transferred to a cryoprotectant solution containing 25% (v/v) 2-Methyl-2,4-pentanediol (MPD), 2% (m/v) polyethylene glycol (PEG) 4,000 and 100 mM Hepes, pH 7.5. The crystals were then mounted in nylon cryo-loops and flash frozen by immersion in liquid nitrogen. A high-resolution native data set was collected at the Advanced Photon Source (APS) at The Argonne National Laboratories (Argonne, IL) on the Structural Biology Center ID-19 (SBC) beamline (Tables 2.1 and 2.2). Intensity data

were collected by using a 3 x 3 array (3072 x 3072 pixels) custom made CCD area detector to a resolution of 2.3 Å. The crystal to detector distance was set to 220 mm, and 160° of data was collected with an oscillation angle of 0.5°. Diffraction data was processed by using the HKL2000 program package (13,14).

2.1.4.3 SIR and SAD data collections

An SIR experiment was also carried out, in which a native crystal was soaked for 18 hours in a solution containing 10% MPD, 0.1M Hepes pH 7.20 with 10 μM of p-chloro mercuri benzoic acid. The crystals were cryoprotected and frozen, as previously described. Data were collected over 160° with oscillations of 1°. A total of 118,955 reflections were measured at our home source by using a Rigaku R-AXIS IV^{++} image plate detector (Table 2.2). Cu Kα radiation was generated by a Rigaku RU-200 rotating anode source operating at 50 kV and 90 mA.

In addition to the SIR data, a SAD experiment was also performed. The Se-Met substituted protein was crystallized and cryoprotected with the same conditions as the native crystals and a SAD experiment was performed at the selenium absorption edge. Anomalous data, to a resolution of 2.5 Å, were collected in a single element 165 mm MAR CCD detector from beamline 17-ID in the facilities of the Industrial Macromolecular Crystallography Association Collaborative Access Team (IMCA-CAT) at the APS. The crystal to detector distance was set to 190 mm, and 180° of data was collected with a 0.5° oscillation (Table 2.2).

The automated heavy atom search routines from the program SOLVE were used to find the four mercury and 27 initial Se positions (5). The program SHARP was used to

calculate an initial set of phases and find a total of 61 Se sites (Table 2.4) (6). An improved electron density map was calculated with the aid of four fold non-crystallographic symmetry averaging using the Dork program package graciously provided by the author, Dr. Greg Van Dyne, University of Pennsylvania. All model building was done using TURBO FRODO (15) (16) and refinement and map calculations were carried out using CNS (Table 2.5) (17).

2.2 Angiostatin

2.2.1 Crystallization and data collection

EntreMed Inc., (Rockville, MD) provided the human angiostatin protein containing Kg1-3. EntreMed Inc., has the patent on angiostatin and is running its clinical trials. This protein was crystallized (Figure 2.6) and a complete data set to a resolution of 1.75 Å was collected. The crystals are tetragonal and belong to the $P4_12_12$ space group with unit cell parameters a = b = 56.94 Å and c = 192.97 Å. Assuming one molecule of angiostatin (29.77 kDa) per asymmetric unit, the crystal volume per protein mass is 2.65, which corresponds to approximately 51% solvent content in the crystal. This value is within the range observed for protein crystals (3). The crystal parameters of the angiostatin crystal are listed in Table 2.6. A synchrotron X-ray diffraction data set to a resolution of 1.75 Å, with an overall I/σ of 19.5, was obtained. The X-ray diffraction data was 92.8% complete with an Rmerge of 7% for 30,370 unique reflections from a total of 217,983 measured reflections. Detailed data collection statistics are found in Table 2.7.

2.2.2 Molecular replacement and structure refinement

The structure was solved by molecular replacement using the structures of Kg1 and Kg2 as models (PDB id 1CEA and 1I5K) (18,19). The molecular replacement method uses phases from a known protein structure as an initial estimate of the phases of the new protein. The challenge of this method lies in finding the correct orientation and position of the model in the unit cell of the new protein. Although unit cell dimensions

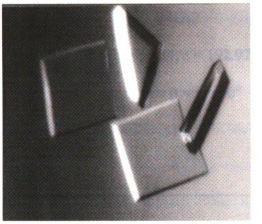


Figure 2.6 Tetragonal crystals of human angiostatin (Kg1-3). The square plate crystals have dimensions of 0.7 x 0.7 x 0.4 mm^3.

Table 2.6 Crystal parameters for the angiostatin crystal

Crystal form	Tetragonal
Space group	$P4_12_12$
Unit cell	a = b = 56.94, c = 192.97 Å
	and $\alpha = \beta = \gamma = 90^o$
Solvent content	51%
Molecules per asymmetric unit	1

Table 2.7 Statistics for the angiostatin X-ray diffraction data collection

Wavelength, Å	1.0332
Resolution range, Å	50 – 1.75
(last resolution shell)	(1.81 – 1.75)
Completeness, %	92.8
(last resolution shell)	(83.0)
I/σ	19.5
(last resolution shell)	(5.1)
R_{merged}, %	7.0
(last resolution shell)	(34.1)
Unique reflections	30,370
Measured reflections	217,983

and symmetry place some constraints, the process is still complicated. For this reason it is divided in two steps; calculation of a rotation function that is followed by a translation function. The rotation function uses Patterson maps to determine the correct orientation of the model. As was previously discussed, a Patterson map is independent of the position of the structure in the unit cell as long as the orientation does not change. The solution process involves the calculation and evaluation of Patterson maps. This is performed by the automated Patterson search routine of the program AMoRe (20). Once the correct orientation of the model has been found the correct position can be determined using the translation function. The translation search is done systematically over the unit cell and evaluated based on the agreement of the calculated structure factors with observed ones.

The structure factors of a properly positioned model are then calculated and a correlation coefficient (C) and an R value will determine if a correct solution have been found. These two values are defined below, where Fobs and Fcalc are the observed and calculated structure factors.

$$R = \frac{\sum_{hkl} \left\| F_{obs} \right| - \left| F_{calc} \right\|}{\sum_{hkl} \left| F_{obs} \right|} \times 100$$

$$C = \frac{\sum_{hkl} \left(\left| F_{obs} \right|^2 - \overline{\left| F_{obs} \right|^2} \right) \left(\left| F_{calc} \right|^2 - \overline{\left| F_{calc} \right|^2} \right)}{\left[\sum_{hkl} \left(\left| F_{obs} \right|^2 - \overline{\left| F_{obs} \right|^2} \right)^2 \sum_{hkl} \left(\left| F_{calc} \right|^2 - \overline{\left| F_{calc} \right|^2} \right)^2 \right]^{1/2}}$$

The human plasminogen Kg1 and Kg2 structures were used as the protein models (18,19). Both models gave rotation solutions that were later used in the translation function. A translation search with Kg2 gave two solutions; a search with Kg1 also gave two solutions, one of which was unique, relative to the initial Kg2 search. Examination of the packing of the Kg2 solutions showed them to be disulfide linked Kg2 and Kg3. These solutions had a correlation factor of 32.6 and 30.8 and an R of 49.1% and 49.3%, respectively. Fixing the positions of Kg2 and Kg3 and calculating an electron density map revealed density corresponding to the unique Kg1 solution, indicating it to be Kg1. This was corroborated with a new translation search that also produced the unique solution that corresponded to Kg1.

This multi kringle angiostatin structure contains the first structure of Kg3 to be determined. The high homology among Kgs (50% average identity) allowed use of Kg1 to approximate Kg3 in the electron density. Residues different from Kg1 were mutated to alanine and an electron density map of Kg1, Kg2 and Kg3 was calculated by using the program CNS (17). Analysis of the map allowed replacement of the previously mutated residues to the corresponding amino acids in Kg3. The interkringle peptides connecting Kg1 to Kg2 and Kg2 to Kg3 were also built.

Multiple rounds of structure refinement using the simulated annealing method followed by the addition of water molecules and resolution extension resulted in the final refinement parameters listed in Table 2.8. The final model includes 253 residues (from amino acid 81-333) and 398 water molecules. In addition to the 398 water molecules, electron density for a bicine molecule, the buffer used in crystallization, was located in

each of the three Kg LBS's. The Ramachandran plot of the angiostatin structure shows no residues in disallowed regions (Figure 2.7). An example of the final $2F_o$ - F_c electron density map is shown in Figure 2.8.

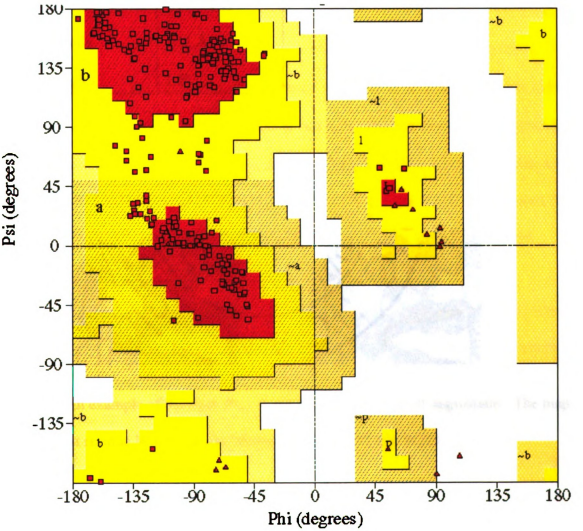


Figure 2.7 Ramachandran plot of angiostatin

Table 2.8 Refinement statistics of angiostatin

R-factor	19.58%
R_{free}	26.25%
Resolution	8.0-1.75 Å
rmsd Bonds	.011 Å
rmsd Angles	1.6°

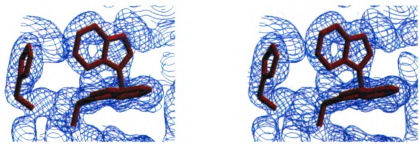


Figure 2.8 An example of the final $2F_o$ - F_c electron density map of angiostatin. The map is centered at residue W315 and also shows residues H317 and W325.

2.2.3 Materials and Methods

The human angiostatin mutant N289E (this mutant lacks N-linked glycosylation) containing Kg1-3 was expressed in *Pichia pastoris* and purified as previously described (21). The purified protein, provided by the company EntreMed Inc. Rockville, MD, was buffer exchanged into saline buffer (0.15 M NaCl) and concentrated to 15 mg/ml.

The protein was extensively screened for crystallization by using the hanging drop vapor diffusion method (Figure 2.1). The search for well diffracting crystals was performed using several sparse crystallization screens at two different temperatures, 298 K and 277 K (11,12). A 4 μL hanging drop in a 1:3 protein to precipitant solution ratio was equilibrated against 650 μL of precipitant solution. The best crystals were obtained at 277 K, with a precipitant solution containing 10% PEG 20,000, 2% (v/v) dioxane and 100 mM N,N Bis(2hydroxyethyl) glycine (bicine) buffer at a pH of 9.0. The crystals first appear overnight and grow to a maximum size of 0.7 x 0.7 x 0.4 mm^3 in three days (Figure 2.6).

The crystals were transferred to a cryoprotectant solution containing 35% (v/v) glycerol, 10% PEG 20,000, 2% dioxane and 100 mM bicine pH of 9.0 and flash frozen by immersion in liquid nitrogen. X-ray diffraction data to a resolution of 1.75 Å was collected at the SBC beamline at the APS. Data were collected by using a custom built 3 x 3 array (3072 x 3072 pixels) CCD area detector. The crystal to detector distance was 150 mm and 120° of data was collected with an oscillation of 0.3°. Diffraction data was processed using HKL2000 (14).

The structure was solved by molecular replacement using the program AMoRe (20). All model building was done by using TURBO FRODO (15,16) and refinement and map calculations were carried out by using CNS (Table 2.8) (17).

2.3 Literature cited

1. Guan, H., Li, P., Imparl-Radosevich, J., Preiss, J., and Keeling, P. (1997) *Arch. Biochem. Biophys.* **342,** 92-100.

2. Binderup, K., Mikkelsen, R., and Preiss, J. (2000) *Arch Biochem Biophys* **377**(2), 366-371.

3. Mathews, B. W. (1968) *J. Mol. Biol* **33,** 491-497.

4. Blundell, T. L., and Johnson, L. N. (1976) in *Mol. Biol., Int. series of books and monographs*, pp. 183-239, Academic Press, New York

5. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Cryst.* **D55,** 849-861.

6. La Fortelle, E., and Bricogne, G. (1997) *Methods Enzymol* **276,** 472-494.

7. Kirkpatrick, S., Gelatt, C., and Vecchi, M. (1983) *Science* **220,** 671-680.

8. Brunger, A. T., Kuriyan, J., and Karplus, M. (1987) *Science* **235,** 458-460.

9. Binderup, K., and Preiss, J. (1998) *Biochemistry* **37**(25), 9033-9037.

10. Guan, H. P., and Preiss, J. (1993) *Plant Phys.* **102,** 1269-1273.

11. Cudney, R., Patel, S., Wesisgraber, K., Newhouse, Y., and McPherson, A. (1994) *Acta Cryst.* **D50,** 414-423.

12. Jancarik, J., and Kim, S. H. (1991) *J. Appl. Cryst.* **24,** 409-411.

13. Otwinowski, Z. (1993) in *Data Collection and Processing* (Sawyer, L., Issacs, N., and Bailey, S., eds), pp. 56-62, SERC Daresbury Laboratory, Daresbury, U.K.

14. Otwinowski, Z. M., W. (1997) *Methods Enzymol* **276,** 307-326.

15. Jones, T. A. (1985) *Methods Enzymol.* **115,** 157-171.

16. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Cryst.* **A47,** 110-119.

17. Brunger, A. T. (1992) *X-PLOR, version 3.1, a System for X-ray Crystallography and NMR.*, Yale University Press, New Haven, CT

18. Mathews, I., Vanderhoff-Hanaver, P., Castellino, F. J., and Tulinsky, A. (1996) *Biochemistry* **35(8),** 2567-2576.

19. Rios-Steiner, J. L., Schenone, M., Mochalkin, I., Tulinsky, A., and Castellino, F. J. (2001) *J Mol Biol* **308**(4)**,** 705-719.

20. Navaza, J. (1994) *Acta Cryst* **A50,** 157-163.

21. Shepard, S. R., Boucher, R., Johnston, J., Boerner, R., Koch, G., Madsen, J. W., Grella, D., Sim, B. K. L., and Schrimsher, J. L. (2000) *Prot. Exp. Purif.* **20**(2)**,** 216-227.

