Molecular Mechanism of Starch Digestion by Family 31 Glycoside Hydrolases:

Structural Characterization and Inhibition Studies of C-terminal Maltase Glucoamylase and Sucrase Isomaltase

by

Kyra Jill Jacques Jones

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Biology
Waterloo, Ontario, Canada, 2014

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Although carbohydrates are a principal component of the human diet, the mechanism of the final stages of starch digestion is not fully understood. One approach to treating metabolic diseases such as type II diabetes, obesity, and congenital sucrase isomaltase deficiency is inhibition of intestinal α-glucosidases and pancreatic α-amylases. Intestinal α-glucosidases, sucrase isomaltase (SI) and maltase glucoamylase (MGAM), are responsible for the final step of starch hydrolysis in mammals: the release of free glucose. MGAM and SI consist of two catalytic subunits: N-terminal and C-terminal, with overlapping, but variant substrate specificities.

The objective of this thesis is to increase the understanding of the differential substrate specificity seen in the catalytic subunits of SI and MGAM. Through inhibitor studies, the structural and biochemical differences between the enzymatic subunits are explored, illustrating that each individual catalytic subunit can be selectively inhibited. In Chapter 3, homology models of ctSI and ctMGAM-N20 are presented, giving insight into the residues hypothesized to impact substrate specificity, enhancing our understanding of the functionality of these enzymatic subunits and overlapping substrate specificity. The structural implications of mutations seen in ntSI in CSID patients and the potential functional and structural implications are discussed in Chapter 4 in addition to the prevalence of SNPs in the SI gene in different populations. The mammalian α-glucosidases are compared to the 3 Å structure of CfXyl31, a Family 31 glycoside hydrolase from *Cellulomonas fimi*. Comparison to Family 31 glycoside hydrolases of known structure gives rise to possible mutations proposed to mimic ntMGAM α-glucosidase activity.

Through inhibitor studies, homology models, examining mutations found in disease states such as congenital sucrase isomaltase deficiency, and investigating a bacterial family 31 glycoside hydrolase from *Cellulomonas fimi*, the active site characteristics and substrate specificities of SI and MGAM are better understood.

Acknowledgements

They say it takes a village to raise a child, but in my case it took a village to finish a PhD. There have been so many people who have helped and supported me during my journey as a graduate student. First, Dr. Todd Holyoak and the members of his lab have been a wonderful source of knowledge and advice, especially William Lotosky and Marc Adeasis. There have also been cohorts of co-op and volunteer students, especially Daniel Rickert who helped me with many purification and cloning projects. Finally, Steve George of the Aucoin lab was incredibly helpful during all baculovirus expression runs.

Members of the Rose lab, past and present, have also helped me along the way. Marcie Chaudet has been my friend and confidant over the past 4 years. We have shared ups and downs as grad students and friends and I appreciate her friendship and support immensely. Similarly, Dr. Megan Barker has been a wonderful friend and resource, both scientific and personal.

My collaborators and all of the "starchies" have been wonderful to work with and I have enjoyed our work together immensely. My supervisory committee, Dr. John Honek, Dr. Bernard Glick, and Dr. Brendan McConkey have been wonderful advisors during my PhD and have been very helpful in guiding my work. Dr. David Rose has been an incredible supervisor. He has shared his expertise with me and taught me more than I could ever imagine. He has supported me in all of my scientific endeavors and encouraged me to take advantage of every opportunity I have had along the way. He also sounded constant reminders to stop worrying and enjoy my work – and my work was always easier and my results better when I followed his advice and had fun. I will always be thankful for his support and guidance, not only has he taught me to be a good scientist, he has taught me how to enjoy being a good scientist.

My brothers (Luke and Noah) and my "new" siblings (Jennifer, Simon, Lisa, and Adrian) have been very supportive, sharing graduate school struggles, You Tube videos, and cousin nights to help ease my stress. Mom and Dad, Martin and Diane, you have been so helpful – from last minute babysitting when I have experiments to tend to and no childcare, to sending home meat pies and inviting us for meals, there are no words to express my gratitude. Without you, I wouldn't have been able to take advantage of many of the opportunities I had during graduate school. Thank you.

Dedication

I would like to dedicate my “book”, as Rylie calls it, to my two children and my husband Phil. Rylie, you were only four months old when I started grad school, and Emmett, you were born at the halfway point. The two of you have never known a life without Mommy going to the “science lab”. I am glad this seems normal to you, because my work is an important part of my life. I will always be grateful that both of you are supportive when I have to work at odd times and that I have never had to hide my work from you. I want you to be fully involved in my life. The two of you have made art to decorate my lab bench and Rylie; you are always there to help me choose the colours for my figures. I want you to know that despite my love of science, you will always be the most important part of my life; nothing, not even the “science lab” is more important than you. Phil - you have been an endless well of support over the past five years. When I was tired and wanting to give up, your support was unwavering. You are my best friend and the love of my life. You know me better than anyone and you are the only one who knows how to talk me out of putting so much pressure on myself when I feel overwhelmed. This is my largest accomplishment aside from our marriage and our children. Without you and the kids, I would never have gotten this far and I am so grateful. This “book” is for you guys.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction and Thesis Overview

1.1 Role of Maltase Glucoamylase and Sucrase Isomaltase in Starch Digestion

Starch comprises a substantial portion of the human diet and plays an integral role in human metabolism, contributing a sizable fraction of total caloric intake. Starch is digested to glucose in the small intestine, contributing to the postprandial rise in blood glucose levels as well as insulin response (Englyst and Kingman 1992). Two macromolecules comprise starch, amylose and amylopectin. Amylose is a linear chain polymer of α-D-glucopyranose units linked by α-1,4 bonds, with only minor branching. Amylopectin is a branched polymer of α-D-glucopyranose linked by α-1,4 linkages with branching α-1,6 bonds (Jones et al. 2011).

The digestion of starch to glucose requires multiple enzymatic reactions in the human body. Glucose is only a minor product of α-amylase digestion of starch, the major products being soluble glucose oligomers, with both linear and branched structures (Robyt and French 1970). The small, linear and branched α-limit oligomers (dextrins) produced by α-amylase hydrolysis cannot be absorbed into the bloodstream and must be further hydrolyzed to free glucose. Maltase- glucoamylase (MGAM) and sucrase-isomaltase (SI) process the small linear oligomers and α-limit dextrins in the small intestine (Figure 1-1) (Nichols et al. 1998; Jones et al. 2011).

Inhibition of intestinal α-glucosidases and pancreatic α-amylases is an approach to controlling blood glucose and serum insulin levels in individuals with Type II diabetes. Each of human MGAM and SI incorporates two Family 31 Glycoside Hydrolases (Cantarel et al. 2009) responsible for the final step of starch hydrolysis. The apparent redundancy in glucosidase activities in the gut is not fully understood and forms the basis for the present studies (Jones et al. 2011).

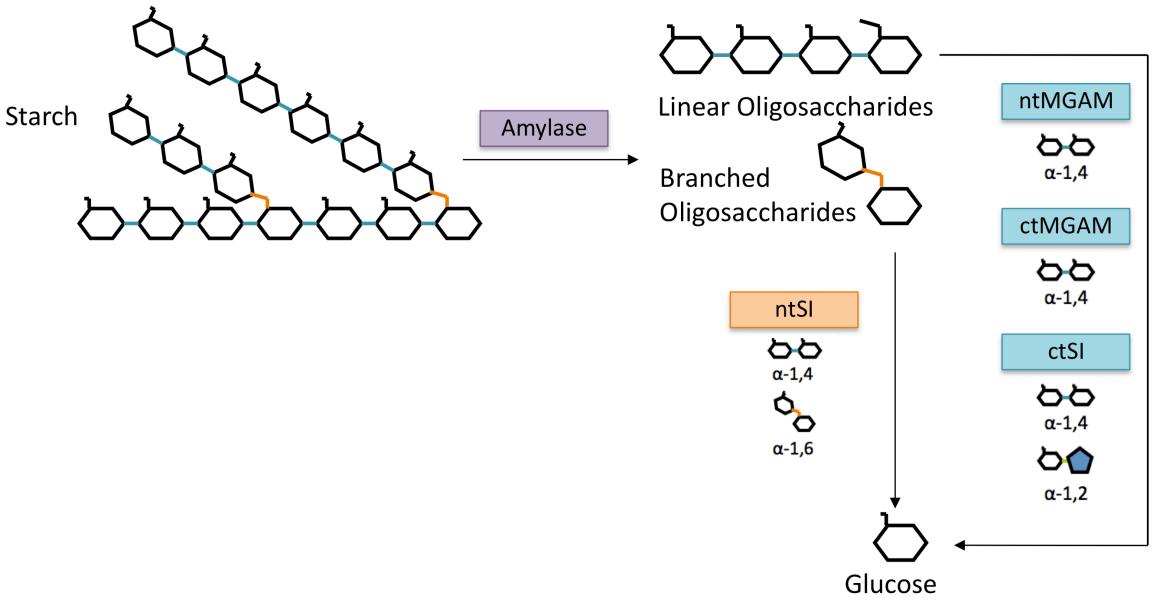

Figure 1-1: Schematic diagram of the process of starch digestion by α-amylase, maltase-glucoamylase, and sucrase isomatase.

SI demonstrates hydrolytic activity on branched α-1,6 linkages, which is complemented by the hydrolytic activity of both SI and MGAM on α-1,4 linkages (Quezada-Calvillo et al. 2008); however, the main substrate of SI is the α-1,4 linkage (Nichols et al. 2003). The complementary activities of the human enzymes allow for digestion of the spectrum of food starches from more than 400 plants that comprise two-thirds of most human diets; SI is more abundant but MGAM displays higher hydrolytic activities (Heymann and Günther 1994; Heymann, Breitmeier, and Günther 1995; Robayo Torres, Quezada Calvillo, and Nichols 2006). In studies of severely malnourished infants, it has been demonstrated that the terminal steps of MGAM α-1,4 and SI α-1,6 starch digestion are decreased. The loss of MGAM message parallels the reduction in villin message and degree of villous atrophy, and also parallels reduction in the SI message. It has also been shown that the correlations between the two enzyme activities (SI and MGAM) exist at the mRNA level (Nichols et al. 2000). Recent studies demonstrate that feeding rats a high fat/carbohydrate ratio diet reduces the jejunal sucrase/isomaltase (S/I) activity ratio. In addition, the SI-linked oligosaccharides show modifications of unsialylated galactose as well as the mannose-linked form when the S/I activity ratio changes, and thus is thought to be regulated by altering the glycosylation of SI (Samulitis-Dos Santos, Goda, and Koldovsky 1992;

Goda and Takase 1994; Goda, Yasutake, and Takase 1994; Honma, Mochizuki, and Goda 2007; Mochizuki, Igawa-Tada, et al. 2010c). These studies emphasize the importance of MGAM and SI activities, regulated both by expression levels and enzyme structure, in human nutrition. Interestingly, protein expression regulated by epigenetic changes resulting from human prenatal malnutrition has been associated with a number of disease conditions (Lillycrop et al. 2007; Heijmans et al. 2008; Jones et al. 2011).

Both MGAM and SI are anchored to the brush-border epithelial cells of the small intestine and each contains two catalytic subunits (Figure 1-2): an N-terminal subunit (ntMGAM and ntSI) adjacent to the membrane-bound end of the enzyme and a C-terminal luminal subunit (ctMGAM and ctSI). Each catalytic subunit acts as a Family 31 Glycoside Hydrolase (GH31) (Cantarel et al. 2009), which operate through a mechanism that results in retention of configuration at the anomeric center (Figure 1-3). SI and MGAM show 59% amino acid sequence similarity and it has been proposed that the shared domain structures imply that MGAM and SI evolved through duplication of an ancestral gene, which itself had already undergone tandem gene duplication (Nichols et al. 2003). Further, the catalytic subunits of MGAM and SI are 40–60% identical in amino acid sequence and contain the catalytic site signature sequence WiDMNE (i indicates a variable residue), characteristic of GH31 subgroup 4 members (Figure 1-4) (Ernst et al. 2006). Each of these proteins includes a small cytosolic domain, a transmembrane domain (TMD), an O-glycosylated linker (O-link), and two homologous catalytic subunits (ntMGAM, ctMGAM, ntSI, and ctSI). Additionally, there are various alternative splicing patterns of ctMGAM in mammals; two spliceforms from mice are discussed in this study: ctMGAM-N2 and ctMGAM-N20 (Jones et al. 2011). ctMGAM-N20 has an 89-amino acid insertion not present in ctMGAM-N2 nor the other catalytic subunits. All three C-terminal catalytic subunits have a 21-amino acid insertion near the catalytic signature site (See alignment (Edgar 2004) in Figure 1-5).

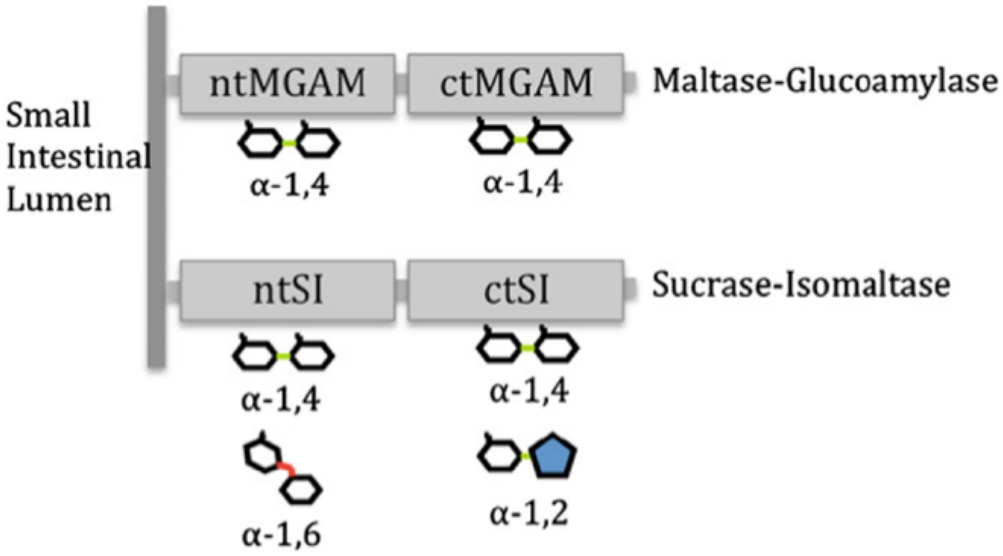

Figure 1-2: Schematic diagram of MGAM and SI indicating hydrolytic activity.

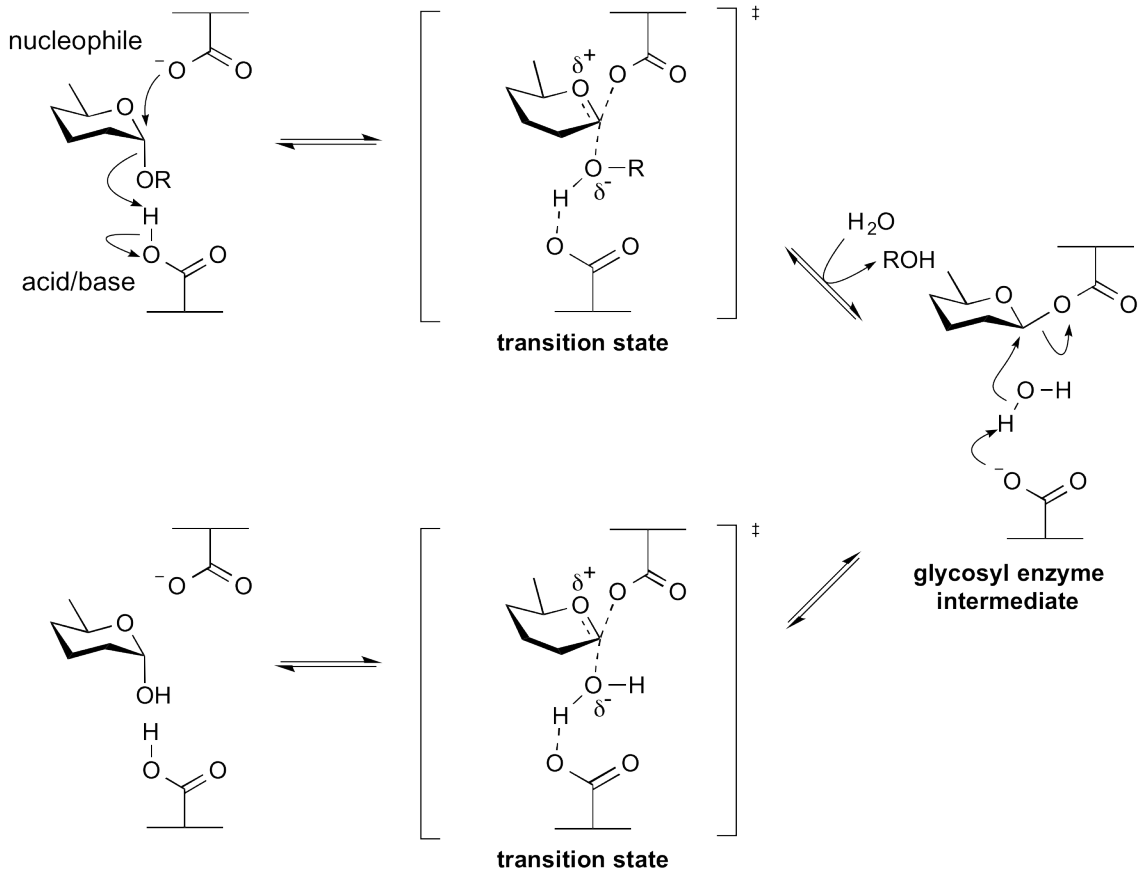

Figure 1-3: Retaining mechanism of Family 31 Glycoside Hydrolases . In the first step of the double-displacement mecnahnism, nucleophilic attack of the anomeric centre displaces the aglycon. This results in the formatiom of a glycosyl intermediate. The glycosidic oxygen is protonated by the acid catalyst. In the second step, the covalent glycosyl-intermediate is hydrolyzed by a water molecule that is deprotonated by the base catalyst. Both steps proceed through oxacarbenium ion-like transition states. (Withers and Williams 2014).

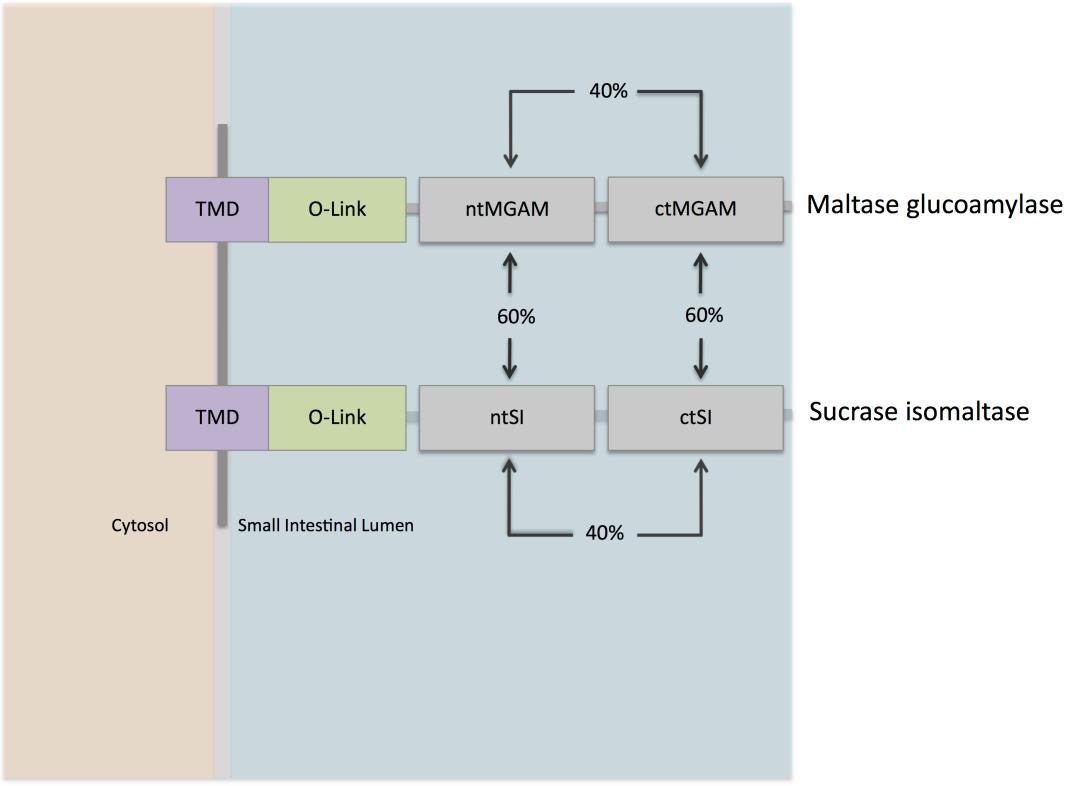

Figure 1-4: Linear representation of protein organization of SI and MGAM. Percentage of sequence identity of catalytic domains is indicated (adapted from Sim et al. 2008).

Figure 1-5: Sequence alignment of SI and MGAM catalytic subunits.

1.2 Control of Blood Glucose Levels with Novel Inhibitory Compounds

Until recently, there has been little investigation into the individual roles of the four catalytic subunits comprising the human intestinal MGAM and SI complexes in the digestion of food starches. Currently, inhibitors targeting intestinal α-glucosidases as well as pancreatic α-amylases are used in the treatment of non-insulin-dependent Type II diabetes. Inhibition of these enzymes decreases postprandial hyperglycemia and hyperinsulinemia, reducing the occurrence of insulin resistance as well as reducing the stress on the beta cells of the pancreas, preventing further insulin-dependent disorders (Jacob 1995; Holman, Cull, and Turner 1999; Chiasson et al. 2002; Scheen 2003; Simpson, Shaw, and Zimmet 2003). Presently, there are two key α-glucosidase inhibitors that have demonstrated value as dietary and treatment supplements for patients with Type II diabetes. These are acarbose and miglitol, which inhibit the enzymes MGAM and SI. These compounds exhibit a broad affinity for α-glucosidases, such as α-amylases as well as the target enzymes, leading to a requirement for relatively large doses, resulting in unpleasant side effects such as gastrointestinal upset. These problems can be mitigated through targeting the core activities that contribute to postprandial glycemic excursions (Godbout and Chiasson 2007; Jones et al. 2011).

The crystal structures of human ntMGAM in complex with inhibitors including acarbose revealed the structural basis for its poor inhibitory properties (Sim et al. 2008; Sim et al. 2009; Sim et al. 2010). These structural results in conjunction with sequence analysis indicate ctMGAM and ctSI have features that extend their substrate-binding clefts to more subsites than in ntMGAM (Figure 1-6). Novel compounds that can distinguish each of the four enzyme subunits of MGAM and SI will be important tools for understanding their individual roles, but also in limiting side effects resulting from unnecessary inhibitory activities (Jones et al. 2011).

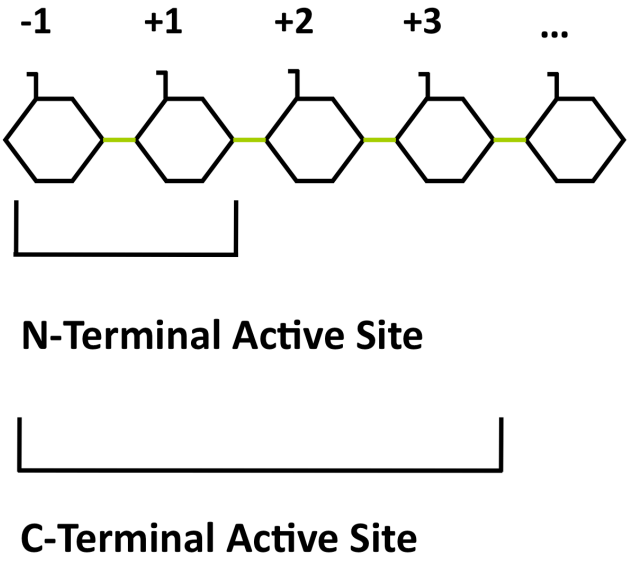

Figure 1-6: Schematic of N-terminal active site and the hypothesized extended C-terminal active site.