CHAPTER III: THE THREE DIMENSIONAL STRUCTURE OF BRANCHING ENZYME

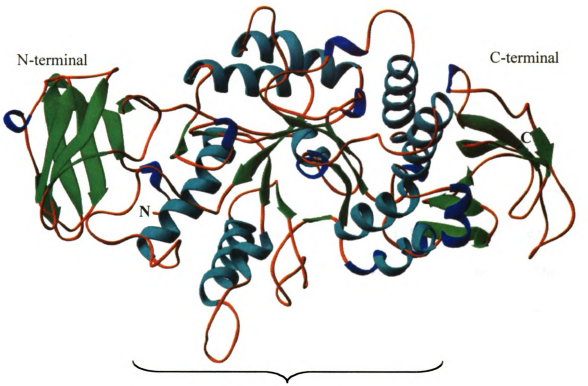
This chapter presents the three dimensional structure of BE from *e. coli*. This is the first structure of any enzyme involved in glycogen biosynthesis and the only member of the amylotic family of enzymes of which the structure was still unknown. Analysis of the structure provided valuable information that allowed us to propose a mechanism for BE catalytic action. This structure and the information obtained from it represent one of the most important pieces of information obtained in the field of glycogen biosynthesis.

3.1Overall structure

The enzyme used in the structure presents a truncation at the amino terminal lacking the first 112 residues. It portrays an altered branching pattern when compared to WT BE. This truncated enzyme has a higher propensity for transferring glucose chains of 12 units while the WT BE has a higher propensity for transferring branches of 6 glucose units (1).

The four molecules in BE asymmetric unit consists of 19,323 residues and 1,142 water molecules. Each molecule extends from residue 117 to residue 728, with the first four residues disordered. There are two disordered regions between residues 361 to 373 and 414 to 429 in all four molecules. The overall elliptical structure of BE has dimensions of 87.7 Å by 42.6 Å and 42.0 Å in depth.

The structure of BE is organized into three domains: the C-terminal, N-terminal and (α/β) barrel (Figure 3.1). The C-terminal domain consists of 116 residues organized



Resolution = 2.3 Å
Rfactor = 20.00%
Rfree = 26.53%

Figure 3.1 Three dimensional structure of *e. coli* BE truncated at the amino terminus at amino acid 113.

in seven β strands. The N-terminal domain contains a β sandwich fold, which was previously predicted by sequence analysis (2). This N-terminal domain is composed of 128 residues arranged in seven β strands. The central (α/β) barrel domain common in members of the α-amylase family of enzymes extends from residue 241 to residue 612 comprising a total of 372 residues. This domain contains the residues involved in catalysis and presents a substrate binding cavity with dimensions of 30.5 Å x 17.7 Å x 17.7 Å big enough to accommodate branched glucose chains.

A complete (α/β) barrel should contain eight α-helices and eight β-strands. The (α/β) barrel domain in BE is missing α5, the α-helix between β sheet number five (β5) and six (β6). This barrel also has three extra helices inserted; α1a, α6a and α7a. The α1a helix located between β1 and α1 and the α7a located between β7 and α7 are both one turn helices. The α6a helix positioned between α6 and β7 is a three turn helix. This variation of the (α/β) barrel domain is also observed in isoamylase but not in other members of the family. There are two loops connecting the domains in BE. The loop that connects the N-terminal domain to the (α/β) barrel is eighteen residues long (223-240), and the loop joining the end of the (α/β) barrel to the C-terminal domain is thirteen residues long (613-625). The organization of the elements of secondary structure is depicted in Figure 3.2 and summarized in the amino acid sequence diagram of Figure 3.3.

There are four molecules in the asymmetric unit of BE, which are oriented as shown in Figure 3.4. There is a two fold rotation that relates molecule A to C, and B to D, but all four molecules are not related by a perfect four fold. Instead the two folds are coupled with a translation that relates both two folds.

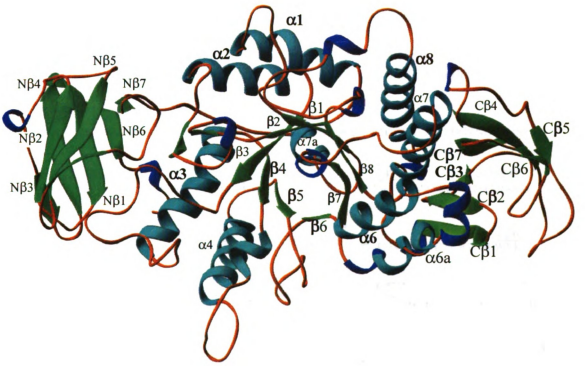


Figure 3.2 The elements of secondary structure in the three domains of BE. The β sheets from the N and C terminals are identified with an N and a C, respectively.

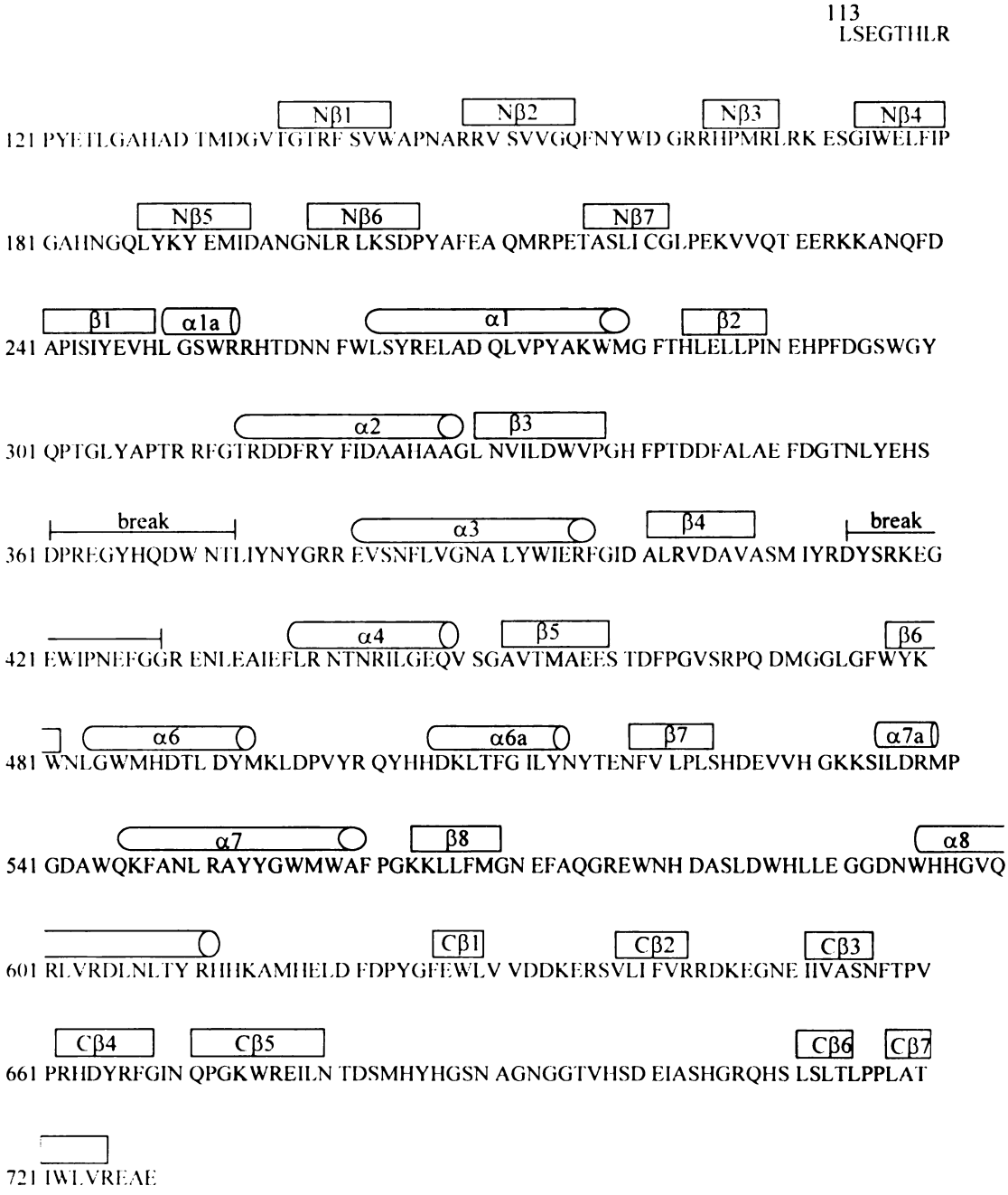


Figure 3.3 The amino acid sequence of the truncated BE with elements of secondary structure.

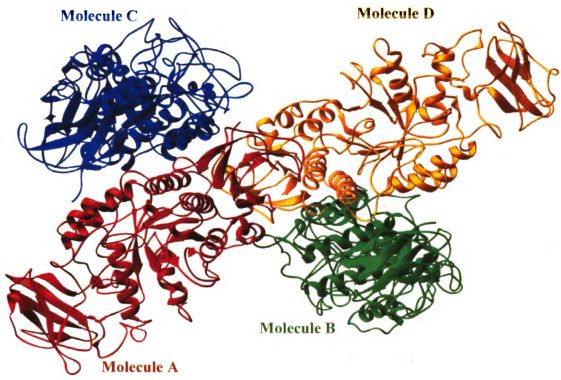


Figure 3.4 There are four molecules in the BE asymmetric unit.

3.2 Structural differences among members of the α-amylase family

BE belongs to the α-amylase family of enzymes, which includes α-amylases, CGTs, isoamylases and BEs. Even though members of this family share some structural features, like the (α/β) barrel domain and the conserved catalytic residues, each enzyme performs different reactions (Figure 3.5). α-Amylase and isoamylase hydrolyze α-1,4 and α-1,6 glucosidic bonds, respectively. Branching enzyme and CGT catalyze transglycosylation reactions with BE being the only one with specificity for two different glucosidic bonds, α-1,4 and α-1,6. CGT catalyzes the formation of cyclodextrins by cleavage and subsequent transglycosylation of α-1,4 links. The similarity among the primary structure of members of the α-amylase family is very low, preserved only among the four regions that contain the conserved residues.

X-ray structures of members of the family showed the existence of a conserved barrel domain (Figure 3.6) (3) (4,5). BE was the only member of the α-amylase family for which the three dimensional structure was still unknown. The structure of BE revealed that indeed BE has a central (α/β) barrel that contains the catalytic residues (Figure 3.1). This domain is similar to the (α/β) barrels of other members of the family (Figure 3.6). Also, the C-terminal domain of BE is structurally similar to the C-terminals of isoamylase and α-amylase while the N-terminal β sandwich domain of BE is analogous to the N-terminal domain in isoamylase (4,5). BE and isoamylase are therefore the more structurally similar members of the α-amylase family. A comparison between domains is presented in Figure 3.7 where each domain in isoamylase has been rotated and positioned in the same orientation as is BE. Isoamylase and BE are the only

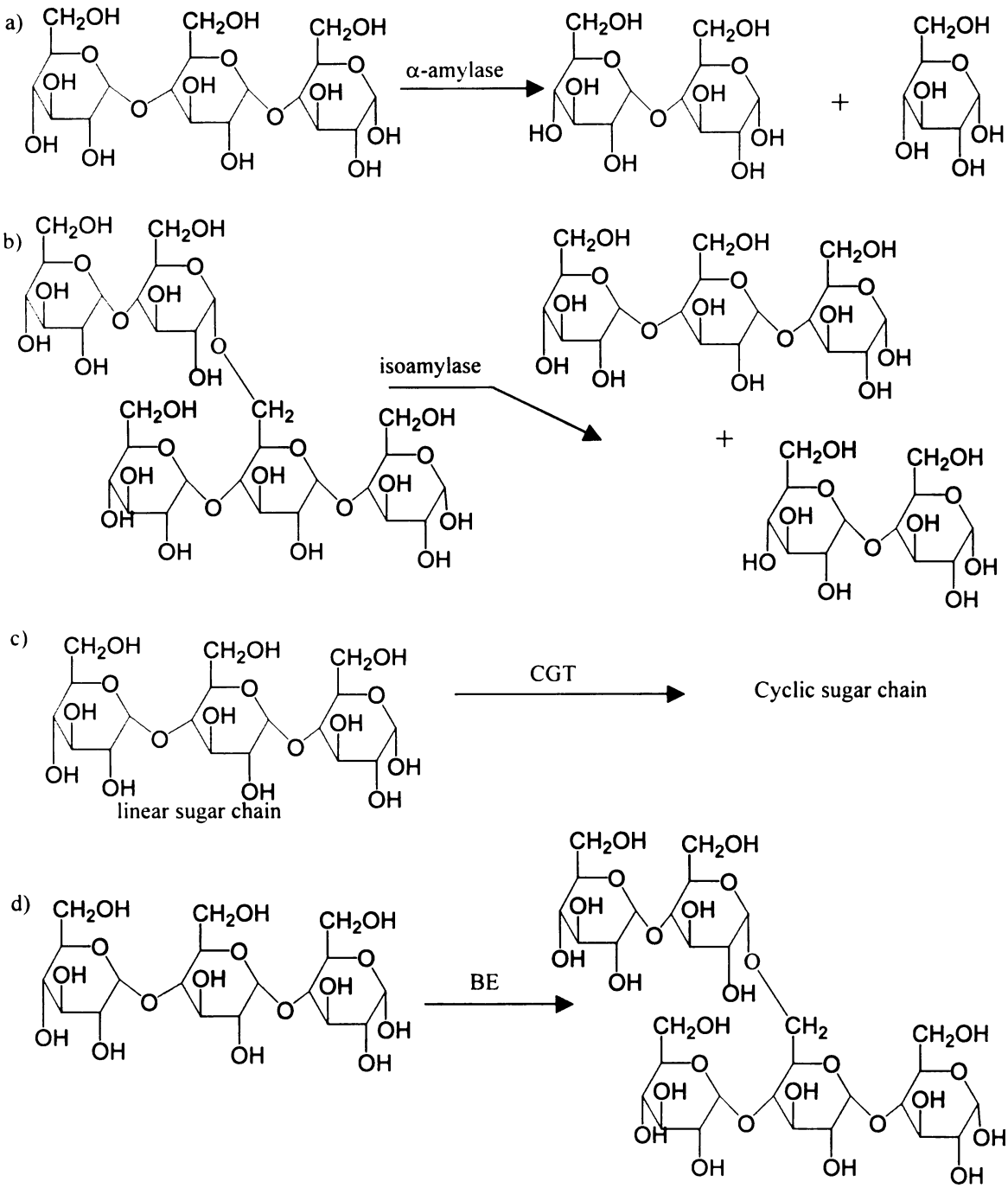


Figure 3.5 The reactions catalyzed by the members of the α-amylase family of enzymes.

a) α-amylase hydrolyses α-1,4 bonds. b) isoamylase cleaves α-1,6 bonds. c) CGT

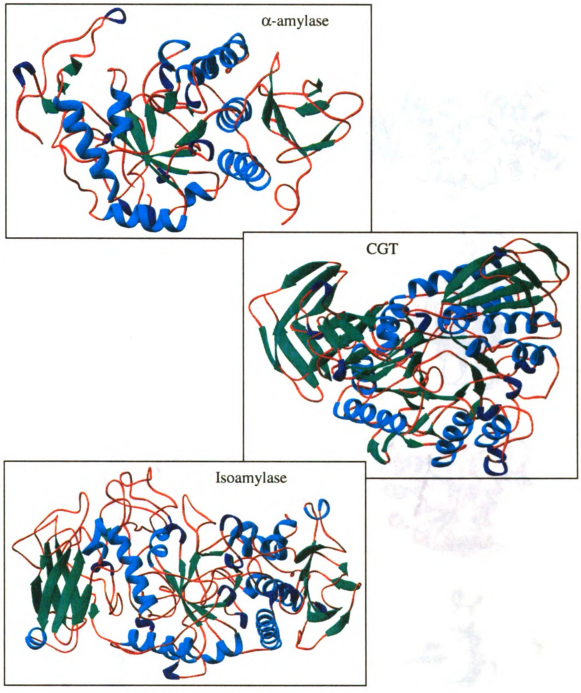


Figure 3.6 X-ray structures of members of the α-amylase family of enzymes (3) (4,5).

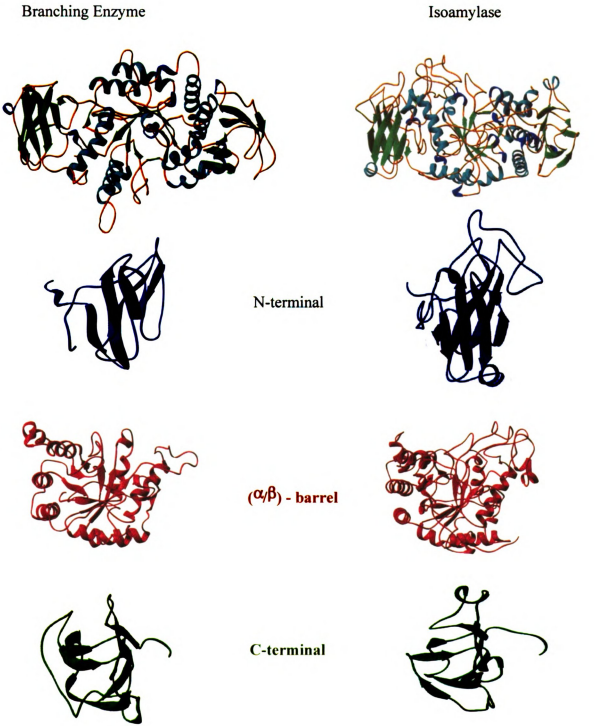


Figure 3.7 Comparison between domains of isoamylase and BE (5). The domains in isoamylase have been rotated to match the orientation of BE.

members in the family catalyzes the formation of cyclodextrins and d) BE catalyzes the formation of α-1,6 branches.that bind sugars in the α-1,6 position. In addition, Figure 3.8 shows a superposition of isoamylase (Figure 3.8a) and α-amylase (Figure 3.8b) onto BE (rmsd ~1Å). It can be observed in the figure that although the positions of the elements of secondary structure lie in proximity to one another, their relative orientation is quite different. Also, the loops connecting those secondary structure elements are different among the structures.

There are marked differences between the loops connecting elements of secondary structure among members of the α-amylase family of enzymes. These loops might be responsible for the distinct catalytic properties between the enzymes. Comparison of the loop structures between members of the family revealed that BE has shorter loops, presenting a more open cavity for the binding of a bulkier sugar as is the case of a branched sugar. Overall BE has a more accessible cavity for sugar binding compared to CGT, isoamylase and α-amylase (Figure 3.9). A model of the α-1,4 cleaved sugar and the incoming sugar oriented to form the branch point was modeled, as shown in Figure 3.9a. When the structures of α-amylase, CGT and isoamylase are overlaid onto BE the modeled sugar oriented to form the branch point run into the extended loops. The incoming α-1,6 sugar collides with the β7/α7 loop of isoamylase and the β5/β6 loops of CGT and α-amylase. It is important to remember that BE not only binds already branched sugars but it also requires access to sugars properly oriented for forming the α-1,6 links.

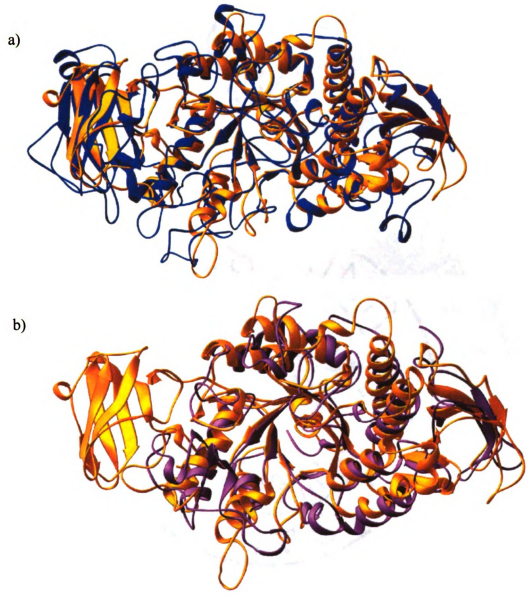


Figure 3.8 a) Superposition of the structure of isoamylase in blue onto BE shown in gold (5). b) Superposition of the structure of α-amylase depicted in lavender onto *e. coli* BE.

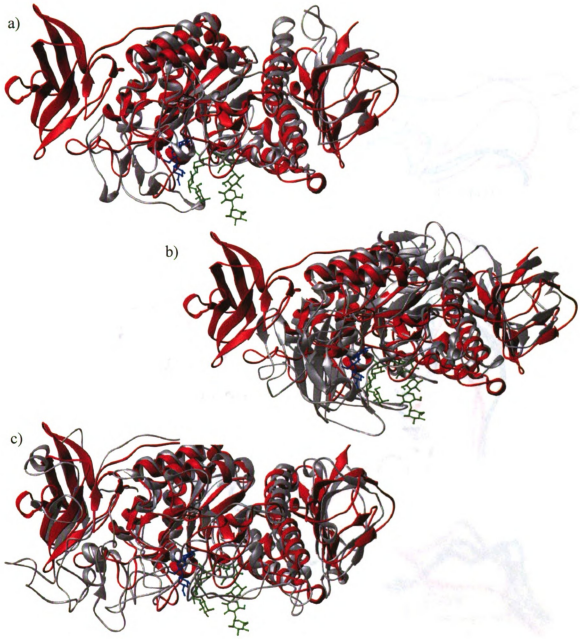


Figure 3.9 The structures of a) α-amylase, b) CGT and c) isoamylase are overlaid onto BE also showing are the α-1,4 cleaved sugar and the incoming sugar oriented to form the branch point. BE is shown in red and α-amylase, CGT and isoamylase in gray. This mimic was based on the substrate and intermediate bound structures of other members of the α-amylase family taking into account the unique loop structure of BE.

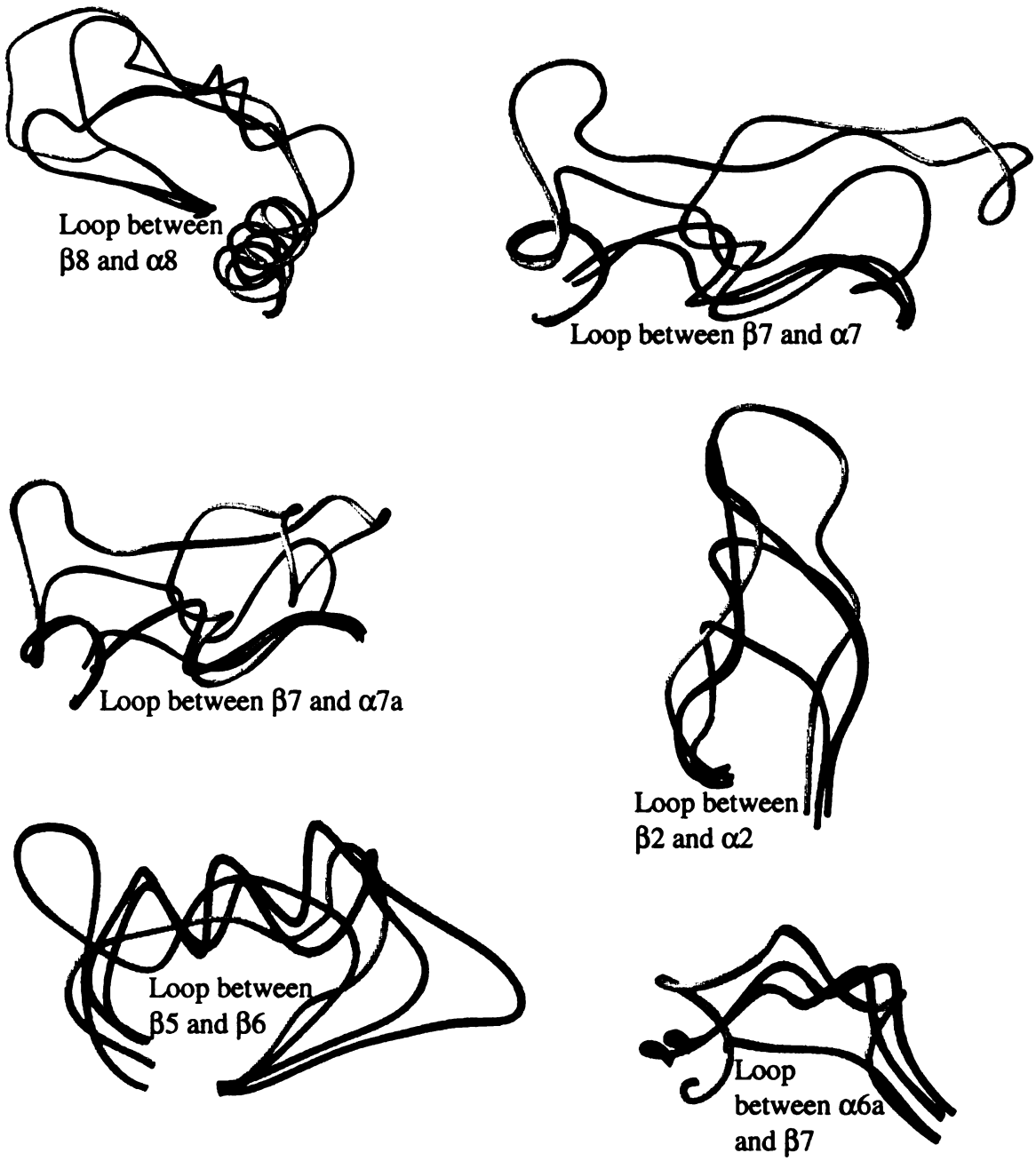


Figure 3.10 Comparison of the loops that surround the (α/β) barrel cavity. BE is shown in red, isoamylase in lavender, CGT in green and α-amylase in blue.

There are six loop structures observed in α-amylase, CGT and isoamylase not present in BE that block the access of incoming sugars. These loops lie between β8/α8, β7/α7, β7/α7a, β2/α2, β3/α3, β5/β6 and α6a/β7. The loop between β5 and β6 in isoamylase is moved away from the sugar binding channel. A comparison of all these loops can be seen in Figure 3.10. There is also an extra domain of approximately sixty residues, named domain B, inserted in the barrel between β3 and α3 (Figure 3.11). This domain is present in α-amylase and CGT but not in BE. Isoamylase depicts a loop extension consisting of as many residues as α-amylase and CGT, but topologically it can not be considered a domain because it lacks elements of secondary structure. Although in the BE structure there are 13 disordered residues in the loop between β3 and α3, this loop is only 40 residues long, not long enough to account for the whole B domain.

Based on the structures of α-amylases and primary structure analysis, seven conserved residues among the amylotic enzymes were defined. These residues are D335, H340, R403, D405, E458, H525 and D526 (Figure 3.12). It was established early on that these conserved residues are involved in catalysis and substrate binding. Upon analysis of the BE active site and after comparison with other members of the family we noticed that although D335 is conserved it does not make any contacts with the substrates in any of the bound structures. We also noticed that Y300 is not only conserved but it also interacts with the modeled substrate and, as will be discussed later, has a crucial role in the mechanism of BE. Figure 3.12 shows the conserved residues and their orientation in the barrel.

The structures of BE, isoamylase, α-amylase and CGT were overlaid and the conserved residues were compared (Figure 3.13) (3) (4,5). A comparison between

a)

b)

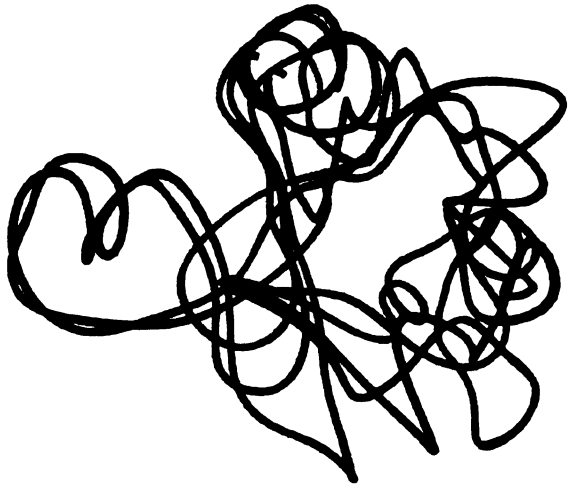


Figure 3.11 The B domain lies between β3 and α3. a) The loop between β3 and α3 in BE is shown in red b) A comparison between the B domain of α-amylase in blue, isoamylase in lavender, CGT in green and BE in red.