In the Ayruvedic medicinal system, extracts from the root and stem of *Salacia reticulata* have been traditionally used to treat Type II diabetes. *Salacia reticulata* is a large, woody climbing plant that is native to Sri Lanka and southern India (Chandrasena 1935; Jayaweera 1981; Vaidyartanam 1993). Human clinical trials on patients with Type II diabetes have been conducted and shown to be effective treatments of Type II diabetes with minimal side effects (Jayawardena et al. 2005). The active compounds of *S. reticulata* were found to be a novel class of sulfonium-ion glucosidase inhibitors, including salaprinol (Yoshikawa et al. 2008), salacinol (Yoshikawa et al. 1997; Yuasa, Takada, and Hashimoto 2000; Ghavami, Johnston, and Pinto 2001), ponkoranol (Johnston, Jensen, and Pinto 2006; Yoshikawa et al. 2008), kotalanol (Matsuda et al. 1998; Jayakanthan, Mohan, and Pinto 2009; Eskandari, Jayakanthan, et al. 2010a), de-O-sulfonated kotalanol (Muraoka et al. 2008; Oe and Ozaki 2008; Ozaki, Oe, and Kitamura 2008; Jayakanthan, Mohan, and Pinto 2009), de-O-sulfonated salacinol (Minami et al. 2008; tanabe et al. 2009), de-O-sulfonated ponkoranol (Eskandari, Kuntz, et al. 2010c; Xie et al. 2011) and de-O-sulfonated salaprinol (Xie et al. 2011), the structures of which comprise a 1,4-anhydro-4- thio-D-arabinitol core and polyhydroxylated acyclic chain (Eskandari, Jones, Ravinder Reddy, et al. 2011a). The anti-diabetic property of these herbal extracts can be attributed, at least in part, to the inhibition of intestinal α-glucosidases by the sulfonium-ion

compounds (Yoshikawa et al. 1997; Ozaki, Oe, and Kitamura 2008; Eskandari, Jones, Rose, and Pinto 2011b) This novel class of α-glucosidase inhibitors is a putative mimic of the proposed intermediate of glucosidase-mediated hydrolysis, the oxacarbenium transition state (Eskandari, Jones, Rose, and Pinto 2011b; Mohan, Sim, Rose, and Pinto 2010; Mohan and Pinto 2007). Thus, collaborators in the Pinto laboratory have successfully synthesized the naturally occurring glucosidase inhibitors salacinol (Ghavami, Johnston, and Pinto 2001), ponkoranol (Johnston, Jensen, and Pinto 2006), kotalanol (Jayakanthan, Mohan, and Pinto 2009), and de-O-sulfonated kotalanol (Jayakanthan, Mohan, and Pinto 2009). An additional compound that was recently synthesized, de-O-sulfonated ponkoranol (Eskandari, Kuntz, et al. 2010c), was subsequently isolated from the same plant (Xie et al. 2011; Eskandari, Jones, Rose, and Pinto 2011b)

1.3 Role of Sucrase Isomaltase in Congenital Sucrase Isomaltase Deficiency (CSID)

The two subunits of SI are implicated in the disease state of congenital sucrase-isomaltase deficiency (CSID), a rare, autosomal, intestinal disease in which patients experience a range of phenotypes from mild decrease to a complete absence of sucrase (ctSI) activity. There is also a marked decrease in isomaltase (ntSI) activity in some CSID patients. Symptoms of CSID include severe diarrhea after ingestion of disaccharides and oligosaccharides, often resulting in malnutrition and failure to thrive in infants and young children (Jacob et al. 2000). These symptoms are typically first identified when sucrose is introduced into the infant diet and carry on into adulthood, although a tolerance to sucrose is often seen in older patients due to the growth of the gut, increasing the absorptive area of the gut (Sterchi, Lentze, and Naim 2004).

CSID is frequently misdiagnosed because symptoms are similar to celiac disease, cystic fibrosis, and other food allergies and intolerances. Originally, CSID was diagnosed based on eliminating sucrose from the diet as well as monitoring blood glucose levels in response to sucrose ingestion. This type of diagnosis resulted in a 25% false-positive response rate (Krasilnikoff, Gudmand-Hoyer, and Moltke 1975). Presently, biopsies of the small bowel followed by sucrase enzyme assays have become the standard method for diagnosis of CSID, but this methodology is highly invasive. A [^{13}C]-sucrose labeled breath test has also been developed in which a small dose of labeled substrate is given to the patient and oxidation of [^{13}C]-sucrose

is monitored (Robayo-Torres et al. 2009). This breath test has been correlated to the small bowel biopsy results and validated (Nichols and Auricchio 2012).

Many factors contribute to the degree of severity of symptoms presenting in patients with CSID. These include the amount of carbohydrates ingested as well as the other meal components, the rate of gastric emptying and small-bowel transit, the degree of residual enzymatic activity of ntSI and ctSI catalytic subunits, the adsorption of the colon, and the fermentation of unabsorbed carbohydrates in the colon by bacteria (Naim, Heine, and Zimmer 2012). The tolerance to sucrose varies substantially from one patient to the next, making one treatment option, life-long diet restriction, difficult to impose, as the degree of diet restriction is difficult to prescribe. An additional challenge is that isomaltase activity is commonly compromised as well, thus CSID patients are often required to exclude starch from the diet in addition to sucrose. Because it is virtually impossible to remove starch and sucrose from the diet effectively, enzymatic supplementation is a treatment option (Nichols and Auricchio 2012). A supplement known clinically as Sucraid is a liquid preparation produced by *Saccharomyces cerevisiae* and has been used as an effective treatment option in CSID patients (Lucke et al. 2009). This treatment was developed and marketed by QOL Medical.

CSID is most commonly seen in Canadian Aboriginals (3-7%), native Alaskans (3%) and Greenland Eskimos (5-10%), whereas the occurrence is relatively low in the general North American population (0.2%) (Bell, Draper, and Bergan 1973; Ellestad-Sayed, Haworth, and Hildes 1978). Due to misdiagnosis, lack of awareness of CSID, and similarities to other digestive disorders, CSID is not commonly diagnosed even in young children. The clinical symptoms of CSID vary significantly and multiple mutant phenotypes of SI have been found to be responsible for the onset of CSID, leading to the hypothesis that CSID is more common than previously thought and affected patients are not being diagnosed. Several mutant phenotypes of SI have been identified that vary in their posttranslational processing, cellular localization, and enzymatic functionality (Alfalah et al. 1999; Jacob et al. 2000; Spodsberg 2001; Ritz et al. 2003; Keiser et al. 2006; Alfalah et al. 2009). SI is synthesized in the rough endoplasmic reticulum (ER), transported to the Golgi apparatus where N-linked and O-linked glycosylation occurs, and sorted to the apical membrane. Full length SI is cleaved by luminal pancreatic proteases into its two catalytically active subunits: ntSI (isomaltase) and ctSI (sucrase) in the apical membrane (Hauri, Quaroni, and Isselbacher 1979; Naim, Heine, and Zimmer 2012).

Current research suggests that an inactive form of SI is present in CSID patients (Freiburghaus et al. 1977), and further studies have demonstrated trafficking defects and intracellular accumulation of SI (Hauri et al. 1985). Mutations have also been indentified from biopsies of CSID patients and grouped into seven distinct phenotypes (Naim, Sterchi, and Lentze 1988). Interestingly, all of the patients in this study presented with similar symptoms, but the mutations in the two catalytically active subunits varied (Naim, Sterchi, and Lentze 1988; Naim, Heine, and Zimmer 2012). Hypotheses about the causative effects of CSID-associated mutations in SI are presented in Chapter 4.

Chapter 2
Characterizing inhibition profiles of ctSI, ctMGAM, ntSI, & ntMGAM

2.1 Chapter Overview

Here, we compare the kinetic effect on the four catalytic subunits by fourteen inhibitors: two of the currently available antidiabetic drugs, acarbose (Balfour and McTavish 1993) and miglitol (Johnston et al. 1994); a novel class of sulfonium-ion glucosidase inhibitors, salacinol (Ghavami, Johnston, and Pinto 2001; Johnston, Jensen, and Pinto 2006), kotalanol (Jayakanthan, Mohan, and Pinto 2009), de-*O*- sulfonated kotalanol (Jayakanthan, Mohan, and Pinto 2009), de-*O*-sulfonated ponkoranol (Eskandari, Kuntz, et al. 2010c), and 5' stereoisomer of de-*O*-sulfonated ponkoranol (Eskandari, Kuntz, et al. 2010c); the selenium analogue of salacinol, blintol (Liu and Pinto 2005); selenium analogues of de-*O*-sulfonated ponkoranol, and its 5' stereoisomer (Eskandari, Jones, Rose, and Pinto 2011b); a derivative of ponkoranol, 3'-*O*-methyl ponkoranol (Eskandari, Jones, Rose, and Pinto 2011c); and finally maltoside derivatives of de-*O*-sulfonated ponkoranol, 3'-*O*-β-maltosyl de-*O*-sulfonated ponkoranol and 5'-*O*-β-maltosyl de-*O*-sulfonated ponkoranol (Eskandari, Jones, Ravinder Reddy, et al. 2011a) (Figure 2-11). The results demonstrate catalytic domain selectivity, helping differentiate the active site requirements of each catalytic subunit. This directly enhances the understanding of the role these enzymes play in glucogenesis of starch *in vivo* (Jones et al. 2011).

2.2 Contributions

All work was performed by the candidate, with the following exceptions. The synthesis of inhibitors was completed by Dr. B. Mario Pinto's laboratory (R. Eskandari, S. Mohan, K. Jayakanthan, H. Jensen, K.R. Reddy, B.D. Johnston, A. Ghavami). Lyann Sim purified a portion of the protein used in assays and her previous structural studies helped guide inhibitor development. Assays involving linear malto-oligosaccharides and branched α-limit dextrins were performed in Bruce Hamaker's laboratory (B-H Lee). Sections of this chapter are modified from or taken verbatim from text written by the candidate in the following publications:

K. Jones, L. Sim, S. Mohan, K. Jayakanthan, H. Lui, S. Avery, H.Y. Naim, R. Quezada-Calvillo, B.L. Nichols, B.M. Pinto, D.R. Rose. Bioorganic & Medicinal Chemistry. 19 (2011) 3929-3924.

R. Eskandari, K. Jones, D.R. Rose, B.M. Pinto. Bioorganic & Medicinal Chemistry Letters. 20 (2010) 5686-5689.

R. Eskandari, K. Jones, D.R. Rose, B.M. Pinto. Bioorganic & Medicinal Chemistry Letters. 21 (2011) 6491-6494.

R. Eskandari, K. Jones, D.R. Rose, B.M. Pinto. Chemical Communications 47 (2011) 9134-9136.

R. Eskandari, K. Jones, K.R. Reddy, K. Jayakanthan, M. Chaudet, D.R. Rose, B.M. Pinto. Chemistry – A European Journal. 17 (2011) 14817-14825.

B-H Lee, R. Eskandari, K. Jones, K.R. Reddy, R. Quezada-Calvillo, B.L. Nichols, D.R. Rose, B.R. Hamaker, B.M. Pinto. The Journal of Biological Chemistry. 287(38) (2012) 31929-31938.

2.3 Materials and Methods

2.3.1 Protein Expression and Purification

2.3.1.1 N-terminal MGAM and SI

2.3.1.1.1 Drosophila Expression System

Cloning, protein expression and purification for a portion of the kinetic assays required for characterization of ntMGAM and ntSI was completed by Elena Rossi and Lyann Sim (Rose Lab), described in detail in previous publications (Ghavami, Johnston, and Pinto 2001; Rossi et al. 2006; Sim et al. 2008) In summary, the N-terminal domain of MGAM was cloned into a Drosophila pMT-BiP-V5-His vector, permitting secretion of ntMGAM into the culture media with a C-terminal hexa-histidine tag for purification. ntMGAM was then purified by copper-chelating and anion-exchange chromatography (Scotter et al. 2006; Jones et al. 2011).

Cloning, expression, and purification of ntSI from Drosophila S2 cells was completed by Lyann Sim (Rose Lab) and reported in Sim et al. 2010. In summary, residues 62–931 of full-length human SI (GenBankTM accession number NP_001032) were cloned into a Drosophila pMT-TEVA expression vector (Scotter et al. 2006) allowing for secretion of ntSI into the culture medium with an N-terminal hexa-histidine tag for purification. ntSI was expressed in Drosophila S2 cells using a similar expression protocol as for ntMGAM (Ghavami, Johnston, and Pinto 2001; Sim et al. 2008; Jones et al. 2011).

2.3.1.2 Baculovirus Expression System

Using the constructs previously used in drosophila cell expression, the ntSI and ntMGAM inserts were amplified by PCR using Phusion High-Fidelity polymerase (New England Biolabs) and ligated into the pAcGP67-His vector in-frame with the start codon. The two N-terminal constructs were sequenced to ensure sequence accuracy and transformed into *Escherichia coli* DH5α. Colonies were screened using the antibacterial resistance conferred by the pAcGP67-His plasmid and the N-terminal constructs were harvested using the QIAprep Spin Miniprep kit (Qiagen). The constructs were named pAcGP67B-His-ntSI and pAcGP67B-His-ntMGAM. The two N-terminal constructs were individually co-infected with linear wild-type baculovirus DNA (BD Biosciences) into *Spodoptera frugiperda* isolate 9 (Sf-9) insect cells. Viral isolates were selected

based on screening secreted protein from each isolate for maltase enzyme activity with the previously described glucose oxidase assay (Sim et al. 2008). Selected viral isolates were amplified to titred Passage 2 (P2) viral stocks.

The viral stocks were used to infect *Spodoptera frugiperda* isolate 9 (Sf-9) cells (GIBCO, Carlsbad, CA, USA) grown in Sf-900 III Serum-free Media (GIBCO, Carlsbad, CA, USA) to a density of approximately 2.5 x 10^6 cells/mL at a multiplicity of infection (MOI) of 3.0. The cells were allowed to grow at a temperature of 27°C shaking at 130 rpm until viability reached approximately 75%, at which point the culture medium was harvested by centrifugation for 10 minutes at 1500 x *g*.

The collected culture media from each infection containing the secreted ntMGAM and ntSI proteins was filtered using a Bottle-top Vacuum Filtration System (0.2 µm, VWR) and purified by affinity chromatography. HisPur Nickel-NTA Resin (Pierce Protein Biology Products) was equilibrated in equilibration buffer (50 mM monosodium phosphate, 300 mM NaCl, pH 7.0). The equilibrated resin was used to bind the protein by batch treatment, adding the resin to the harvested media and allowed to mix gently overnight at 4°C. The supernatant was decanted by passing the slurry through a BioRad column (2.5 cm diameter). The resin was washed with 25 column volumes of wash buffer (50 mM monosodium phosphate, 300 mM NaCl, 25 mM imidazole, pH 7.0) and the protein was eluted stepwise with elution buffer (50 mM monosodium phosphate, 300 mM NaCl, pH 7.0) with increasing concentrations of imidazole (50 mM, 100 mM, and 200 mM). Unfortunately, the protein did not effectively bind to the column and the majority of the N-terminal protein was eluted from the column during the 25 mM wash. Despite the low affinity of the N-terminal proteins to the Ni-NTA resin, it was recoverable by concentrating the collected flow-through during the wash stage of the purification, concentrating the protein, and exchanging the protein into a Tris buffer (20 mM Tris, 50 mM NaCl, pH 7.0) to rid the samples of imidazole using a centrifugal concentrator (Amicon). This was repeated multiple times and the protein was then further concentrated to ~10 mg/mL. The protein was flash frozen by submersion in liquid nitrogen in 50-100 µl aliquots and stored at -80°C. Figure 2-1 shows the resulting reducing SDS-PAGE gels from ntMGAM and ntSI nickel column purifications. Overall, the proteins demonstrated low affinity for the Ni-NTA resin. After a great deal of troubleshooting, it was discovered that there was an error in the sequence of the pAcGP67B-His vector, causing the HIS-tag to be out of frame with the vector's start codon. Thus,

the protein was expressed but with no HIS-tag. Fortunately, enough protein was recovered to perform some biochemical assays with a yield of approximately 1 mg/L. The recovered protein was tested for maltase activity using the glucose oxidase assay reported previously (Sim et al. 2008). The ntSI and ntMGAM sequences were later re-cloned for baculovirus expression (unpublished, work completed by Jin Duan, Rose Lab).

Table 2-1: Properties of expressed recombinant proteins.

Protein	Size	pI	Yield (/L of culture)
ctSI	~102 kDa (Migrates at ~120 kDa due to glycosylation)	6.14	7 mg
ctMGAM-N2	~120 kDa (Migrates at ~130 kDa due to glycosylation)	5.99	3 mg
ctMGAM-N20	~130 kDa (Migrates at ~150 kDa due to glycosylation)	5.65	5 mg
ntSI	~102 kDa (Migrates at ~100kDa)	5.50	1 mg
ntMGAM	~98 kDa (Migrates at ~100 kDa)	5.17	1 mg

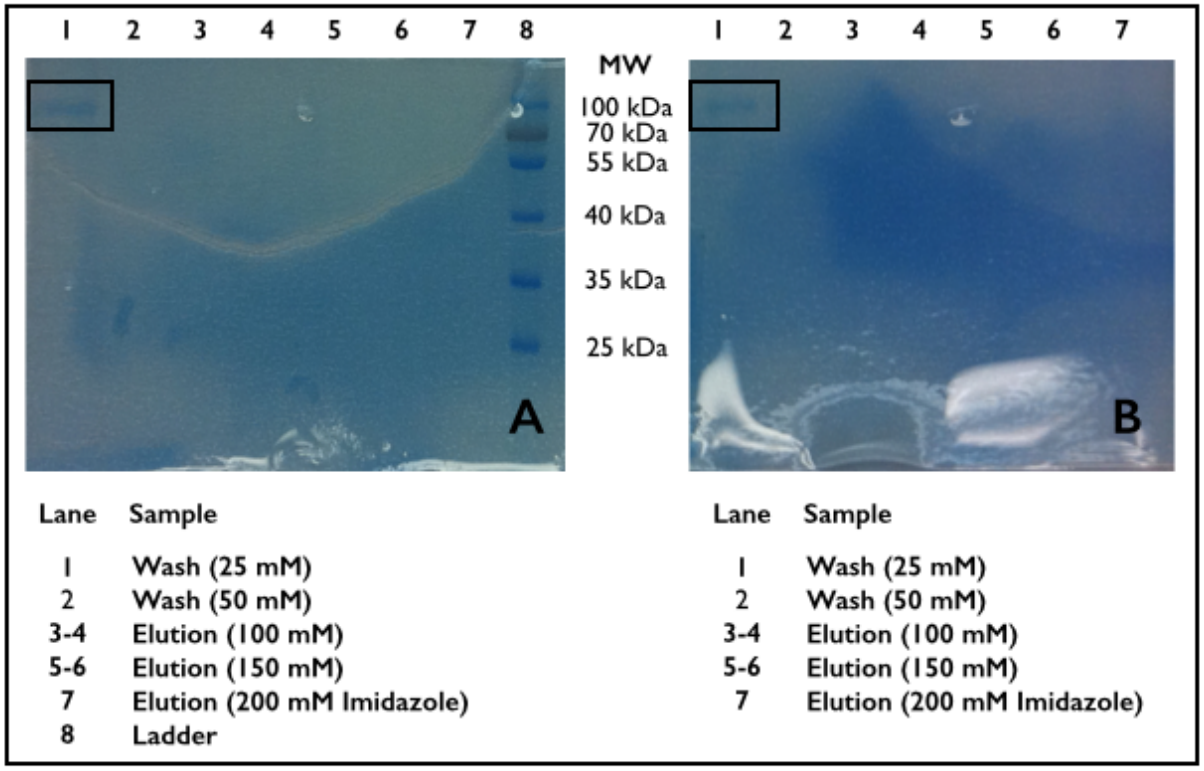

Figure 2-1: ntSI and ntMGAM Nickel-NTA affinity (SDS-PAGE). The above reducing SDS-PAGE gel (12%) illustrates the lack of affinity of ntSI and ntMGAM to the Nickel-NTA resin due to the lack of an expressed histidine tag as the protein did not bind to the column.

2.3.1.3 C-terminal MGAM and SI

ctMGAM and ctSI were originally cloned, expressed and purified by Stephen Avery and Buford Nichols and reported (Jones et al. 2011). In summary, nucleotides 2432-6215 from Genbank sequence for MGAM (EU073529) and nucleotides 2645-5373 from GenBank sequence for SI (EU937530) were amplified by PCR, purified by gel extraction, and ligated into pT7-blue T-vector. The constructs were sequenced to select error free inserts (Applied Biosystems, 373A automated sequencer), which were ligated into pAcGP67 baculovirus vectors (BD Biosciences), modified to ten histidine residues at the 3' end of expressed proteins (pAcGP67B_His-1). The constructs (pAcGP67 His-CtSI and pAcGP67 His-ctMGAM) were sequenced and separately co-infected with linearized wild type baculovirus DNA (BD Biosciences) into *Spodoptera frugiperda* isolate 9 (Sf-9) insect cells. Isolates from each co-infection were obtained by plaque purification.

Secreted proteins were screened for maltase enzyme activity with the glucose oxidase assay described previously (Sim et al. 2008). Viral isolates were selected based on the screening assay results and amplified to titred Passage 2 (P2) viral stocks (Jones et al. 2011).

The viral stocks were used to infect *Spodoptera frugiperda* isolate 9 (Sf-9) cells (GIBCO, Carlsbad, CA, USA) grown in Sf-900 III Serum-free Media (GIBCO, Carlsbad, CA, USA) to a density of approximately 2.5 x 10^6 cells/mL at a multiplicity of infection (MOI) of 3.0. The cells were allowed to grow at a temperature of 27°C shaking at 130 rpm until viability reached approximately 75%, at which point the culture medium was harvested by centrifugation for 10 minutes at 1500 x *g*.

The C-terminal proteins were expressed using Sf-9 insect cells grown in SF-900 III Serum-free Media. Culture media containing the secreted protein was filtered and purified by affinity chromatography as described in section 2.2.1.2. Following the affinity chromatography, the samples were exchanged into 20 mM Tris pH 8.0 and further purified using fast protein liquid chromatography (FPLC) (BioRad). Anion exchange chromatography using the BioRad Uno Q1 Column was performed. The protein was eluted using a gradient of 1.0 M NaCl, with ctSI eluting from the column at a concentration of approximately 0.1M NaCl and ctMGAM-N2 and ctMGAM-N20 eluting from the column at a concentration of approximately 0.6M NaCl. The fractions containing the elution peak were pooled, exchanged into 20 mM Tris, 50 mM NaCl pH 8.0 and further concentrated to ~10 mg/mL. The purified proteins divided into 50-100 uL volume aliquots, flash frozen by submersion in liquid nitrogen, and stored at -80°C. Four different c-terminal protein constructs were expressed (Table 2-1). The faint band on the SDS-PAGE gel after affinity chromatography at ~130 kDa was confirmed using His-Stain (Invitrogen) as well as glucose oxidase assays. ctSI, ctMGAM-N2, and ctMGAM-N20 were used in biochemical assays to determine the potency of various inhibitory compounds. Figure 2-2 and 2-3 illustrates a representative purification of ctSI and Figure 2-4 illustrates a summary of two ctMGAM-N2 and ctMGAM-N20 purifications.

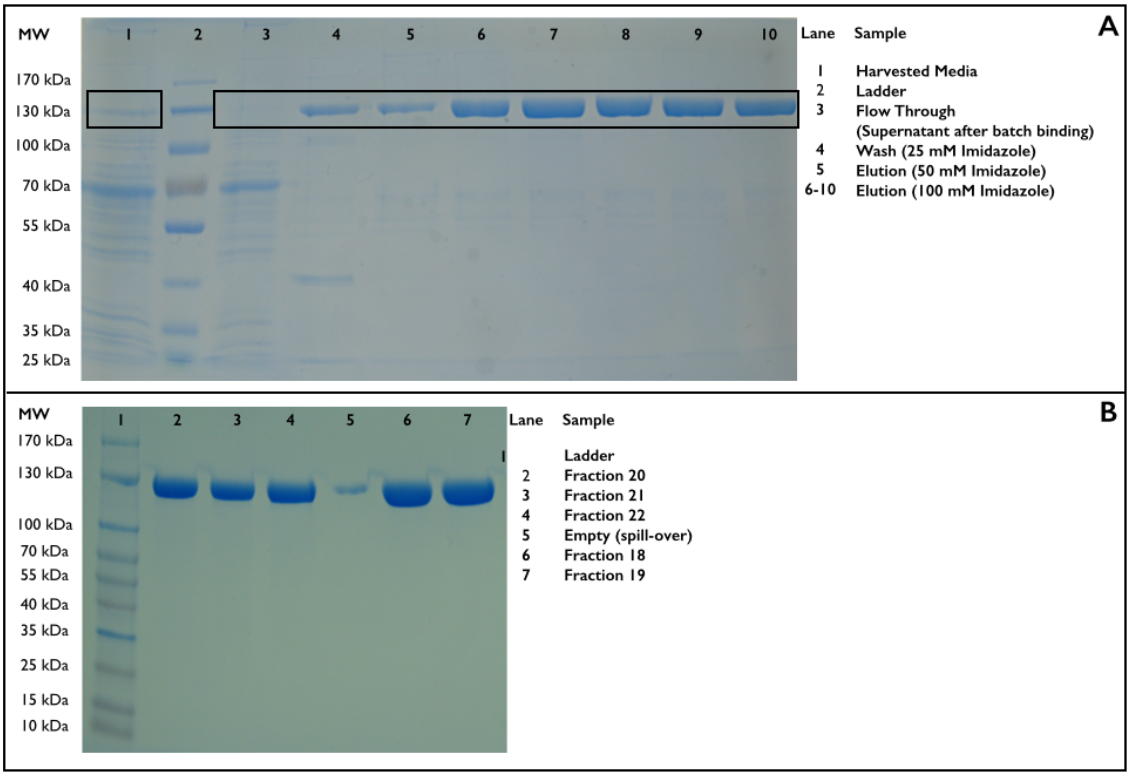

Figure 2-2: SDS-PAGE gels following Nickel-affinity chromatography (A) and anion-exchange chromatography (B) of ctSI.