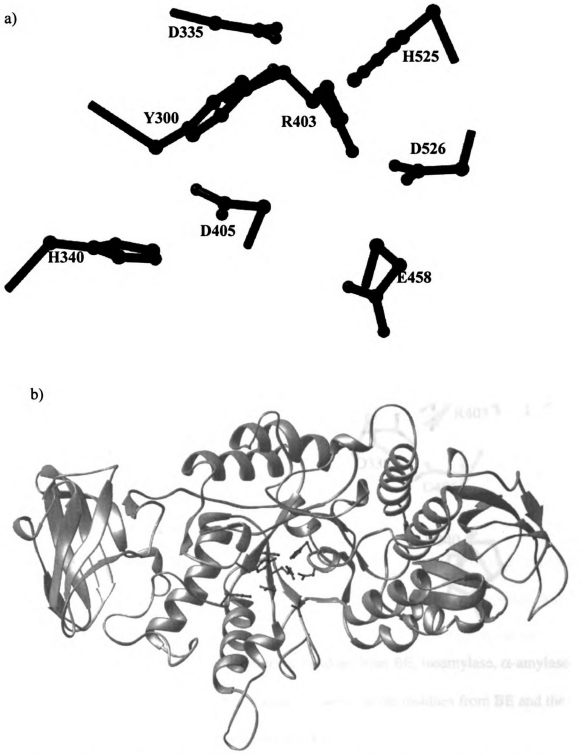


Figure 3.12 a) Residues involved in BE catalysis. b) Position of these residues in the barrel

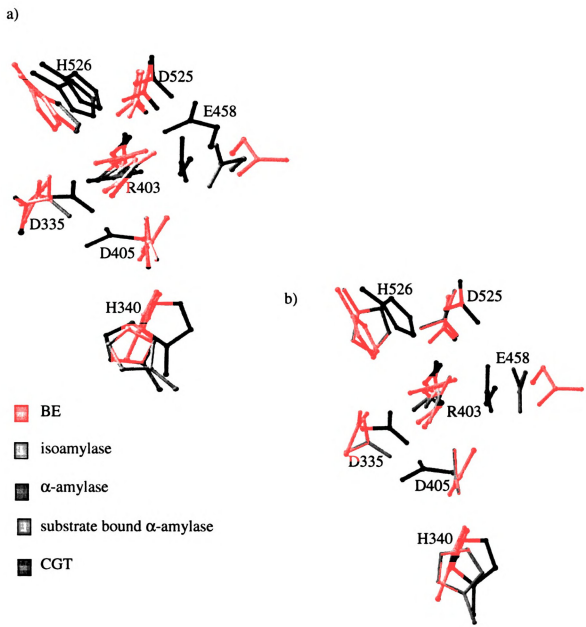


Figure 3.13 a) Superposition of the conserved residues from BE, isoamylase, α-amylase and CGT not bound to substrate. b) Comparison between the residues from BE and the ones from α-amylase apo and substrate bound (4-6).

structures not bound to substrate is shown in Figure 3.13a. The position and orientation of residues R403, H525 and D526 are conserved in all structures. Inspection of residues E458, D405 and H340 show a drastic motion of their side chain in BE. We also compared those residues in BE with apo and substrate bound α-amylase. Interestingly, the position of those three residues does not change in both α-amylase structures after the substrate binds (Figure 3.13b). The change in position of these residues is only observed in BE. This may be important for the type of reaction that BE catalyzes. We must await further studies of substrate bound BE to understand this difference in orientation.

3.3 Residues associated with the GSDIV

Mutations in BE are responsible for the GSDIV, a genetic disease that produces an inactive BE causing the glycogen to precipitate in the cell (7,8). With few exceptions, GSDIV is a progressive and lethal disease (9,10). The residues involved in these mutations are V273, Y306, Y377, G555 and K546 (*e. coli* numbering) depicted in Figure 3.14. Table 3.1 presents a summary of the residues in human and *e. coli* numbering, the respective mutation and their location in the structure. These residues are conserved in *e.coli* with the exception of R524 in humans, which is a glycine in *e.coli*. None of these residues are located within the catalytic cavity, but rather are spread throughout the structure. K546 is located at the beginning of helix α7 and an although exposed it makes a salt bridge with the carbonyl of a proline in the turn of the helix holding the loop in position. The two tyrosines, Y306 and Y377, hold the loops connecting β2 to α2 and β3 to α3, respectively. The valine is located in α1, interacting with residues in α2 and holding both helices next to each other. In conclusion, these mutations, in addition to the

deletions and truncations, are likely to cause the unfolding of the protein, inactivating the enzyme.

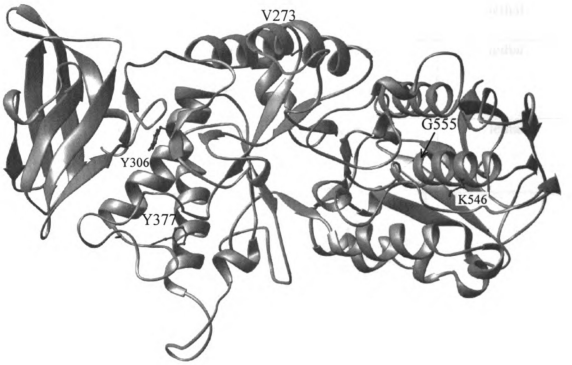


Figure 3.14 Residues responsible for causing the GSDIV (7,8) (9,10).

Table 3.1 Residues responsible for causing the GSDIV their location and effect (7-10).

homo sapiens	*e. coli*	mutation	Location	Effect
L224	V273	P	α1	slow progressive
F257	Y306	L	β2/α2	lethal
Y329	Y377	S	β3/α3	slow progressive
R515	K546	C	α7	lethal
R524	G555	N or truncation	α6a/β7 α7a to C-terminal	lethal
residues 262-331	residues 311-379	deletion	α1a to α2	lethal

3.4 Proposed mechanism

Based on the proposed mechanism for CGT, the only member of the α-amylase family besides BE that catalyzes a transglycosylation reaction, we have been able to propose a mechanism for BE. The CGT mechanism was deduced from the analysis of the substrate and intermediate bound X-ray structures (6). The BE structure was superimposed on both CGT structures by overlaying the conserved catalytic residues (rmsd ~1.1 Å). The position and interactions observed when modeling the substrate and intermediate presented a possible model for BE catalytic action. The proposed mechanism for the reaction catalyzed by BE is shown in Figure 3.15. As previously described, BE performs a transglycosylation reaction in which an α-1,4 bond is cleaved and an α-1,6 bond is subsequently formed (Figure 3.5d).

Before substrate binding, E458 and D526 are held in position by a hydrogen bond network between the conserved waters 1017 and 809 and residue Y300 (Figure 3.16a). Once the substrate binds in the cavity it pushes the water molecules away. The carbonyl side chain of E458 rotates and is now oriented properly for interaction with the glycosidic oxygen. The side chain of D526 also rotates to be able to interact with the oxygens in the sugar. The sugar protein interactions are shown in Figure 3.16b and listed in Table 3.2. The hydrolysis of the α-1,4 bond is then initiated by the protonation of the glycosidic oxygen by E458 acting as the proton donor to form an oxocarbenium ion.
This is followed by the nucleophilic attack of D405 to the C1 of the sugar, forming a covalent bond in the intermediate formation (Figures 3.15 and 3.16c). This covalent intermediate has been observed by X-ray crystallography in CGT and by NMR in α-amylase (6) (11). Once the sugar is cleaved it diffuses away and a new polysaccharide

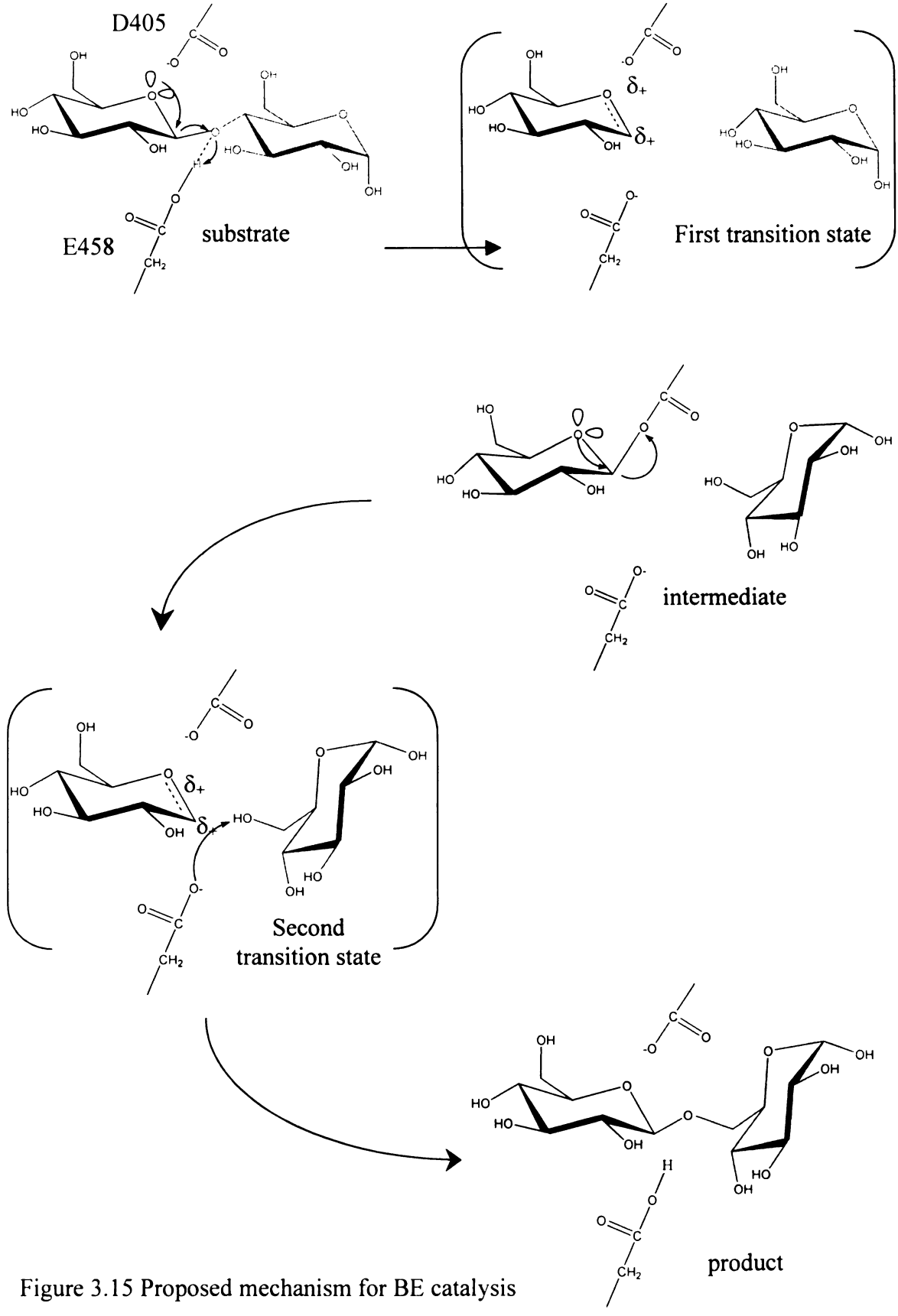


Figure 3.15 Proposed mechanism for BE catalysis

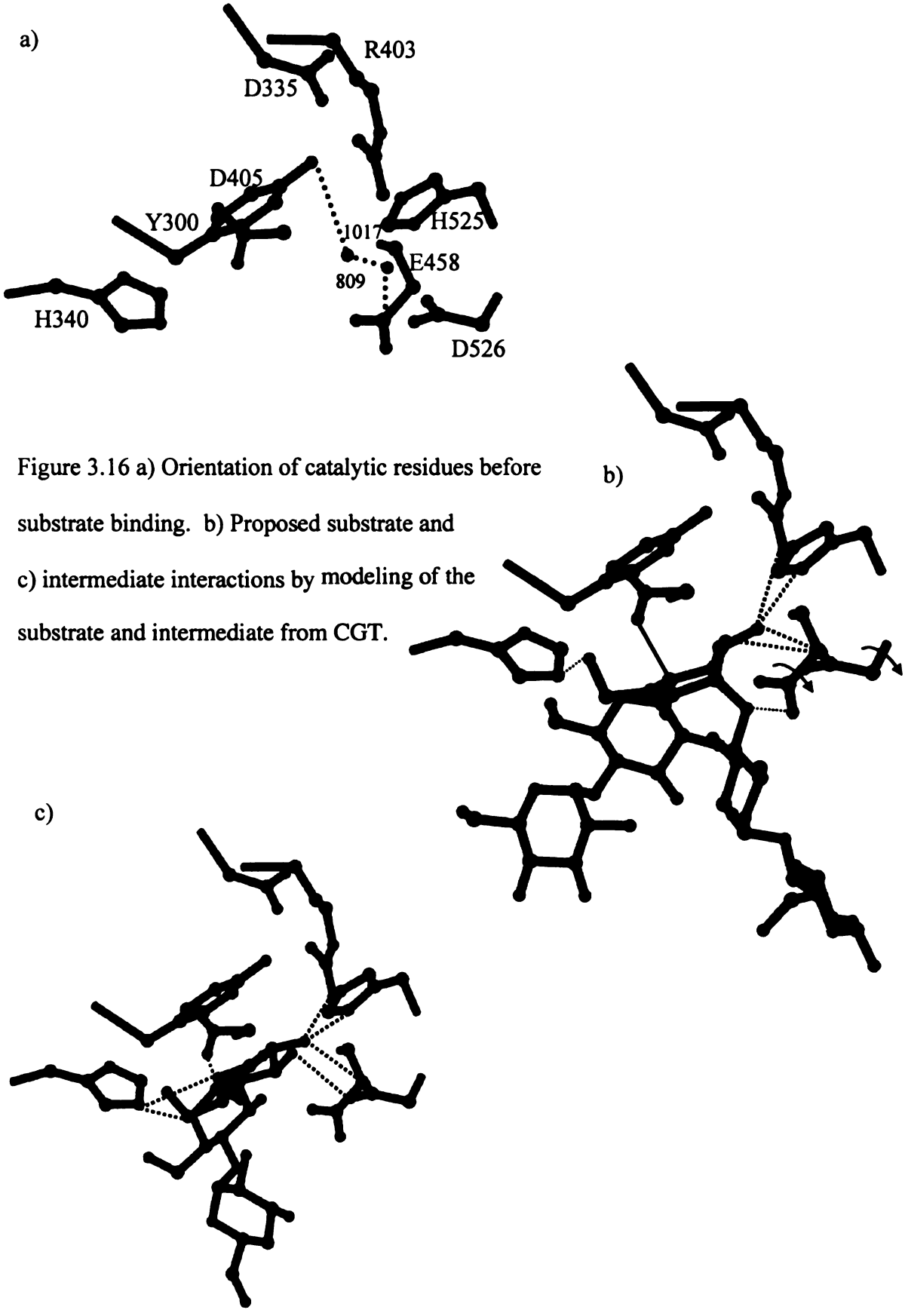


Figure 3.16 a) Orientation of catalytic residues before substrate binding. b) Proposed substrate and c) intermediate interactions by modeling of the substrate and intermediate from CGT.