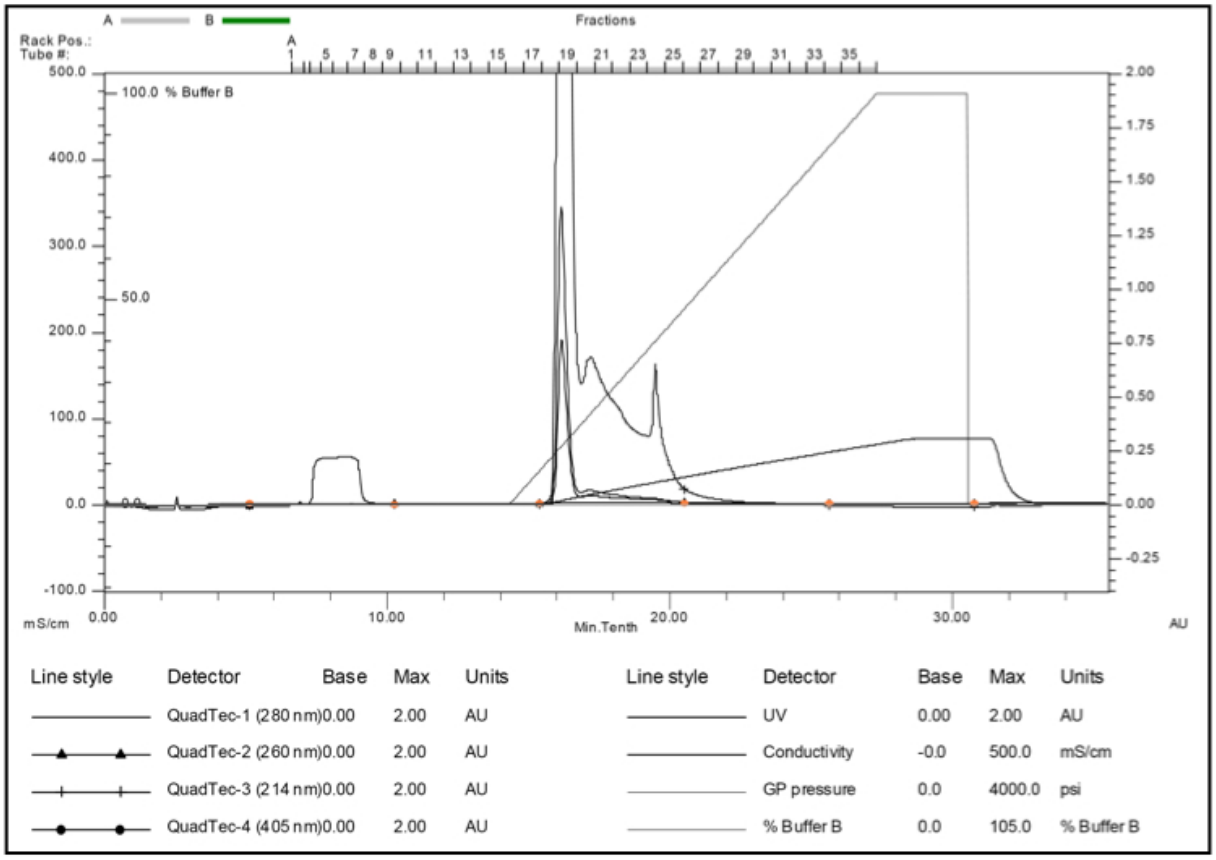

Figure 2-3: Chromatograph illustrating the separation of ctSI from contaminating proteins during anion-exchange chromatography. ctSI began to elute from the anion-exchange column at a concentration of approximately 0.1M NaCl.

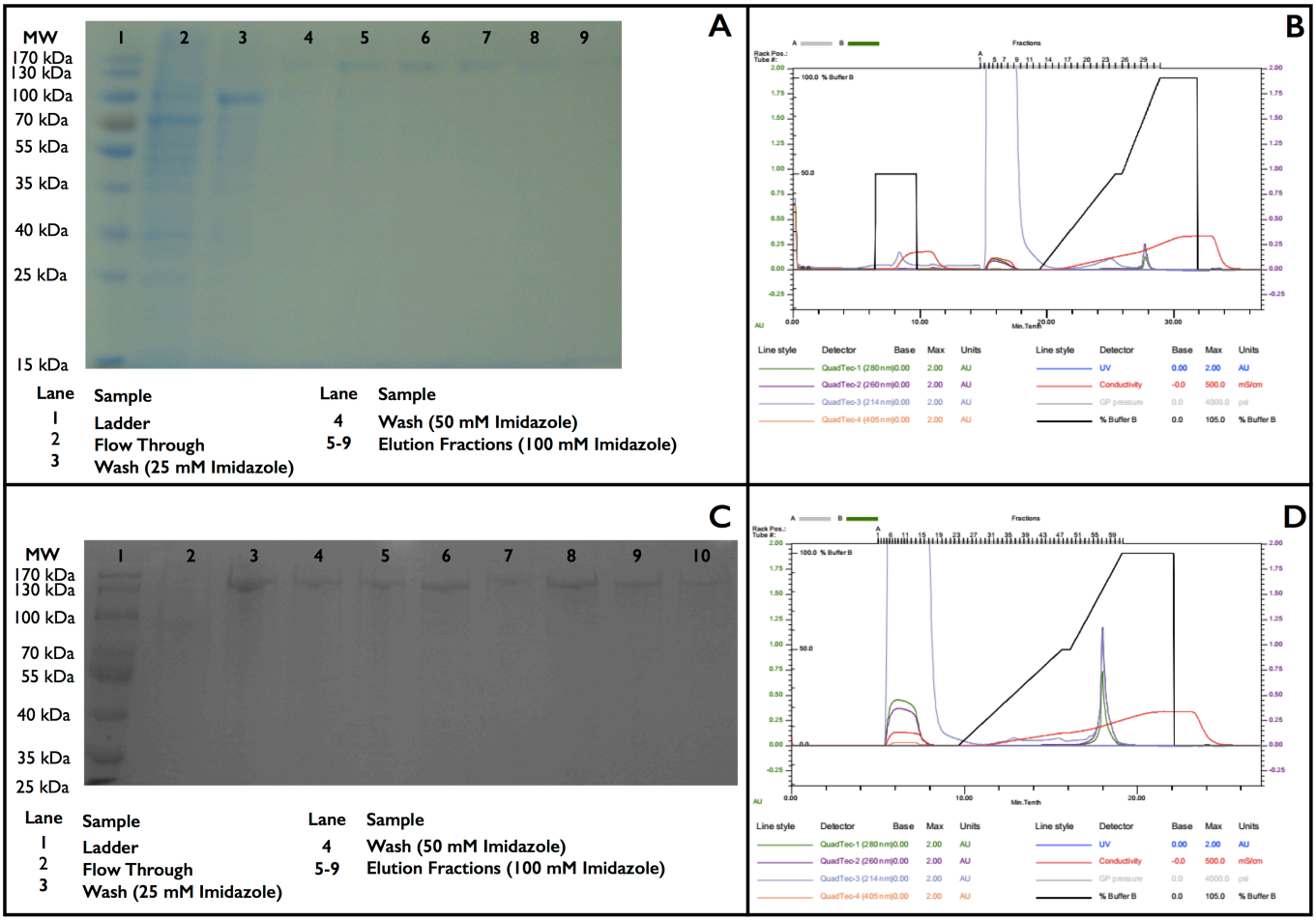

Figure 2-4: SDS-PAGE gel following Ni-affinity chromatography of ctMGAM-N2 (A) and chromatograph illustrating the separation of ctMGAM-N2 from contaminating proteins during anion-exchange chromatography (B). ctMGAM-N2 eluted from the anion exchange column at a concentration of approximately 0.6M NaCl. SDS-PAGE gel following Ni-affinity chromatography of ctMGAM-N20 (C) and chromatograph illustrating the separation of ctMGAM-N20 from contaminating proteins during anion-exchange chromatography (D). ctMGAM-N20 eluted from the anion exchange column at a concentration of approximately 0.6M.

2.3.2 Inhibitory Compounds

With the exception of acarbose and miglitol (Toronto Research Chemicals), synthesis of inhibitory compounds described in this chapter was completed by members of Dr. B. Mario

Pinto's laboratory at Simon Fraser University, specifically Sankar Mohan, Jayakanthan Kumarasamy, Kongara Ravinder Reddy, and Razieh Eskandari (Ghavami, Johnston, and Pinto 2001; Liu and Pinto 2005; Johnston, Jensen, and Pinto 2006; Jayakanthan, Mohan, and Pinto 2009; Eskandari, Jones, et al. 2010b; Eskandari, Kuntz, et al. 2010c; Eskandari, Jones, Ravinder Reddy, et al. 2011a; Eskandari, Jones, Rose, and Pinto 2011b; Eskandari, Jones, Rose, and Pinto 2011c) (Figure 2-5).

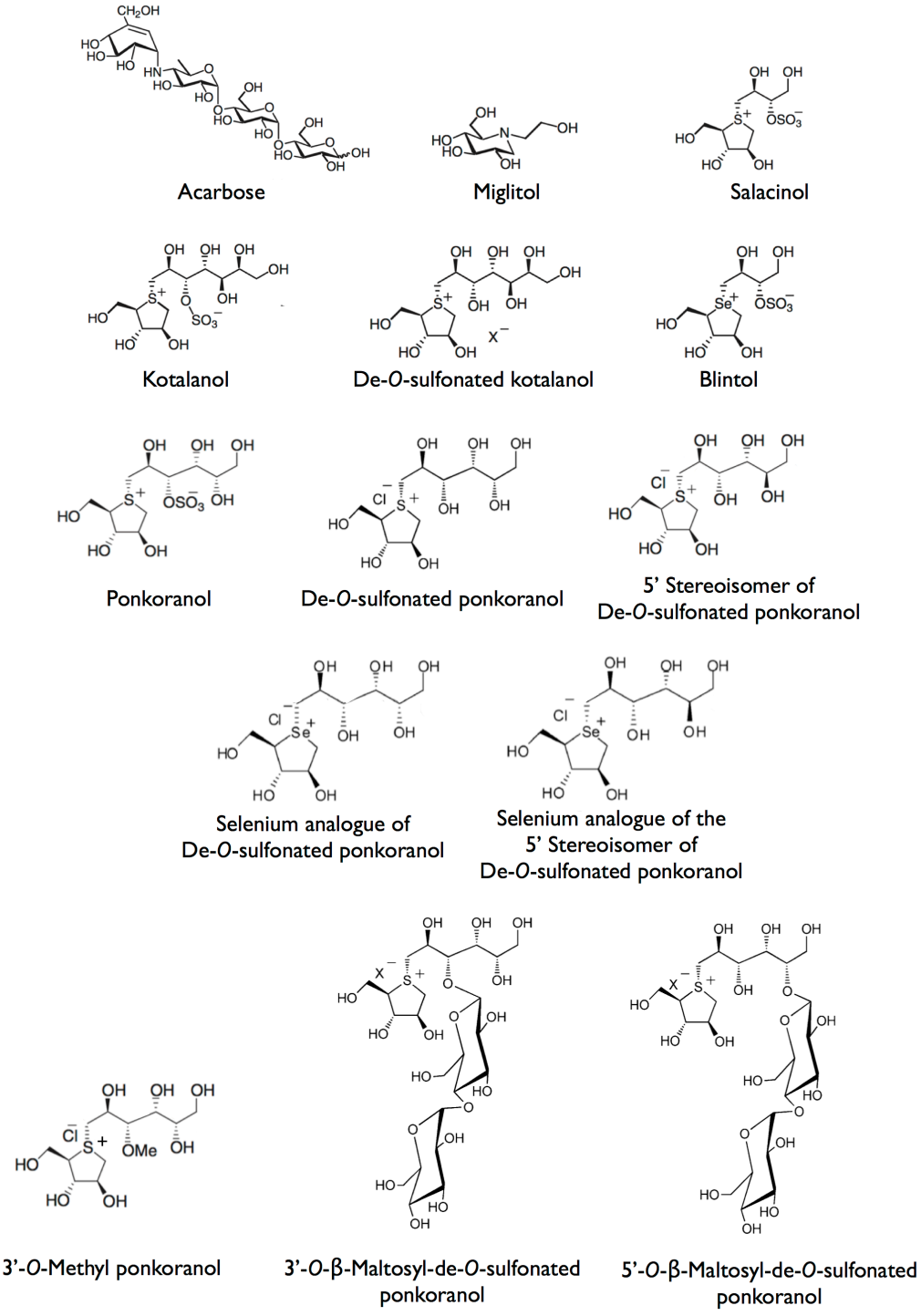

Figure 2-5: Inhibitory compounds.

2.3.3 Kinetic Characterization

Enzymatic activity of ntSI, ntMGAM, ctSI, ctMGAM-N2, and ctMGAM-N20 with maltose substrate in the presence of inhibitory compounds were determined using the end point glucose oxidase assay, previously reported (Dahlqvist 1964; Rossi et al. 2006; Nasi, Sim, Rose, and Pinto 2007a; Nasi, Sim, Rose, and Pinto 2007b; Sim et al. 2009). Briefly, maltose was provided as substrate in increasing concentration (0.5-30 mM) in 100 mM MES buffer at pH 6.5 for the enzyme (5.0×10^{-5} mg/mL ntMGAM and ntSI and 1.0×10^{-6} mg/mL ctSI, ctMGAM-N2, and ctMGAM-N20) in a 96-well microtitre plate. The plate was incubated at 37°C for 45 minutes and reactions were quenched with 3.0 M Tris, pH 6.9. The reactions were linear within the 45-minute reaction time frame. Liberated glucose was quantified using glucose oxidase reagent, incubating the reagent at 37°C for 30 minutes and measuring absorbance at 405 nm.

This assay quantifies the amount of glucose produced upon the addition of enzyme at increasing maltose concentrations (from 0.5 to 30 mM). All reactions were performed in quadruplicate, and absorbance measurements were averaged to give the final value. The program KaleidaGraph4.1 was used to fit the data to the Michaelis–Menten equation by non-linear regression to estimate K_m and V_{max} of the catalytic subunits. K_i values for each inhibitor were determined using the equation $K_i = [I]/((K_{mobs}/K_m) - 1)$, where K_{mobs} is the apparent K_m in the presence of the inhibitor. The inhibitors were characterized to determine the mechanism of inhibition, and competitive inhibition was determined to be the mode of inhibition for all inhibitors tested to date. The reported K_i values for each inhibitor were determined by averaging the K_i values from three different inhibitor concentrations (Jones et al. 2011).