Table 3.2 Protein interactions with modeled substrate

Residue	Substrate atom
H340	O6
R403	O2
D405	C1
D405	O5
E458	Glycosidic-O
H525	O2
D526	O2
D526	O3

comes in properly oriented to form the α-1,6 bond. At this point E458, which acted as a proton donor in the first step of the reaction, will now act at as proton acceptor, deprotonating the hydroxyl group in the C6 of the new incoming sugar. Once this occurs, the α-1,6 glycosidic bond is formed and the glutamic acid is regenerated.

In the case of isoamylase and α-amylase, where there is not a bond formation, a water molecule acts as the proton donor hydroxylating the C1. This brings up the question of what makes CGT and BE perform a transglycosylation reaction and not just a hydrolysis with a water molecule acting as the proton donor. A possible explanation for this could be that in isoamylase and α-amylase, a water molecule is sequestered and clamped in position. It has been mentioned that residue D526 could be the one to activate the water molecule that hydroxylates the C1 of the sugar (12). However, a comparison of the water molecules in the vicinity of D526 and the C1 of the sugar,

among members of the family, does not provide any conclusive information, as a water molecule in this region that would only be particular to the hydrolases was not observed. Another explanation would be the simultaneous binding of the substrate and the sugar that will act as the proton donor.

In Figure 3.17 the different sugars involved in the reaction have been modeled onto the BE structure. We base this mimic on the substrate and intermediate bound structures of other members of the α-amylase family, taking into account the unique loop structure of BE. We also modeled the position and orientation that the incoming sugar must be in order to be able to form the α-1,6 branch.

3.5 Electrostatic potential surface

An analysis of the surface charge distribution of BE was performed by the calculation of the electrostatic potential surface (EPS) (Figure 3.18). The overall surface of BE is quite electronegative and the cavity formed in the (α/β) barrel domain is the most electronegative feature of the surface. This cavity contains four electronegative residues; D335, D405, E458 and D526 known to be involved in substrate binding and catalysis. The EPS of α-amylase, CGT and isoamylase was also calculated (Figure 3.19). It can be observed in Figure 3.19 that all members of the α-amylase family present an overall electronegatively charged surface. The highly electronegative character of the (α/β) barrel domain in all four members of the family indicates that this negatively charged catalytic cavity is important for sugar-protein interactions.

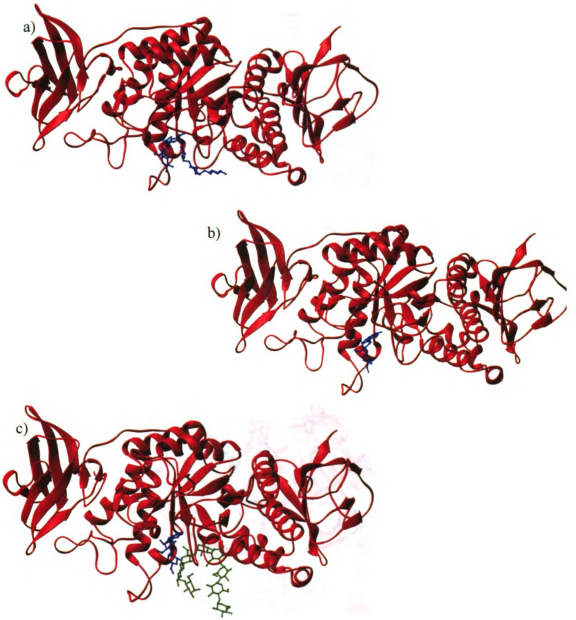


Figure 3.17 Proposed mode of action for BE catalysis. a) substrate binding b) intermediate formation and c) model of the position that the incoming sugar must have to form the α-1,6 branch.

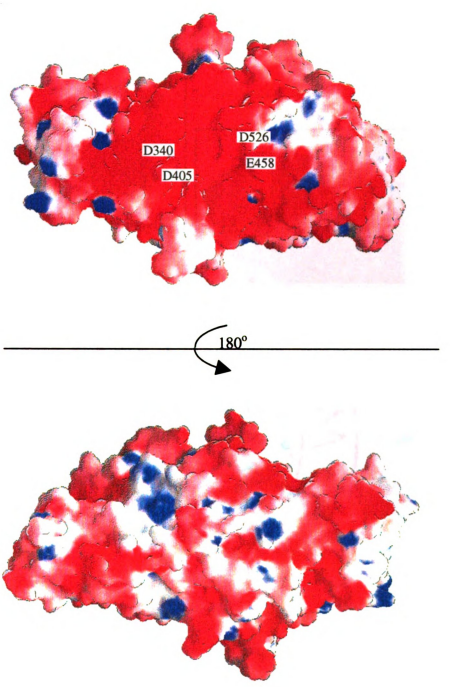


Figure 3.18 Electrostatic potential surface picture of BE a) Looking down the barrel and b) rotated 180°. The EPS calculation corresponds to 10kT/e for the blue color, -10kT/e for red and an EPS ~ 0 is white, where 10kT ~ 6 kcal/mol.

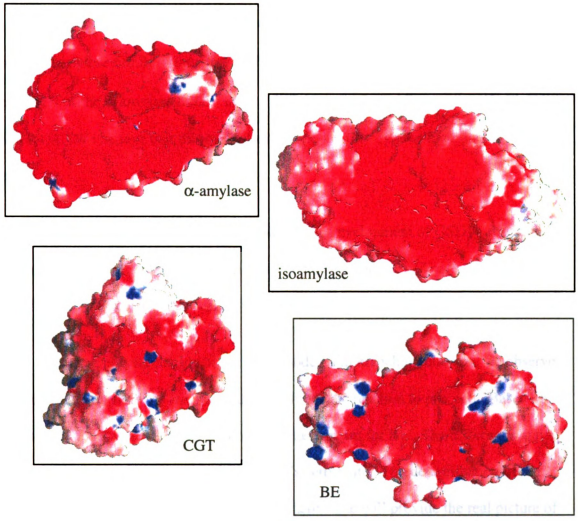


Figure 3.19 EPS of members of the α-amylase family of enzymes. The structures are oriented looking straight into the central barrel domain.

3.6 Conclusions

Branching enzyme has a central catalytic (α/β) domain like all the other members of the α-amylase family. Comparison of the conserved catalytic residues in BE with other members of the α-amylase family show a different orientation for residues H340, D405 and E458. We believe that this might be important for the type of reaction that BE catalyzes.

Analysis of the residues responsible for the development of GSDIV revealed that these mutations, in addition to the deletions and truncations, are likely to cause the unfolding of the protein, inactivating the enzyme.

When a polysaccharide molecule is modeled in branching enzyme, we observe that the catalytic residues are oriented properly for the reaction to proceed. Based on these results we have been able to propose a mechanism for BE. The different sugars involved in the reaction were modeled in BE's active site. The determination of the structure of BE with a substrate bound in the active site will provide the real picture of the protein sugar interaction.

3.7 Literature cited

1. Binderup, K., Mikkelsen, R., and Preiss, J. (2001) *Arch. Biochem. Biophys.* , in press

2. Jespersen, H. M., MacGregor, E. A., Sierks, M. R., and Svensson, B. (1991) *Biochem J* **280,** 51-55.

3. Uitdehaag, J. C., van Alebeek, G. J., van Der Veen, B. A., Dijkhuizen, L., and Dijkstra, B. W. (2000) *Biochemistry* **39**(26), 7772-7780.

4. Brzozowski, A. M., and Davies, G. J. (1997) *Biochemistry* **36**(36), 10837-10845.

5. Katsuya, Y., Mezaki, Y., Kubota, M., and Matsuura, Y. (1998) *J Mol Biol* **281**(5), 885-897.

6. Uitdehaag, J. C., Mosi, R., Kalk, K. H., van der Veen, B. A., Dijkhuizen, L., Withers, S. G., and Dijkstra, B. W. (1999) *Nat Struct Biol* **6**(5), 432-436.

7. DiMauro, S., and Tsujino, S. (1994) in *Myology* (A.G., E., and C., F.-A., eds) Vol. 2, pp. 1554-1576.

8. Bao, Y., Kishnani, P., Wu, J. Y., and Chen, Y. T. (1996) *J Clin Invest* **97**(4), 941-948.

9. Chen, Y. T., and Burchell, A. (1995) in *The metabolic and molecular basis of inherited diseases* (C.R., S., A.L., B., W.S., S., and D., V., eds) Vol. 1, pp. 935-965.

10. Schroder, J. M., May, R., Shin, Y. S., Sigmund, M., and Nase-Huppmeier, S. (1993) *Acta Neuropathol* **85**(4), 419-430.

11. Tao, B. Y., Reilly, P. J., and Robyt, J. F. (1989) *Biochim Biophys Acta* **995**(3), 214-220.

12. Svensson, B. (1994) *Plant Mol Biol* **25**(2), 141-157.

CHAPTER IV: THE THREE DIMENSIONAL STRUCTURE OF ANGIOSTATIN

4.1 Overall structure of angiostatin.

The three dimensional structure of angiostatin represents an important source of information for the understanding of the recently-born field of angiogenesis. Also, this is the first multikringle structure to be solved and the first structure of Kg3. The complete structure of human angiostatin consists of 253 residues that extend from amino acids 81 to 333. In addition, it contains 398 waters and three bicine molecules that form part of the crystal lattice.

The overall structure of angiostatin can be described as a triangle with dimensions of 60 x 57 x 45 Å and 32 Å in depth (Figures 4.1 and 4.2). Each Kg domain contains three β strands connected by a series of loops. The kringle is held together by disulfide links; there are three disulfides per Kg, as shown in Figure 4.2. These disulfides are formed between C84-C162, C105-C145 and C133-C157 in Kg1; C166-C243, C187-C226 and C215-C238 in Kg2; and C256-C333, C277-C316 and C305-C328 in Kg3. There is also an inter-kringle (inter-Kg) disulfide between C169 in Kg2 and C297 in Kg3. These Kg domains are connected to each other by inter-Kg peptides (Figures 4.1a and 4.2). The short inter-Kg peptide connecting Kg1 to Kg2 consists of three glutamates and the longer inter-Kg peptide between Kg2 and Kg3 contains 12 residues.

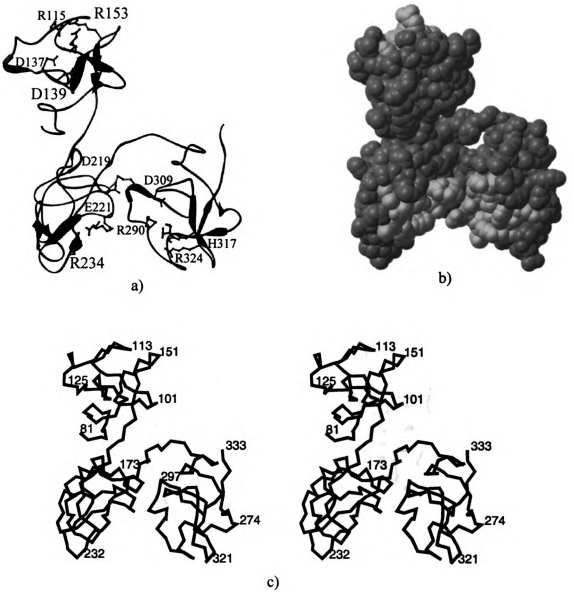


Figure 4.1 Three different representations of the overall structure of angiostatin. (a) Ribbon picture showing Kg1, orange; Kg2, magenta; Kg3, cyan; inter-kg peptide between Kg1 and Kg2, blue; inter-Kg peptide between Kg2 and Kg3, green; bicines, green with atoms in atom colors (nitrogen, blue and oxygen, red); intKg disulfide, yellow. LBS side groups also in atom colors. (b) Space filling view of angiostatin. The LBS in each of the three kringles is colored gold. All other atoms are red. (c) Stereo view of the Cα trace.

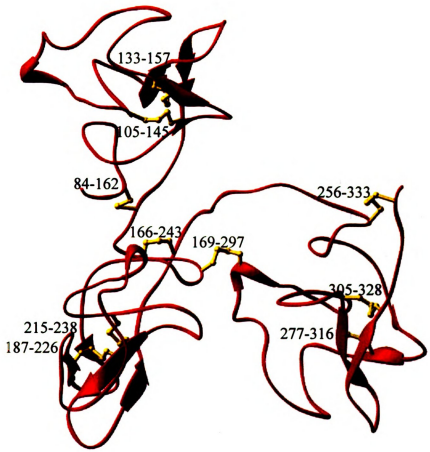


Figure 4.2 The disulfide links in angiostatin. Angiostatin shown in red with all disulfide bonds in yellow

The Kg domains are homologous, not only throughout their amino acid sequence, but also structurally (Figure 4.3). A superposition of the Kgs gives rmsd values between 0.40 to 0.46 Å, and the same is observed between the individual Kgs and the Kgs in angiostatin (Figure 4.3). The results of the C_α superposition are listed in Table 4.1. Beside the structural homology between Kgs, this also shows that the structures of individual Kgs are good representations of the Kgs found in multi-Kg structures like angiostatin.

In angiostatin, we observe that Kg2 and Kg3 are oriented with their LBSs facing each other, forming a 20 Å cavity. We believe that this cavity may be an important binding site in angiostatin, which will be elaborated on later. Their LBS's are related by a 112° rotation about an axis between them, coupled with a 1.6 Å translation (Figure 4.1). The anionic centers between Kg2 and Kg3 are about 13.5 Å apart, while the cationic ones are separated further at 25 Å. The corresponding transformations between the other kringles of angiostatin are: Kg1 to Kg2: 136°, 0.9 Å; Kg1 to Kg3: 163°, 1.2 Å. The LBS of Kg1 is facing towards the back of the molecule as shown in Figure 4.1.