2.4 Results

The inhibition constant (K_i) for each catalytic subunit with each of inhibitor studied was determined using the glucose oxidase assay with maltose as a substrate (Table 2-2).

Table 2-2: K_i values for each α-glucosidase inhibitor with ntMGAM, ctMGAM, ntSI, and ctSI with standard error indicated (μM). All ntMGAM and some ntSI values were previously determined and reported (Ghavami, Johnston, and Pinto 2001◆; Rossi et al. 2006*; Sim et al. 2010^)(Jones et al. 2011).

	ctMGAM-N2	ctMGAM-N20	ntMGAM	ctSI	ntSI
K_m (Maltose)	1.6 mM ± 0.18	1.9 mM ± 0.23	4.3 mM ± 1.2*	2.5 mM ± 0.5	7.1 mM ± 1.3^
Acarbose	0.009 ± 0.002	0.028 ± 0.005	62 ± 13*	0.246 ± 0.005	14 ± 1^
Salacinol	0.21 ± 0.018	0.058 ± 0.003	0.19 ± 0.02*	0.047 ± 0.008	0.28 ± 0.068
Blintol	0.02 ± 0.004	0.013 ± 0.002	0.49 ± 0.05◆	0.029 ± 0.005	0.16 ± 0.01
Kotalanol	0.064 ± 0.006	0.092 ± 0.007	0.19 ± 0.03*	0.042 ± 0.007	0.60 ± 0.06^
De-*O*-sulfonated kotalanol	0.026 ± 0.004	0.078 ± 0.009	0.03 ± 0.01*	0.026 ± 0.008	0.012 ± 0.001
Miglitol	0.211 ± 0.072	0.230 ± 0.076	1.0 ± 0.1*	0.130 ± 0.010	0.148 ± 0.011
De-*O*-sulfonated ponkoranol	No Inhibition	0.096 ± 0.015	0.043 ± 0.001*	0.10 ± 0.037	0.302 ± 0.123
3'-O-methyl ponkoranol	0.60 ± 0.015	0.055 ± 0.014	0.50 ± 0.04*	0.007 ± 0.002	0.035 ± 0.013
5' stereoisomer of de-*O*-sulfonated ponkoranol	No Inhibition	0.138 ± 0.068	0.015 ± 0.001	0.13 ± 0.047	0.14 ± 0.012
Selenium analogue of de-*O*-sulfonated ponkoranol	No Inhibition	0.047 ± 0.014	0.038 ± 0.008	0.018 ± 0.004	0.013 ± 0.008
Selenium analogue of 5' stereoisomer of de-*O*-sulfonated ponkoranol	No Inhibition	0.041 ± 0.027	0.025 ± 0.014	0.019 ± 0.010	0.010 ± 0.002
3'-*O*-β-Maltosyl de-*O*-sulfonated ponkoranol	No Inhibition	0.66 ± 0.063	0.039 ± 0.025	0.062 ± 0.005	0.046 ± 0.018
5'-*O*-β-Maltosyl de-*O*-sulfonated ponkoranol	0.077 ± 0.015	0.067 ± 0.012	0.008 ± 0.002	0.045 ± 0.001	0.019 ± 0.008

2.4.1 Acarbose

The current antidiabetic compound, acarbose, was found to be a micromolar inhibitor of both ntMGAM (K_i = 62 ± 13 μM) (Sim et al. 2010) and ntSI (K_i = 14 ± 1 μM) (Sim et al. 2010). In

contrast, acarbose is a ~1000-fold better inhibitor of ctMGAM-N2 (K_i = 0.009 ± 0.002 μM), ctMGAM-N20 (K_i = 0.028 ± 0.005 μM) and ~100-fold better for ctSI (K_i = 0.246 ± 0.005 μM). This supports the notion that the tetrasaccharide analogue occupies extended binding subsites in the C-terminal catalytic subunits as compared to the N-terminal domains (Rossi et al. 2006; Sim et al. 2008; Sim et al. 2010; Jones et al. 2011).

2.4.2 Salacinol

Salacinol is a submicromolar inhibitor of ntMGAM (K_i = 0.19 ± 0.02 μM) (Rossi et al. 2006) and ntSI (K_i = 0.277 ± 0.068 μM). The inhibition is about fivefold better for ctMGAM-N20 (K_i = 0.058 ± 0.003 μM) and ctSI (K_i = 0.047 ± 0.008 μM). Interestingly, salacinol exhibits some selectivity (fourfold) against ct MGAM-N2 as compared to the N20 splice form (Jones et al. 2011).

2.4.3 Blintol

Surprisingly, blintol, the Se analogue of salacinol, shows an improved inhibition of ctMGAM-N2 (K_i = 0.018 ± 0.004 μM), with little change against the other enzymes in comparison to salacinol: ctMGAM-N20 (K_i = 0.013 ± 0.002 μM), and ctSI (K_i = 0.029 ± 0.005 μM), ntMGAM (K_i = 0.49 ± 0.05 μM) (Ghavami, Johnston, and Pinto 2001) and ntSI (Ki = 0.16 ± 0.01 μM) (Jones et al. 2011).

2.4.4 Kotalanol

Kotalanol generally shows some selectivity (4-to 6-fold) for inhibiting C-terminal enzymes, as compared to N-terminal: ctMGAM-N2 (K_i = 0.064 ± 0.006 μM), ctMGAM-N20 (K_i = 0.092 ± 0.007 μM), and ctSI (K_i = 0.042 ± 0.007 μM) over ntMGAM (K_i = 0.19 ± 0.03 μM) and ntSI (K_i = 0.60 ± 0.06 μM) (Sim et al. 2010; Jones et al. 2011).

2.4.5 De-*O*-sulfonated kotalanol

De-*O*-sulfonated kotalanol demonstrates very similar inhibitory action against all five enzymatic subunits, ntSI (K_i = 0.012 ± 0.0006 μM), ctMGAM-N2 (K_i = 0.026 ± 0.004 μM), ctMGAM-N20 (K_i = 0.078 ± 0.009 μM), ctSI (K_i = 0.026 ± 0.008 μM), and ntMGAM (K_i = 0.03 ± 0.01 μM) (Rossi et al. 2006; Jones et al. 2011).

2.4.6 Miglitol

Miglitol, another compound currently used as an α-glucosidase inhibitor in patients with Type II Diabetes, is a high micromolar inhibitor with little distinction between the catalytic subunits.

With respect to ntMGAM, the K_i value is approximately fourfold higher than the other four enzymes (Jones et al. 2011).

2.4.7 De-*O*-sulfonated ponkoranol

De-*O*-sulfonated ponkoranol demonstrates no inhibition against ctMGAM-N2, but interestingly is a micromolar inhibitor of ctMGAM-N20 (K_i = 0.096 ± 0.015). There is little discrimination between the other catalytic subunits: ntMGAM (K_i = 0.043 ± 0.001), ctSI (K_i = 0.103 ± 0.037), and ntSI (K_i = 0.302 ± 0.123) (Eskandari, Jones, Rose, and Pinto 2011b).

2.4.8 5' stereoisomer of de-*O*-sulfonated ponkoranol

The 5' stereoisomer of de-*O*-sulfonated ponkoranol is similar to de-*O*-sulfonated ponkoranol in that it differentiates between ctMGAM-N2 (no inhibition) and ctMGAM-N20 (K_i = 0.138 ± 0.068). The inhibition seen with this inhibitor and ctMGAM-N20 is comparable with ctSI (K_i = 0.096 ± 0.015) (Eskandari, Jones, Rose, and Pinto 2011b).

2.4.9 Selenium analogue of de-*O*-sulfonated ponkoranol and Selenium analogue of 5' stereoisomer of de-*O*-sulfonated ponkoranol

The substitution of the ring sulfur atom in de-*O*-sulfonated ponkoranol increases the inhibitory effectiveness of this compound on ntSI by 13-fold (K_i = 0.013 μM ± 0.008). Similarly, substitution of sulfur by selenium leads to an increase in inhibitory activity against ctSI (K_i = 0.018 μM ± 0.004). In the 5' stereoisomer of de-*O*-sulfonated ponkoranol, the substitution of the ring sulfur atom increases the inhibitory effectiveness of this compound on ntSI by 23-fold (K_i = 0.010 μM ± 0.002). Similarly, an increase in inhibitory activity against ctSI (K_i = 0.019 μM ± 0.010) is observed. Little discrimination is observed for inhibition of ntMGAM. Interestingly, de-*O*-sulfonated ponkoranol, its 5' stereoisomer, and the selenium analoges of these two compounds differentiate between the spliceforms of ctMGAM, showing no inhibition of ctMGAM-N2 (Eskandari, Jones, Rose, and Pinto 2011b).

2.4.10 3'-O-methyl ponkoranol

3'-*O*-Methyl ponkoranol inhibits both ctMGAM spliceforms similarly, demonstrating inhibitory activity in the nanomolar range for both ctMGAM-N2 (K_i = 0.060 ± 0.015 μM) and ctMGAM-N20 (K_i = 0.055 ± 0.014 μM). 3'-*O*-Methyl ponkoranol is a weak inhibitor of ntMGAM (K_i = 0.5 ± 0.04

μM)(Eskandari, Jones, et al. 2010b) and is found to be 10 times more potent an inhibitor of ctMGAM-N2 and ctMGAM-N20 than ntMGAM. With respect to ctSI, 3'-*O*-methyl ponkoranol exhibits remarkable inhibitory activity in the nanomolar range (K_i = 0.007 ± 0.002 μM). This compound is the most potent inhibitor of ctSI, approximately 70 times more potent than with ntMGAM. ntSI was inhibited in the nanomolar range as well, with a K_i value of 0.035 ± 0.013 M. 3'-*O*-methyl ponkoranol exhibited a modest differentiation between ctSI and ntSI as well, the compound is 5 times more potent against ctSI than ntSI (Eskandari, Jones, et al. 2010b; Eskandari, Jones, Rose, and Pinto 2011c).

2.4.11 3'-*O*-β-Maltosyl de-*O*-sulfonated ponkoranol

3'-*O*-β-Maltosyl de-*O*-sulfonated ponkoranol is a good inhibitor of ctSI, ntSI, and ntMGAM, with K_i values in the nanomolar range (ctSI (K_i = 0.067 μM ± 0.012), ntSI (K_i = 0.045 μM ± 0.001), ntMGAM (K_i = 0.077 μM ± 0.015). A significant finding is that 3'-*O*-β-maltosyl de-*O*-sulfonated ponkoranol differentiates between the two isoforms of ctMGAM; it is a very poor inhibitor of ctMGAM-N20 and shows no inhibition against ctMGAM-N2 (Eskandari, Jones, Ravinder Reddy, et al. 2011a).

2.4.12 5'-*O*-β-Maltosyl de-*O*-sulfonated ponkoranol

5'-*O*-β-Maltosyl de-*O*-sulfonated ponkoranol demonstrates the ability to inhibit all of the catalytic subunits very well with little distinction between four of the catalytic subunits: ctMGAM-N20 (K_i = 0.067 μM ± 0.012), ctSI (K_i = 0.045 μM ± 0.001), ctMGAM-N2 (K_i = 0.077 μM ± 0.015), and ntSI (K_i = 0.019 μM ± 0.008). Interestingly, 3'-*O*-β-maltosyl de-*O*-sulfonated ponkoranol is the most potent inhibitor of ntMGAM to date with a K_i of 0.008 μM ± 0.002 (Eskandari, Jones, Ravinder Reddy, et al. 2011a).

2.5 Discussion

Mammals have evolved a system of multiple enzymes for the derivation of glucose from various sources of dietary starch. We hypothesize that the four GH31 enzyme units act in a complementary manner depending on the organism's nutritional sources and requirements, as well as their physical and environmental conditions. To test this hypothesis, we developed compounds that can regulate (toggle) the activities of the enzyme units individually. This report

builds the foundation for that approach by identifying differential characteristics of the enzyme units, as reflected in the inhibitory properties of glucosidase inhibitors (Jones et al. 2011).

While the results demonstrate a wide range of inhibitory activities, this represents the first report of the profile of the effectiveness of this family of inhibitors against the individual mammalian intestinal glucosidase activities. These inhibitors, as a group, show the highest inhibitory activity reported against MGAM and SI (Jones et al. 2011). Kinetic analysis confirmed the enzyme activity of the recombinant proteins, and inhibition analysis confirmed classic competitive inhibition by the α-glucosidase inhibitors through analysis of double-reciprocal plots. The results indicate that despite the overall similarity between the subunits in terms of amino acid sequence (Table 1), they exhibit different biochemical and structural properties (Eskandari, Jones, Ravinder Reddy, et al. 2011a).

The results provide support to the notion that the salacinol/kotalanol family will be useful in deriving compounds that have the ability to inhibit selectively each of the glucosidase enzyme units. Relatively small structural changes in a compound can result in significant changes in its ability to selectively inhibit one enzyme unit over the others. For example, salacinol, itself, is a 4 to 5-fold better inhibitor of ctSI as compared to the N-terminal enzymes and ctMGAM-N2. Replacement of the ring S of salacinol by Se, to make blintol, induces a 10-fold improvement in its inhibition of ntMGAM-N2, while having little effect on the other enzyme units. A similar effect is seen upon de-*O*-sulfonation of kotalanol, which results in a 10-fold change in activity against ntMGAM, with little effect on the other enzyme units (Jones et al. 2011).

The inhibitory properties of de-*O*-sulfonated ponkoranol and its 5' stereoisomer were studied against ntMGAM (Eskandari, Kuntz, et al. 2010c). These compounds inhibited ntMGAM with K_i values of 43 ± 3 nM and 15 ± 1 nM, respectively. This is considerably lower than the inhibitory action of the parent compound, ponkoranol (K_i = 170 ± 30 nM) (Rossi et al. 2006). The de-*O*-sulfonation appears beneficial and because the K_i values of de-*O*-sulfonated ponkoranol and its 5' stereoisomer are comparable to that seen with de-*O*-sulfonated kotalanol (K_i = 30 ± 1 nM), it seems the configuration at the C-5' is not critical for inhibition of ntMGAM. The de-*O*-sulfonation is thought to alleviate the steric compression of the sulfate anion in the hydrophobic pocket of the ntMGAM active site (Eskandari, Kuntz, et al. 2010c).

As the 5 -stereoisomer of de-O-sulfonated ponkoranol has proven to be a potent (K_i = 15 nM) inhibitor of ntMGAM (Eskandari, Kuntz, et al. 2010c), to continue evaluating the effect of heteroatom substitution in the sugar ring on glucosidase inhibitory activity, selenium congeners of de-O-sulfonated ponkoranol and its 5' stereoisomer were synthesized and studied as potential glucosidase inhibitors. When the sulfur atom is replaced by the heavier cognate atom selenium, the selenonium ion, inferred from the crystal structure of ntMGAM in complex with kotalanol, in a $^{3}T_{2}$ conformation superimposes well on the $^{4}H_{3}$ conformation of the proposed oxacarbenium transition state, the proposed intermediate in glucosidase-catalyzed reactions (Figure 2-6). Thus, this inhibitor is thought to be an effective mimic of the proposed transition state (Eskandari, Jones, Rose, and Pinto 2011b).

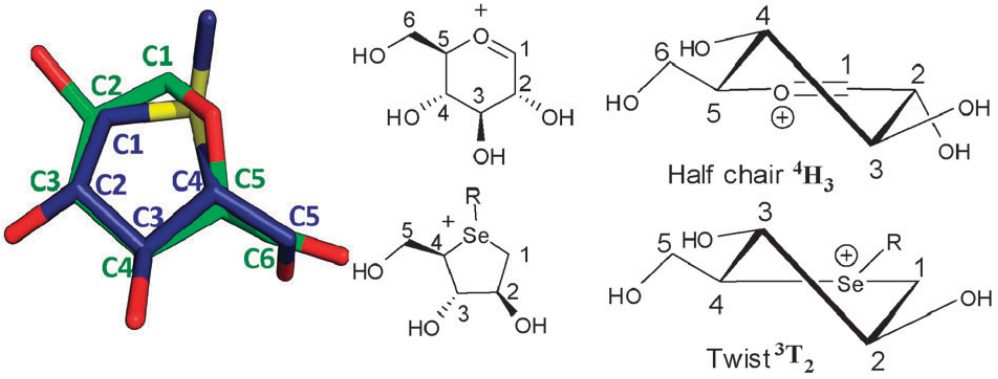

Figure 2-6: Superimposition of the ring carbon atoms of the proposed intermediate in glucosidase-catalyzed reactions (in green) and the selenium ion (in blue) (Eskandari et al. 2011b).

The substitution of the ring sulfur atom in de-O-sulfonated ponkoranol and its 5' epimer by selenium results in a 13-fold (K_i = 0.013 ± 0.008) and 23-fold (K_i = 0.010 ± 0.002) improvement, respectively, in the inhibition of ntSI. ntSI exhibits the broadest substrate specificity of the four catalytic subunits, hydrolyzing both α-1,4 and α-1,6 linkages (Jones et al. 2011). Substitution of sulfur by selenium also results in an increase in inhibitory activity against ctSI (de-O-sulfonated ponkoranol, K_i = 0.018 ± 0.004 μM and its 5' stereoisomer, K_i = 0.019 ± 0.010 μM). Comparatively, there is very little discrimination observed between the congeners in the inhibition of ntMGAM and ctMGAM-N20. All four compounds, de-O-sulfonated ponkoranol, its 5' stereoisomer, and the respective selenium analogues differentiate between the spliceforms of

ctMGAM, with no inhibitive effect on ctMGAM-N2. Currently, the reason for this selectivity is unknown and must await structural analysis of the two isoforms of ctMGAM (Eskandari, Jones, Rose, and Pinto 2011b).

The crystal structure of ntMGAM in complex with acarbose indicates that acarbose is bound to the ntMGAM active site primarily through side chain interactions with its acarvosine unit and there is very little interaction through the other sugar rings. It was proposed that additional subsite interactions with the acarbose sugar rings would significantly increase its inhibitory properties for ctMGAM (Quezada-Calvillo et al. 2008; Sim et al. 2008). We looked to examine whether appending glucose residues to the polyhydroxylated carbon chain of salacinol-based compounds would lead to differential inhibition of the activities of the four catalytic subunits. The extension of the acyclic carbon chain beyond six carbons in salacinol-based compounds is not essential for inhibition of the enzymatic subunits and de-*O*-sulfonated analogues of salacinol-based compounds are more potent inhibitors in comparison to parent compounds (Eskandari, Kuntz, et al. 2010c). Thus de-*O*- sulfonated ponkoranol was selected for modification at C-3'-OH and C-5'-OH of its side chain (Eskandari, Jones, Ravinder Reddy, et al. 2011a).

Interestingly, elongation of the scaffold, as seen in 5'-*O*-β-maltosyl de-*O*-sulfonated ponkoranol and 3'-*O*-β-maltosyl de-*O*-sulfonated ponkoranol, does not result in a significant gain in binding energy, but does result in interesting selectivity in inhibitory activities. These compounds were intended to probe whether one could differentiate between the different enzymes, that is, to toggle between the different enzymes even though they are not strictly analogous to acarbose (Eskandari, Jones, Ravinder Reddy, et al. 2011a). The higher potency of 3'-*O*-β-maltosyl de-*O*-sulfonated ponkoranol against ntMGAM, ctSI, and ntSI compared to both isoforms of ctMGAM, suggest that this inhibitor shows specificity towards these three α-glucosidases. Since ctMGAM is expressed in more than one spliceform (Jones et al. 2011), it provides further complexity to the system. With 3'-*O*-β-maltosyl de-*O*-sulfonated ponkoranol, we are able to maintain the activity of ctMGAM while decreasing the activity of the remaining catalytic subunits (Eskandari, Jones, Ravinder Reddy, et al. 2011a). Additionally, acarbose is an effective inhibitor of ctMGAM in comparison to the ntMGAM and SI subunits (Jones et al. 2011). Using acarbose, we can now turn "off" the ctMGAM unit and study the effect of the remaining catalytic subunits during starch digestion. The ability to selectively inhibit one enzyme unit over

the others can result from relatively small changes in the structure of the compound. This is a significant advancement, as the design of α-glucosidase inhibitors for the treatment of Type II diabetes likely requires specificity for enzymes late in the starch digestion pathway to reduce unwanted side effects (Eskandari, Jones, Ravinder Reddy, et al. 2011a).

Results of the inhibition assays involving the maltose-extended compounds are promising and the relationship between the structure of the catalytic subunits and the inhibitory activity can only be speculative. However, the enzyme selectivity observed in this study will allow us to improve the specificity and affinity of these compounds for potential development as antidiabetic agents, irrespective of whether or not the binding of these compounds occurs through a mechanism analogous to that of acarbose (Eskandari, Jones, Ravinder Reddy, et al. 2011a).

De-*O*-sulfonated ponkoranol (Eskandari, Jones, Rose, and Pinto 2011b) and 3'-O-methyl ponkoranol (Eskandari, Jones, Rose, and Pinto 2011c) also show interesting selectivity profiles for the various catalytic subunits. Comparison of the inhibitory activities of de-*O*-sulfonated kotalanol and some of its stereoisomers against ntMGAM in comparison to kotalanol and the corresponding sulfated stereoisomers indicate that the de-*O*-sulfonated analogs are more potent inhibitors than the parent compounds (Jayakanthan, Mohan, and Pinto 2009; Mohan et al. 2010). In addition, de-*O*-sulfonated ponkoranol (K_i = 0.043 ± 0.01 μM) and its 5' stereoisomer (K_i = 0.015 ± 0.01 μM) are more potent inhibitors of ntMGAM than the parent compound ponkoranol (K_i = 0.17 ± 0.03 μM) (Eskandari, Kuntz, et al. 2010c). Previous structural studies of ntMGAM in complex with kotalanol and de-*O*-sulfonated kotalanol show that removal of the sulfate group affects the conformation of the rest of the polyhydroxylated chain (Sim et al. 2009). In these studies, it was concluded that although the stereoconfiguration at C3' does not affect inhibitory activity, the proximity of the sulfate group to the large hydrophobic groups (Y299, W406, and F575) likely restrict its conformational freedom. To relieve the positional constraint imposed by the bulky hydrophobic residues surrounding the C3' group, the sulfate group was removed, allowing the polyhydroxylated chain to make optimal contacts with the ntMGAM active site (Sim et al. 2009). Thus, we hypothesized that replacing the sulfate group with a hydrophobic methyl ether in ponkoranol may increase its inhibitory properties in comparison to de-*O*-sulfonated ponkoranol by taking advantage of the hydrophobic interactions in the active site. 3'-*O*-Methyl ponkoranol inhibited ntMGAM with a K_i

value of 0.50 ± 0.04 μM whereas de-O-sulfonated ponkoranol and its 5'-stereoisomer inhibited ntMGAM with K_i values of 43 ± 3 and 15 ± 1 nM, respectively. Thus, the hydrophobic interactions between the methyl group and the hydrophobic residues Y299, W406, and F575 in the active site are not as optimal as the interactions of the latter groups with the remainder of polyhydroxylated chain in the absence of the methyl ether. This parallels the binding of the sulfated compound, ponkoranol with a K_i of 0.17 ± 0.03 μM (Eskandari, Kuntz, et al. 2010c; Eskandari, Jones, et al. 2010b).

The inhibitory activity of 3'-*O*-methyl ponkoranol was also examined against ntSI, ctSI, and both isoforms of ctMGAM, N2 and N20 (Eskandari, Jones, Rose, and Pinto 2011b). 3'-*O*-Methyl ponkoranol was a potent inhibitor of all three catalytic subunits. This compound displayed similar inhibitory properties against both ctMGAM isoforms with K_i values in the nanomolar range for both ctMGAM-N2 (K_i = 0.060 ± 0.015 μM) and ctMGAM-N20 (K_i = 0.055 ± 0.014 μM), which is 10 times higher potency compared to the inhibitory action against ntMGAM (K_i = 0.5 ± 0.04 μM) (Eskandari, Jones, et al. 2010b). 3'-*O*-Methyl ponkoranol exhibited remarkable inhibitory activity when tested with ctSI in the nanomolar range (K_i = 0.007 ± 0.002 μM) approximately 70 times more active than with ntMGAM. Similarily, ntSI was inhibited by 3'-*O*-methyl ponkoranol in the nanomolar range, with a Ki value of 0.035 ± 0.013 μM. Finally, this inhibitor demonstrates differentiation between ctSI and ntSI, with 3'-*O*-methyl ponkoranol being 5 times more potent against ctSI than ntSI. These results demonstrate the variation in biochemical and structural properties of the enzymes despite their similarity in amino acid sequence (Eskandari, Jones, Rose, and Pinto 2011c).

When the inhibitory activity of 3'-*O*-methyl ponkoranol is compared with that of de-O-sulfonated pokoranol, the results indicate that addition of the methyl group at the 3' position of de-*O*-sulfonated ponkoranol contributes to a significant inhibition of ctMGAM-N2 (K_i = 0.060 ± 0.015 μM), whereas de-*O*-sulfonated ponkoranol without this addition, does not inhibit ctMGAM-N2 (Eskandari, Kuntz, et al. 2010c; Eskandari, Jones, Rose, and Pinto 2011b). Both 3'-*O*-methyl ponkoranol and de-*O*-sulfonated ponkoranol inhibit ctMGAM-N20 with similar K_i values. Inhibition of ntSI was an order of magnitude greater by 3'-*O*-methyl ponkoranol compared to de-*O*-sulfonated ponkoranol (Eskandari, Jones, Rose, and Pinto 2011c).