4.2 The electrostatic surface of angiostatin

An analysis of the surface charge distribution of angiostatin was performed by the calculation of the electrostatic potential surface (EPS) (Figure 4.4). The EPS calculation

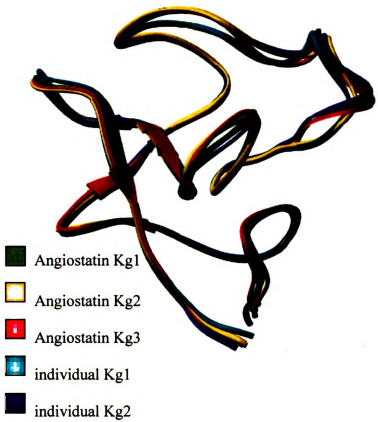


Figure 4.3 Superposition of various Kgs from plasminogen. This figure includes Kg1, Kg2 and Kg3 from angiostatin and the individual Kg1 and Kg2 (1,2).

Table 4.1 Rmsd values of the superposition of the Cα positions of individual Kgs and the Kgs in angiostatin

	Kg1-angiostatin	Kg2-angiostatin	Kg3-angiostatin	Kg1-individual	Kg2-individual
Kg1-angiostatin	-	0.45	0.40	0.41	0.37
Kg2-angiostatin	0.45	-	0.37	0.37	0.46
Kg3-angiostatin	0.40	0.37	-	0.40	0.41
Kg1-individual	0.41	0.37	0.40	-	0.52
Kg2-individual	0.37	0.46	0.41	0.52	-

corresponds to 10kT/e for the blue color, -10kT/e for red and an EPS ~ 0 for white, where 10kT ~ 6 kcal/mol. The EPS shows a neutral overall structure with prominent bipolar character in the LBSs of Kg1 and Kg2. The most outstanding electronic feature is the highly electropositive LBS of Kg3, compared to the other Kgs. In Figure 4.4b, the EPS was rotated 180° to display the back of the structure showing the LBS of Kg1. Inspection of this side of the EPS reveals the dipolar character of the Kg1 LBS, an electronegative charge cluster corresponding to the Kg1-Kg2 linker (with three consecutive glutamates) and the non-polar faces of the Kg2 and Kg3. In addition, there is an electropositive crescent created by R223 and R242 of Kg2 (Figure 4.4b). The central cavity formed by all three Kgs seems to form a non-polar binding surface that might be important for interacting with other ligands.

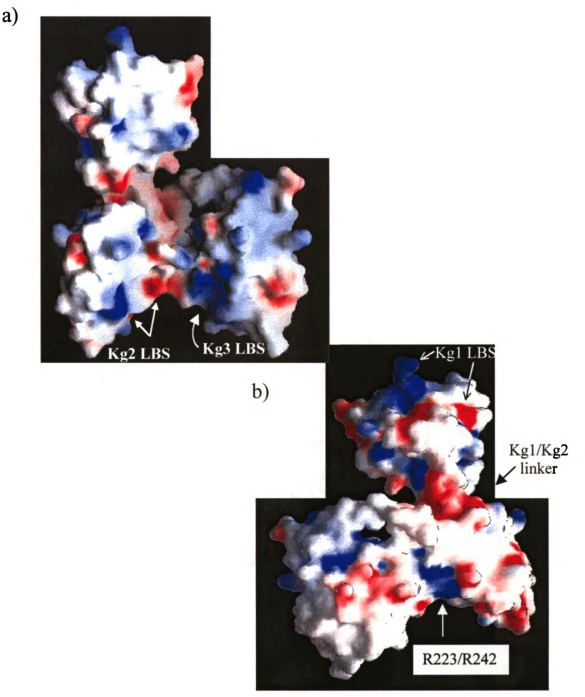


Figure 4.4 Electrostatic potential surface (EPS) of angiostatin with the bicines omitted.

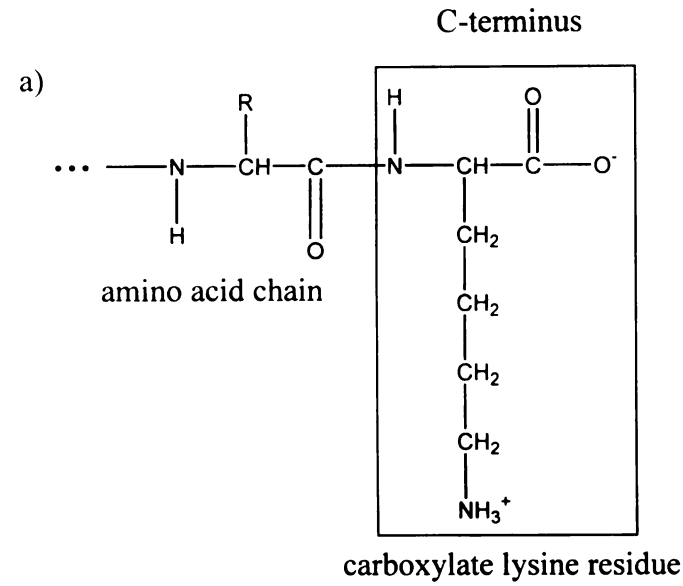
a) Same orientation as in Figure 4.1. b) Rotated 180° to show Kg1's LBS

4.3 Ligand specificity of the kringle LBS

The LBS is defined by residues 115-119, 137-139, 144-146, and 153-155 in Kg1; residues 201-207, 219-221, 224-228, and 234-237 in Kg2, and residues 287-291, 308-312, 314-318, and 324-327 in Kg3. In Figure 4.1b, the residues in the LBS of each Kg are depicted in yellow, while all other residues are shown in red and in Figure 4.2 the LBS of each Kg is marked by magenta bars in the sequence alignment. Each LBS in angiostatin has a bound bicine molecule; where bicine was the buffer used for crystallization. The three bicines were modeled in the electron density and in the refined structure.

Although bicine is not a carboxylate lysyl analog, it does have a carboxyl end (Figure 4.5). The carboxyl group of the bicine interacts with the cationic center in the LBS in the same way EACA does, as shown in the Kg1/EACA structure (Figure 4.6a) (1). The bound bicine molecules provided important information that helped in understanding the various binding affinities of each LBS for EACA, a mimic for the amino terminal lysine in fibrin. Among the three Kgs, Kg1 has the highest binding affinity for EACA with a K_D of 15.5 μM, while Kg2 has a K_D of only 401 μM. Kg3 shows no affinity for EACA (3-5). The binding affinity of Kg1 for bicine was measured to be in the low millimolar range (6).

In the Kg1/EACA structure, R117 and R153 stabilize the carboxylate and D137 and D139 stabilize the ammonium end of EACA, while the hydrophobic residues F118, W144, Y154, and Y156 stabilize the five carbon methene chain by hydrophobic interactions (Figure 4.6a). Analysis of the interactions between angiostatin's Kg1 and



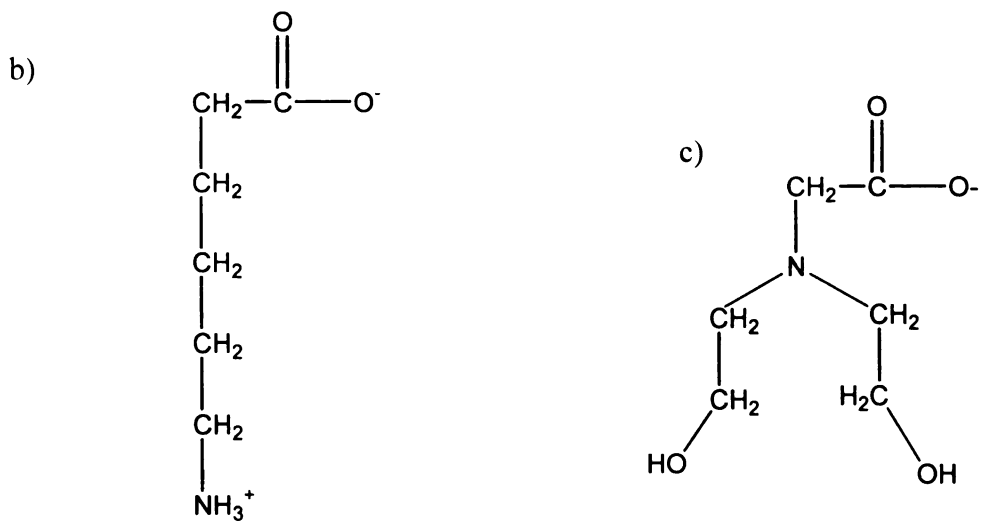


Figure 4.6 a) Carboxylate lysine residue. b) The carboxylate lysine analog, ε-aminocaproic acid. c) Bicine

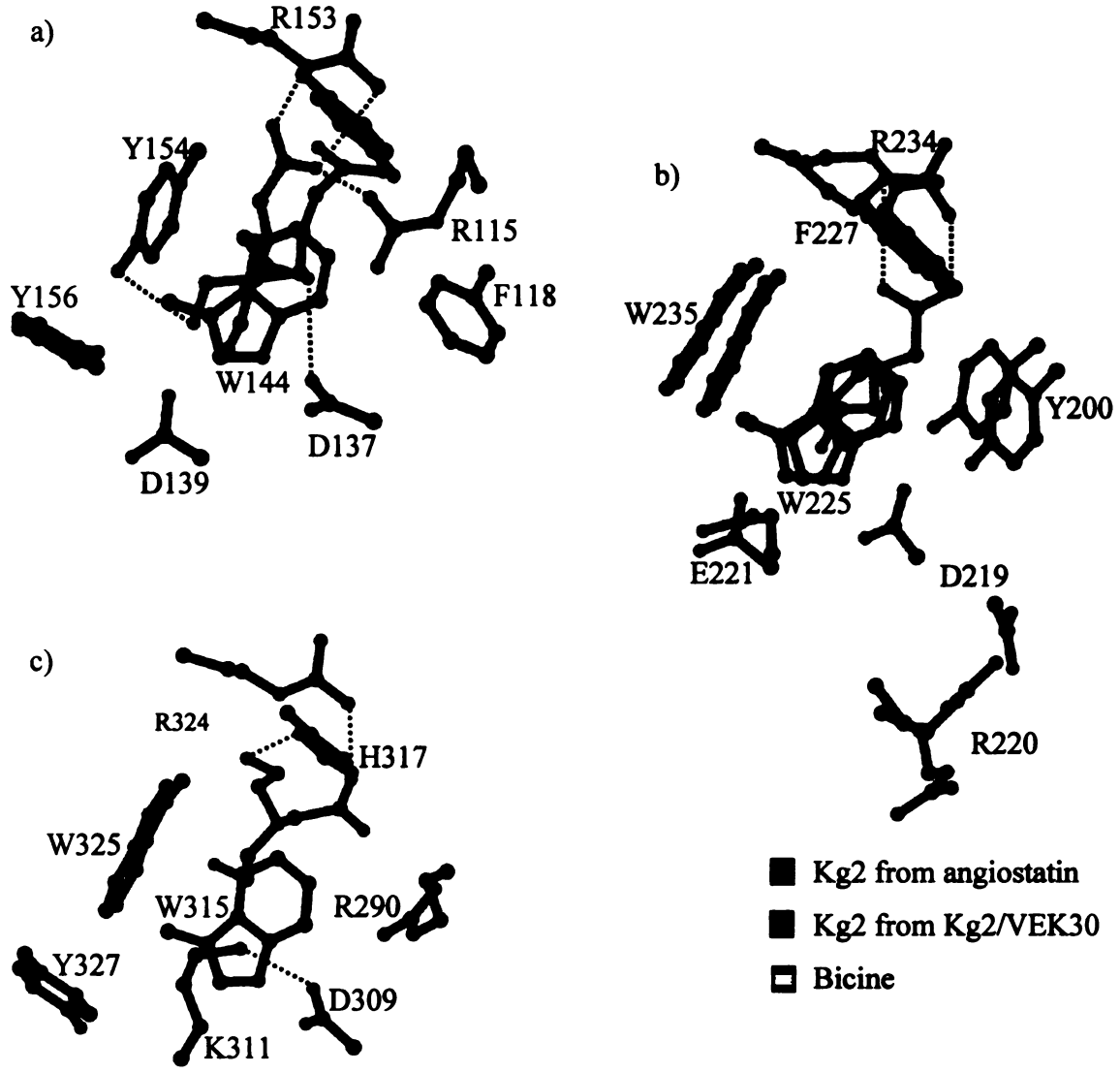


Figure 4.6 Interaction of the three angiostatin LBS's with bicine. All three depictions are in the same orientation. Hydrogen bonds and salt bridge contacts are shown by dotted lines. Residues from angiostatin are shown in green with atom colors, bicines are shown in yellow. (a) Comparison of the binding of Kg1 to bicine and EACA. EACA is shown in lavender (1). (b) Angiostatin Kg2 and Kg2 from the Kg2/VEK-30 structure are overlaid. Residues from the VEK-30 peptide are not shown for clarity. (2) (c) The angiostatin Kg3 LBS with bound bicine.

bicine shows that the catalytic center formed by R115 and R153 stabilizes the carboxylate head of the bicine by ionic interactions (Figure 4.6a). This carboxylate head of the bicine is oriented like EACA in the Kg1/EACA, as shown in Figure 4.6a (1). The hydroxyl tails of bicine are stabilized by interactions with W144, Y154, F118, and D137.

The bicine molecule in Kg2 is similarly oriented to the bicine in Kg1 with R234 forming a salt bridge with the carboxylate end of bicine (Figure 4.6b). Nonetheless there is a difference in conformation between the conserved aspartates of Kg1 (D137) and Kg2 (D219). In Kg1, D137 is oriented towards the LBS, ready to make an ionic interaction with the ammonium group of EACA. However, the D219 in Kg2 interacts with the non-conserved R220 (an aspargine in Kg1 and a glycine in Kg3) rotating its side chain out of the LBS. The position of the D219 side chain is too far away to interact with the ammonium group of EACA. This leaves an incomplete LBS with only one functional residue, E221, in the anionic center and a cationic center with only one residue (R234) as well. We believe this to be the reason why Kg2 has a lower EACA affinity compared to Kg1. The residues W225, W235 and E221 interact with the bis ethyl hydroxy end of the bicine molecule (Figure 4.6b).

One of the most interesting results obtained from the structure of angiostatin comes from the analysis of the LBS in Kg3. The limitation of not having a structure of Kg3 made it impossible to explain why Kg3 is the only Kg of plasminogen with no affinity for EACA (5). Even though Kg3 has no affinity for EACA, its LBS is occupied by a bicine molecule. The carboxylate part of the bicine interacts ionically with the cationic center formed by R324 and R290. The LBS of Kg3 is highly electropositive

with a lysine substituting for one of the aspartates in the anionic center. The positively charged amino terminal tail of EACA will cause electrostatic repulsion between EACA and the highly electropositive LBS of Kg3. Also, inspection of the LBS shows that this lysine (K311) fills half of the LBS, preventing the binding of long molecules like EACA. This can be observed in Figure 4.7a where an EACA molecule was modeled onto the Kg3 LBS by overlaying the structure of Kg1 onto Kg3. There is a salt bridge between K311 and D309 that holds K311 in position. The bicine binds in a totally different conformation compared to the other bicines of angiostatin to avoid steric clashes (Figure 4.7). The bicine in Kg3 is rotated 90° around its C_β atom as shown in Figure 4.7b. The LBS in Kg3 is reduced in size, quite positively charged and without bipolar character. This explains the non-affinity of Kg3 for EACA.

Studies performed on the K311D mutant of Kg3 shows some affinity for EACA and other small molecule C-terminal lysine analogs, indicating that K311 inhibits binding of the molecules (4,5). These observations suggest that the Kg3 LBS is ideally suited to bind only carboxylate-containing ligands such as aspartate or glutamate and not extended bipolar ligands such as EACA or C-terminal lysine residues. This represents a new binding mode, specific to Kg3 like kringles, that results from the highly electropositive nature of this LBS.

It is not clear what role the LBS plays in the angiogenesis inhibitory activity of angiostatin. Previous studies show no correlation between the antiangiogenic activity of individual Kgs and their EACA binding affinity. For example, Kg2 has a higher affinity

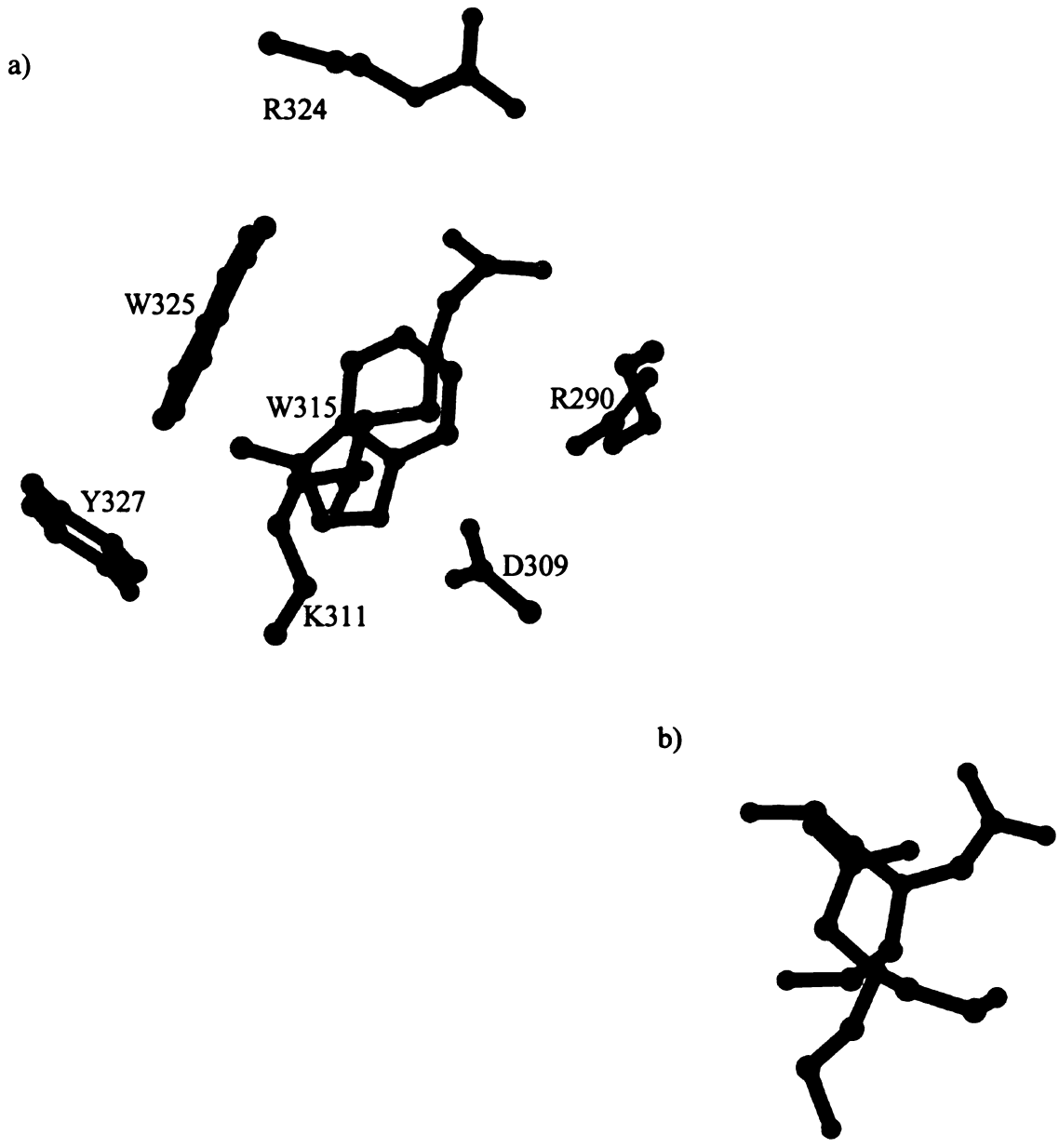


Figure 4.7 a) EACA molecule was modeled onto the Kg3 LBS by overlaying the structures of Kg1 onto Kg3. b) Different conformation of the bicine in Kg1 of angiostatin depicted in red and the bicine in Kg3 shown in blue.

for EACA than Kg3, but its inhibitory activity is lower than Kg3 (7). Other studies performed in the angiostatin C169S/C297S double mutant resulted in the loss of EACA binding by Kg2 without altering its antiangiogenic activity (8). It is important to mention that the lysine affinity monitored is that of EACA, which is a good model for carboxy terminal lysines. It may be that angiostatin binds internal lysine residues, or it could even have some other binding specificity thus far unknown.

4.4 Angiostatin binding to protein domains

It has been determined that Kgs not only bind six carbon zwitterions such as lysine and EACA, but also protein domains (2,9,10). The recently determined structure of Kg2, bound to a peptide sequence (VEK30) of the *Streptococcal* surface protein PAM, presented a model for protein binding at the surface of bacteria (2). Since angiostatin offers a more realistic model of the physiological target of the *Streptococcal* surface protein, we modeled the VEK30 peptide onto angiostatin. In order to do this, we overlaid the Kg2 from the Kg2/VEK30 structure onto angiostatin's Kg2 (rmsd = 0.41) (Figure 4.8). The α-helix of the VEK30 peptide is accommodated without collisions in the 20 Å cavity between Kgs with enough space left to fit another VEK30 helix. Kg2 from the Kg2/VEK30 structure was also superimposed onto angiostatin's Kg3 (rmsd 0.40), showing that both helices fit well, with Kg3 interacting with VEK30 similarly to Kg2. The VEK30 peptide has an arginine and a glutamate separated by one turn of a helix. This resembles a carboxylate lysine with a positive (R101) and a negative (E104) end that interacts with Kg2's E221 and R234, respectively. Interestingly, upon VEK30

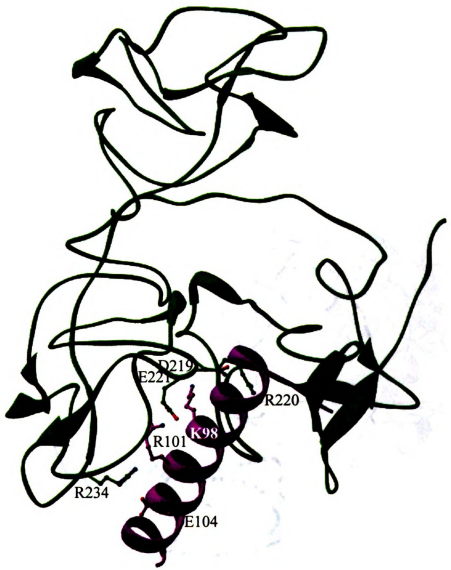


Figure 4.8 A Ribbons depiction of the modeled angiostatin/VEK30 complex. The Kg2 of the Kg2/VEK30 complex was overlaid on angiostatin Kg2 (2). Angiostatin is colored green while VEK30 is colored lavender. Side groups are labeled appropriately.

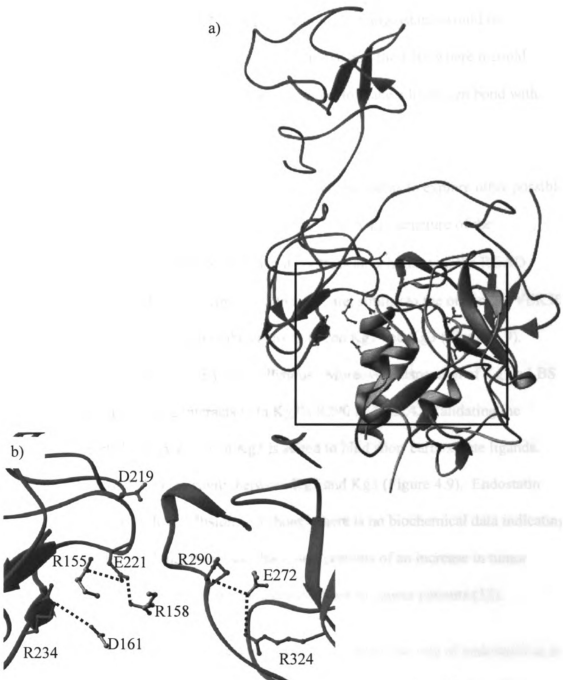


Figure 4.9 Endostatin modeled onto the angiostatin (11). This was done by overlaying the helices of endostatin and VEK30. Endostatin is colored purple and angiostatin green. b) Close view of the section of angiostatin that harbors the residues that may be involved in endostatin binding.

binding, the salt bridge between R220 and D219 in Kg2 of angiostatin would be disrupted by the helix. This would force D219 to move into the LBS where it could interact with VEK30 K98. The R220 could also move to make a hydrogen bond with VEK30 Q95.

The ability of angiostatin to bind a protein domain led us to explore other possible molecules relevant to angiostatin's mode of action. The X-ray structure of the angiogenesis inhibitor endostatin reveals that it contains an α-helix with the RGAD sequence (11). R158 and D161 form a pseudo lysyl site, similar to the one in the VEK30 peptide. Endostatin was modeled in the cavity between Kg2 and Kg3 (Figure 4.9). Endostatin fills the cavity with very few collisions. Moreover, inspection of Kg3's LBS shows that endostatin's E272 interacts with Kg3's R290 and R324. Validating the previously suggested observation that Kg3 is suited to bind short carboxylate ligands. Endostatin was modeled in the cavity between Kg2 and Kg3 (Figure 4.9). Endostatin fills the cavity with very few collisions. Although there is no biochemical data indicating that these two molecules bind, there have been observations of an increase in tumor reduction when both inhibitors are given in combination to cancer patients (12).

Marneros and Olsen proposed a mechanism to explain the role of endostatin as an angiogenesis inhibitor (13). This mechanism is based on the binding affinity of the endostatin domain of collagen XVIII. Collagen XVIII is involved in the activation of cell migration by interactions with extracellular components through its endostatin domain. It is proposed in this model that endostatin binding competes for the binding of matrix components, inhibiting endothelial cell migration and consequently angiogenesis. We

propose that in the same way, angiostatin can interact with the endostatin domain of collagen XVIII by the interactions previously proposed, and this could possibly be the inhibitory mechanism of angiostatin.

Recent studies indicate that angiostatin binds the β_3 subunit of the endothelial cell surface receptor, $\alpha_v\beta_{3\,(14)}$. The interaction is inhibited by EACA, but only at concentrations high enough to occupy Kg2. This reinforces the importance of the cavity between Kg2 and Kg3 for interaction with protein domains. We have identified a sequence KKVEE in an exposed α-helix of the β_3 subunit (Figure 4.10). We believe that this sequence can represent a pseudo-lysyl site similar to the one in the VEK30 peptide. However, the Kg2-Kg3 cavity must open further for the $\alpha_v\beta_3$ integrin to fit in the cavity. It is believed that the binding of angiostatin to $\alpha_v\beta_3$ might perturb a critical signaling pathway for endothelial cells, therefore inhibiting angiogenesis. Some of these results reinforce the importance of the Kg2-Kg3 cavity in the protein domain binding. This could mean that the angiostatin inhibitory activity resides in this cavity and not in the LBS.

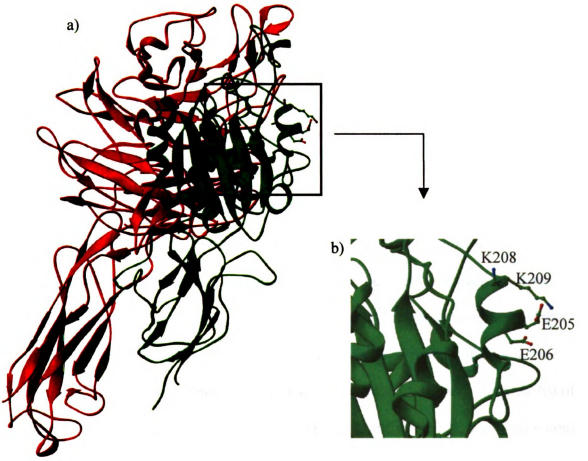


Figure 4.10 a) Structure of $\alpha_v\beta_3$ integrin (15). The α_v subunit is shown in red and the β_3 subunit in green; also shown are the residues that could be involved in angiostatin binding. b) Close view of the section of β_3 that harbors the residues involved in possible angiostatin binding.

4.5 The inter-kringle disulfide bond

Studies performed on angiostatin as an inhibitor of endothelial cell proliferation revealed that the Kg2-Kg3 fragment exhibits inhibitory activity, similar to Kg2 alone (7). Also, an enhancement in inhibitory activity is observed with individual Kg2 and Kg3 versus the Kg2-Kg3 fragment (7). It was initially suggested that it is necessary to open the Kg2-Kg3 disulfide bond in order to obtain maximum antiangiogenic activity. With this in mind, the angiostatin double mutant C169S/C297S, which eliminates the inter-Kg disulfide bond, was constructed (8). This mutant had little effect in antiangiogenesis activity. Analysis of the interactions between Kg2 and Kg3 in angiostatin revealed that there are numerous contacts between Kg2 and Kg3 and that disruption of the disulfide bond should not alter the structure of angiostatin. Appendix 4.1 shows a complete list of interactions between Kgs and their inter-Kg peptides. Some of the Kg2/Kg3 interactions include a salt bridge between E163 and H168, four hydrogen bonds between the inter-Kg peptide between Kg2 and Kg3 and Kg2 and Kg3. These are only some of the 102 contacts between all three Kgs and inter-Kg peptides (Table 4.2). Various interactions within this structure indicate that angiostatin forms a molecular entity much like a single domain protein that might function cooperatively.

4.6 Conclusions

The numerous interactions between the Kgs in angiostatin produce a unique domain that harbors a recognition site important for angiogenesis inhibition. This

recognition site could be the cofacial orientation of the LBSs in Kg2 and Kg3, which forms a cavity 20 Å wide. The models of angiostatin with VEK30 and endostatin illustrate how the interaction may occur. This could mean that the inhibitory activity of angiostatin does not reside in the individual LBSs, but in this cavity, with the Kg domains working in concert. Also, the orientations of all three LBSs in angiostatin demonstrate that the LBSs in this multi-Kg structure remain functionally viable.

The structure of angiostatin has also provided an explanation for the inability of Kg3 to bind EACA. The structure showed that the LBS of Kg3 is reduced in size, quite positively charged and without bipolar character. Additionally, the low affinity of Kg2 for EACA is explained by the rotation of the D219 side chain outside the LBS.

Table 4.2 Summary of the Kg-Kg interactions and inter-Kg peptide Kg interaction of angiostatin. The interactions are determined with a cutoff distance of <4.0 Å.

Interactions between	Number of interactions
Kg2/Kg3	28
Kg2/inter-Kg(Kg1-Kg2) peptide	19
Kg2/inter-Kg(Kg2-Kg3) peptide	12
Kg3/inter-Kg(Kg2-Kg3) peptide	43

4.7 Literature cited

1. Mathews, I., Vanderhoff-Hanaver, P., Castellino, F. J., and Tulinsky, A. (1996) *Biochemistry* **35**(8), 2567-2576.

2. Rios-Steiner, J. L., Schenone, M., Mochalkin, I., Tulinsky, A., and Castellino, F. J. (2001) *J Mol Biol* **308**(4), 705-719.

3. Chang, Y., Mochalkin, I., McCance, S. G., Cheng, B., Tulinsky, A., and Castellino, F. J. (1998) *Biochemistry* **37**(10), 3258-3271.

4. Burgin, J., and Schaller, J. (1999) *Cell Mol Life Sci* **55**(1), 135-141.

5. Marti, D., Schaller, J., Ochensberger, B., and Rickli, E. E. (1994) *Eur J Biochem* **219**(1-2), 455-462.

6. Castellino, F. J. (2001), University of Notre Dame, personal communication

7. Cao, Y., Ji, R. W., Davidson, D., Schaller, J., Marti, D., Sohndel, S., McCance, S. G., O'Reilly, M. S., Llinas, M., and Folkman, J. (1996) *J Biol Chem* **271**(46), 29461-29467.

8. Lee, H., Kim, H. K., Lee, J. H., You, W. K., Chung, S. I., Chang, S. I., Park, M. H., Hong, Y. K., and Joe, Y. A. (2000) *Arch Biochem Biophys* **375**(2), 359-363.

9. Moser, T. L., Stack, M. S., Asplin, I., Enghild, J. J., Hojrup, P., Everitt, L., Hubchak, S., Schnaper, H. W., and Pizzo, S. V. (1999) *Proc Nat Acad Sci USA* **96**(6), 2811-2816.

10. Troyanovsky, B., Levchenko, T., Mansson, G., Matvijenko, O., and Holmgren, L. (2001) *J Cell Biol* **152**(6), 1247-1254.

11. Hohenester, E., Sasaki, T., Olsen, B. R., and Timpl, R. (1998) *Embo J* **17**(6), 1656-1664.

12. Yokoyama, Y., Dhanabal, M., Griffioen, A. W., Sukhatme, V. P., and Ramakrishnan, S. (2000) *Cancer Research* **60**(8), 2190-2196.

13. Marneros, A. G., and Olsen, B. R. (2001) *Matrix Biol* **20**(5-6), 337-345.

14. Tarui, T., Miles, L. a., and Takada, Y. (2001) *J. Mol. Biol.* **276**(43), 39562-39568.

15. Xiong, J. P., Stehle, T., Diefenbach, B., Zhang, R., Dunker, R., Scott, D. L., Joachimiak, A., Goodman, S. L., and Arnaout, M. A. (2001) *Science* **294**(5541), 339-345.

APPENDIX

Appendix 4.1 Kg/Kg and Kg/intKg peptide interactions of angiostatin. The interactions are determined with a cutoff distance of 4.0 Å.

kg1/intKg(Kg1-Kg2)

#	name	atom	#	name	atom	distance (Å)
161	GLU	O	163	GLU	N	3.60
162	CYS	N	163	GLU	N	3.61
162	CYS	CA	163	GLU	N	2.43
162	CYS	CA	163	GLU	CA	3.84
162	CYS	C	163	GLU	N	1.34
162	CYS	C	163	GLU	CA	2.47
162	CYS	C	163	GLU	CB	3.58
162	CYS	C	163	GLU	CG	3.82
162	CYS	C	163	GLU	C	3.44
162	CYS	C	164	GLU	N	3.82
162	CYS	O	163	GLU	N	2.27
162	CYS	O	163	GLU	CA	2.84
162	CYS	O	163	GLU	C	3.80
162	CYS	O	164	GLU	N	3.85
162	CYS	CB	163	GLU	N	3.24
162	CYS	CB	164	GLU	OE2	3.83

kg2/intKg(Kg1-Kg2)

#	name	atom	#	name	atom	distance (Å)
166	CYS	N	164	GLU	C	3.41
166	CYS	N	164	GLU	O	3.61
166	CYS	N	165	GLU	N	2.84
166	CYS	N	165	GLU	CA	2.40
166	CYS	N	165	GLU	CB	3.12
166	CYS	N	165	GLU	C	1.32
166	CYS	N	165	GLU	O	2.24
166	CYS	CA	165	GLU	CA	3.78
166	CYS	CA	165	GLU	C	2.43
166	CYS	CA	165	GLU	O	2.78
166	CYS	C	165	GLU	C	3.59
166	CYS	O	163	GLU	CA	3.86
166	CYS	O	163	GLU	CB	3.47
166	CYS	O	163	GLU	C	3.71
166	CYS	O	163	GLU	O	3.95
166	CYS	O	165	GLU	C	3.80
166	CYS	CB	165	GLU	C	3.28
166	CYS	CB	165	GLU	O	3.52
167	MET	CA	163	GLU	OE1	3.53
167	MET	CB	163	GLU	OE1	3.91
167	MET	C	163	GLU	OE1	3.80
168	HIS	N	163	GLU	CD	3.75
168	HIS	N	163	GLU	OE1	3.10
168	HIS	CB	163	GLU	OE2	3.73
168	HIS	ND1	163	GLU	CD	3.96
168	HIS	ND1	163	GLU	OE2	3.77

173	ASN	CB	163	GLU	OE2	3.71
173	ASN	C	163	GLU	OE1	3.67
173	ASN	O	163	GLU	CB	3.76
173	ASN	O	163	GLU	CD	3.75
173	ASN	O	163	GLU	OE1	3.64
174	TYR	N	163	GLU	OE1	3.77
174	TYR	CA	163	GLU	OE1	3.89
174	TYR	CD1	163	GLU	O	3.69
174	TYR	CE1	164	GLU	O	3.78
174	TYR	C	163	GLU	O	3.91
175	ASP	N	163	GLU	O	2.85
175	ASP	CA	163	GLU	O	3.49
175	ASP	CB	163	GLU	O	3.81
175	ASP	C	163	GLU	O	3.44
175	ASP	O	164	GLU	CD	3.89
175	ASP	O	164	GLU	OE1	3.17
176	GLY	N	163	GLU	O	3.32
176	GLY	CA	164	GLU	CA	3.95
176	GLY	CA	164	GLU	O	3.61

kg2/intKg(Kg2-Kg3)

#	name	atom	#	name	atom	distance (Å)
168	HIS	CB	248	PRO	CD	3.93
168	HIS	CG	248	PRO	CD	3.63
168	HIS	CD2	247	PRO	CA	3.56
168	HIS	ND1	248	PRO	CD	3.74
168	HIS	CE1	246	PRO	O	3.82
168	HIS	NE2	245	THR	OG1	3.79
168	HIS	NE2	245	THR	O	3.70
168	HIS	NE2	246	PRO	C	3.83
168	HIS	NE2	246	PRO	O	3.24
168	HIS	NE2	247	PRO	CA	3.73
242	ARG	C	244	THR	N	3.91
242	ARG	O	244	THR	N	3.51
243	CYS	N	244	THR	N	3.54
243	CYS	CA	244	THR	N	2.42
243	CYS	CA	244	THR	CA	3.80
243	CYS	C	244	THR	N	1.34
243	CYS	C	244	THR	CA	2.44
243	CYS	C	244	THR	CB	3.69
243	CYS	C	244	THR	C	3.28
243	CYS	C	245	THR	N	3.49
243	CYS	O	244	THR	N	2.27
243	CYS	O	244	THR	CA	2.78
243	CYS	O	244	THR	C	3.42
243	CYS	O	245	THR	N	3.86
243	CYS	O	245	THR	O	3.95
243	CYS	CB	244	THR	N	3.29

kg2/kg3

#	name	atom	#	name	atom	distance (Å)
169	CYS	N	297	CYS	SG	3.51
169	CYS	CA	296	PRO	CG	3.91
169	CYS	CA	297	CYS	SG	3.29

169	CYS	O	296	PRO	CD	3.91
169	CYS	O	296	PRO	CG	3.72
169	CYS	CB	297	CYS	CB	3.84
169	CYS	CB	297	CYS	SG	3.10
169	CYS	SG	297	CYS	CB	3.01
169	CYS	SG	297	CYS	SG	2.05
170	SER	CB	294	ASN	O	3.93
222	LEU	CD1	289	GLU	OE1	3.96
222	LEU	CD2	289	GLU	CB	3.69
222	LEU	CD2	289	GLU	OE2	3.71
222	LEU	CD2	289	GLU	O	3.97
222	LEU	CD2	294	ASN	ND2	3.84
222	LEU	CD2	295	PHE	CZ	3.94
242	ARG	CZ	293	GLU	O	3.66
242	ARG	NH1	293	GLU	O	3.71
242	ARG	NH2	293	GLU	O	3.94

Kg3/intKg(Kg2-Kg3)

#	name	atom	#	name	atom	distance (Å)
256	CYS	N	254	TYR	CB	3.80
256	CYS	N	254	TYR	C	3.10
256	CYS	N	254	TYR	O	3.49
256	CYS	N	255	GLN	N	2.55
256	CYS	N	255	GLN	CA	2.43
256	CYS	N	255	GLN	CB	3.56
256	CYS	N	255	GLN	C	1.34
256	CYS	N	255	GLN	O	2.26
256	CYS	CA	255	GLN	N	3.99
256	CYS	CA	255	GLN	CA	3.83
256	CYS	CA	255	GLN	C	2.47
256	CYS	CA	255	GLN	O	2.84
256	CYS	C	255	GLN	C	3.58
256	CYS	O	254	TYR	CB	3.29
256	CYS	O	254	TYR	C	3.90
256	CYS	O	254	TYR	O	3.99
256	CYS	O	255	GLN	C	3.83
256	CYS	CB	255	GLN	C	3.41
256	CYS	CB	255	GLN	O	3.72
256	CYS	SG	254	TYR	CB	3.92
258	LYS	CG	254	TYR	CD2	3.88
258	LYS	CD	254	TYR	CD2	3.38
258	LYS	CD	254	TYR	CE2	3.18
258	LYS	CD	254	TYR	CZ	3.78
258	LYS	CE	254	TYR	CE2	3.77
258	LYS	CE	254	TYR	CZ	3.83
263	ASN	CB	252	PRO	O	3.45
263	ASN	CG	254	TYR	CD2	3.75
263	ASN	OD1	254	TYR	CD2	3.92
263	ASN	ND2	252	PRO	O	3.97
263	ASN	ND2	254	TYR	CD2	3.60
263	ASN	ND2	254	TYR	CE2	3.55
263	ASN	C	254	TYR	N	3.96
263	ASN	O	253	THR	CA	3.10
263	ASN	O	253	THR	CB	3.15

263	ASN	O	253	THR	CG2	3.49
263	ASN	O	253	THR	C	3.46
263	ASN	O	254	TYR	N	2.88
263	ASN	O	254	TYR	O	3.89
264	TYR	CA	254	TYR	O	3.44
264	TYR	CD1	254	TYR	O	3.15
264	TYR	CD1	255	GLN	C	3.93
264	TYR	CE1	254	TYR	O	3.78
264	TYR	CE1	255	GLN	CA	3.89
264	TYR	CE1	255	GLN	C	3.52
264	TYR	CE1	255	GLN	O	3.52
264	TYR	C	254	TYR	O	3.60
265	ARG	N	254	TYR	C	4.00
265	ARG	N	254	TYR	O	2.83
265	ARG	CA	254	TYR	O	3.81
265	ARG	CG	253	THR	CB	3.38
265	ARG	CG	253	THR	OG1	3.77
265	ARG	CG	253	THR	CG2	3.47
265	ARG	C	254	TYR	O	3.93
265	ARG	C	255	GLN	NE2	3.96
265	ARG	O	255	GLN	CD	3.97
265	ARG	O	255	GLN	OE1	3.82
265	ARG	O	255	GLN	NE2	3.52
266	GLY	N	254	TYR	O	3.66
266	GLY	N	255	GLN	NE2	3.88
266	GLY	CA	255	GLN	NE2	3.27
297	CYS	SG	247	PRO	CB	3.85
333	CYS	SG	254	TYR	CB	3.98

MICHIGAN STATE UNIVERSITY LIBRARIES
3 1293 02334 2383