Remarkably, 3'-*O*-methyl ponkoranol inhibited ctSI with K_i values of 7 ± 2 nM, which is significantly lower than that for de-O-sulfonated ponkoranol itself (K_i = 103 ± 30 nM) (Eskandari, Kuntz, et al. 2010c; Eskandari, Jones, Rose, and Pinto 2011b). Thus, 3'-*O*-methyl ponkoranol is the most potent compound against ctSI to date in this class of molecules. We speculate ctSI will have a hydrophobic pocket in the catalytic site that better accommodates the methyl group and provides favorable hydrophobic interactions compared to the other enzymatic subunits in which steric interactions between the methyl group and active site residues likely lead to unfavorable contacts, similar to the unfavourable contacts seen in the binding of the sulfated derivative, kotalanol compared to its de-O-sulfonated analogue (Sim et al. 2009; Eskandari, Jones, Rose, and Pinto 2011c).

Improved understanding of the interactions between the inhibitory compounds developed and the active site of the enzymes await further structural studies, in particular the interaction with the C-terminal active sites. Gaining insight into the structure of the active sites of the C-terminal catalytic subunits as well as the N-terminal catalytic subunits in complex with potent inhibitors will provide valuable tools to design and synthesize α-glucosidase inhibitors effective against and specific to each catalytic subunit. These inhibitors are promising potential candidates to be used as oral agents in the treatment and prevention of non-insulin dependent Type II diabetes.

To investigate the ability to toggle the enzyme activity and the potential of sulfonium ion-based inhibitors to modulate starch digestion, the four enzymatic subunits were assayed individually providing α-amylolyzed waxy corn starch as the substrate in the presence of four different inhibitors. Four representative inhibitors were chosen due to the selectivity that has been seen in previous studies (Jones et al. 2011; Eskandari, Jones, Ravinder Reddy, et al. 2011a) acarbose, de-*O*-sulfonated kotalanol, 3'-β-maltosyl de-*O*-sulfonated ponkoranol, and 5'-β-maltosyl de-*O*-sulfonated ponkoranol. α-amylolyzed waxy cornstarch consists mainly of maltose, maltotriose, and branched α-limit dextrins and is representative of the products of α-amylase hydrolysis present in the small intestinal lumen. Because all four enzymatic subunits are required for effective digestion of the linear malto-oligosaccharides and branched α-limit dextrins, toggling the activity of the catalytic subunits may be a good approach to moderating blood glucose excursions (Lee et al. 2012). To avoid gastrointestinal upset, a partial, controlled

inhibition is preferred because at high levels of inhibition, the bulk of the starch will enter the large intestine undigested (Lee et al. 2012).

This study illustrated that controlling the rate of starch digestion using differential inhibition is indeed possible. The four catalytic subunits required different concentrations of acarbose to inhibit the enzymatic activity. For example, 5 nM of acarbose effectively inhibits ctMGAM and ctSI, whereas ntMGAM and ntSI were virtually uninhibited. At concentrations of only 50 pmol, de-O-sulfonated kotalanol exhibited 20% inhibition of ctMGAM activity and 40% inhibition of ctSI activity, while N-terminal subunits remained uninhibited. This demonstrates that de-O-sulfonated kotalanol can turn the C-terminal catalytic subunits off, especially ctSI, which has sucrase activity, with little effect on the activity of the N-terminal subunits. 3'-β-Maltosyl de-*O*-sulfonated ponkoranol demonstrated almost no inhibitory effect below 50 pmol. Each α-glucosidase showed a different susceptibility to 3'-β-maltosyl de-*O*-sulfonated ponkoranol at 500 pmol. ctMGAM and ntMGAM were inhibited by 50% at 500 pmol, however the inhibitory pattern of ntSI and ctSI subunits varied with 70% inhibition of ctSI and 20% inhibition of ntSI. 5'-β-Maltosyl de-*O*-sulfonated ponkoranol did not show any inhibitory effect below 50 pmol and all catalytic activity was abruptly decreased at concentrations of 500 pmol.

ctMGAM is the catalytic subunit with the highest activity. This catalytic subunit also demonstrated high-binding and hydrolytic properties for larger malto-oligosaccharides (Quezada-Calvillo and Robayo-Torres 2007; Quezada-Calvillo et al. 2008; Sim et al. 2008). Due to these properties, selectively inhibiting the C-terminal domains is a possible approach to slowing glucose release, leaving the task of glucogenesis from starch to the slower and more specific N-terminal catalytic subunits (Lee et al. 2012). In this case, de-*O*-sulfonated kotalanol would be a good candidate, as it can inhibit the C-terminal subunits but leaves the N-terminal subunits active. This compound also inhibits ctSI especially well, and due to the activity of ctSI on sucrose, has potential for use in modulation of sucrose digestion as well. 5'-β-Maltosyl de-*O*-sulfonated ponkoranol displayed similar inhibitory effects on all four catalytic subunits, and thus would not be a good candidate for differential inhibition of α-glucosidases. 3'-β-Maltosyl de-*O*-sulfonated ponkoranol showed selective inhibition of ctSI or can be used at lower concentrations to inhibit ctSI, ctMGAM, and ntMGAM, and leave only ntSI catalytically active. Compounds that have an inhibitory effect on one or two subunits have potential to be used to modulate starch digestion and glucose delivery *in vivo* (Lee et al. 2013).

2.6 Conclusion

The results of our inhibitor studies illustrate the differences both structurally and biochemically between the enzyme subunits, despite overall similarity in amino acid sequence and structural architecture. Compounds capable of selectively inhibiting catalytic subunits independently have been developed and selectivity has been demonstrated with linear malto-oligosacharides and α-limit dextrins. Both C-terminal subunits have stronger affinity for longer oligomers, and this is reflected in the results of the inhibitor studies, as ctMGAM-N2, ctMGAM-N20, and ctSI are more sensitive to the longer inhibitor acarbose than ntSI and ntMGAM. According to previous studies by Sim et al. 2008 involving structural analysis of ntMGAM, subsite mapping with acarbose, and sequence alignments, it is likely that the higher affinity seen in the C-terminal subunit is due to an extended substrate-binding site. The structure of ntMGAM indicates that this subunit lacks +2 and +3 subsites. Further, in ctMGAM there is a substitution of an alanine to a phenylalanine as well as a 21-residue extension near the catalytic site in ctMGAM. We speculate that the different hydrolytic properties of the MGAM and SI catalytic subunits and the observed overlapping substrate specificity aid in accommodating the diverse composition of starches in the human diet (Jones et al. 2011).

Results of these studies also further support the concept that sulfonium ion-based inhibitors will be valuable tools in developing compounds that can independently toggle the activities of each enzyme in turn. Ultimately, we look to apply the supported notion that the rate of starch digestion can be controlled through selective inhibition of ntMGAM, ntSI, ctMGAM, and ctSI, mimicking the effect of slowly digestible starch that is digested throughout the small intestine (Zhang and Hamaker 2009). Slowing the digestion of starch will help to lessen the initial spike in glucose levels and prolong postprandial blood glucose delivery, which may be beneficial in the treatment of non-insulin dependent Type II diabetes. Ultimately, compounds that can independently toggle the activities of each enzyme in turn will open the door to understanding their roles in starch digestion individually and in all possible combinations (Jones et al. 2011).

Chapter 3
C-terminal Homology Models

3.1 Overview

To help understand the substrate specificity as well as the results of the inhibition studies discussed in Chapter 2, homology models of the C-terminal subunits ctSI and one murine isoform of ctMGAM (N20) were produced. Previous studies (Sim et al. 2010; Jones et al. 2011; Lee et al. 2013) indicate that the C-terminal subunits are capable of binding longer substrates in comparison to the N-terminal subunits. This ability was hypothesized to be due to of the 21-amino acid insertion present in the C-terminal subunits not found in the N-terminal subunits. This hypothesis was supported by the structure of ctMGAM recently determined by Ren et al. 2011. Based on sequence analysis, this structure is the human equivalent of the murine isoform ctMGAM-N2.

Most recently, selectivity between two isoforms of ctMGAM (N2 and N20) has been seen with shorter inhibitors such as the de-*O*-sulfonated ponkoranol and its derivatives. The selectivity seen between ctMGAM-N2 and ctMGAM-N20 is hypothesized to be due to the 89-amino acid insertion found in ctMGAM-N20, but not present in ctMGAM-N2. Although crystal structures are a long-term goal, we hope to increase our understanding of the seeming duplicity and active site requirements of the four catalytic subunits through examining the homology models of murine ctSI and ctMGAM-N20.

In this chapter, homology models of the C-terminal subunits ctSI and ctMGAM-N20 will be presented and structural differences between the catalytic subunits will be explored in the context of catalytic function and substrate specificity.

3.2 Contributions

All work presented in this chapter was completed by the candidate. Sections of this chapter are modified from or taken verbatim from a report prepared by the candidate for BIOL 614: Bioinformatics – Tools and Techniques. Materials and Methods

3.3 Materials and Methods

3.3.1 Template Selection and Sequence Alignment

3.3.1.1 ctSI

The ctSI sequence was used in a BLAST (Altschul et al. 1997) query of the Protein Data Bank (PDB) to find homologous sequences with known structures to use as templates. As expected, the query results were the other enzymatic subunits, ctMGAM, ntSI, and ntMGAM, as native structures as well as with bound ligands. A multiple sequence alignment of the target sequence (ctSI) and template sequences (ctMGAM: PDB 3TON (Ren et al. 2011) and ntSI: PDB 3LPO (Sim et al. 2010)) was calculated using MUSCLE (Edgar 2004). The ctMGAM (PDB: 3TON) structure (Ren et al. 2011) contains the 21 amino acid insertion that is found in ctSI but absent from ntSI. The first 57 amino acids of the ctSI construct sequence were removed as well as the expression tag region to optimize the alignment. According to analysis by Signal P 4.1 (Petersen et al. 2011), there is a cleavage site between residues 38 and 39. In addition, domain prediction by DomPred Protein Domain Prediction Server (Marsden, McGuffin, and Jones 2002) indicated that the first approximately 40 amino acids are not part of the domains of interest in the ctSI enzyme. In summary, the first 57 amino acids were removed because they did not align with the template, are likely part of a signal peptide, and are not a component of the domains of interest.

3.3.1.2 ctMGAM-N20

The structure of the N20 isoform of ctMGAM has not yet been determined experimentally. This is a murine isoform with an 89-amino acid insertion that is absent from the sequence and structure of ctMGAM (PDB: 3TON) (Ren et al. 2011) as well as the other murine isoforms. The methodology for template selection and alignment is the same as that for ctSI (3.2.1.1). The first 136 amino acids of the ctMGAM-N20 sequence were removed as well as the expression tag region to optimize the alignment. Similar to ctSI, these amino acids were removed because they did not align with the template and are not a component of the domains of interest.

3.3.2 Modeller

3.3.2.1 ctSI

Using the multiple sequence alignment, homology modeling was completed with Modeller (Sali and Blundell 1999) using the command line version as well as the Modeller Plug-in through the UCSF Chimera (Pettersen et al. 2004) interface. Five models were produced for each target and the model that corresponded to the overall topology and fold pattern of the templates in addition to having the lowest discrete optimized protein energy (DOPE) score was selected. An additional criterion was to have a modeled loop corresponding to that seen in the structure of ctMGAM (PDB: 3TON) (Ren et al. 2011) of the 21-amino acid insertion. Although the program itself will perform energy minimization, the selected model was submitted to ModRefiner (Xu and Zhang 2011) for consistency.

3.3.2.2 ctMGAM-N20

The protocol used to model ctMGAM-N20 was identical to that discussed in Section 3.2.2.1 except for the following modifications. After homology modeling was complete using Modeller, the 89-amino acid insertion remained unstructured. Thus, this region was selected for *ab initio* structure prediction (Xu and Zhang 2012) using QUARK. QUARK produced 10 different models, and each model was compared to the structure of the 89-amino acids preceding this unstructured loop, as there was high sequence identity between these two regions. The structure of the unordered loop was predicted to be similar to the preceding structural element of high sequence similarity. Each possible fold of the 89-amino acid insertion produced by QUARK was highly similar to the preceding structural element, as predicted. All fold possibilities produced by QUARK were closely compared to the structural element of high sequence similarity, and the model that had the most similar structure to the structural element with high sequence similarity and acceptable statistical likelihood was chosen for use in further modeling work.

The unstructured insertion (residues 606-694) was removed from the initial model of ctMGAM-N20 and both the structures were visually inspected to determine if the two models could possibly be oriented appropriately to create one model. Visually, this seemed possible, and thus the Chimera (Pettersen et al. 2004) function Join Models was used to join residue 606 in the model of the 89-amino acid insertion to residue 605 in the ctMGAM-N20 model. The

function Join Loops in Chimera (Pettersen et al. 2004) creates a bond between two atoms in different models, moving one of the models (in this case, the 89-amino acid insertion), to form the appropriate bond. This created a new model with the structured 89-amino acid insertion and the N-terminal "half" of the N20 homology model, residues 1-605. This newly created model included the structured 89-amino acid insertion with residues numbered 1-694 (Model A). This left the C-terminal "half" of the ctMGAM-N20 homology model (residues 695-986, Model B) unconnected from the newly created model (Model A). This new model (Model A), was refined using the ModRefiner (Xu and Zhang 2011) server. Following this, residue 694 in the newly refined Model A and residue 695 in the C-terminal half of the ctMGAM-N20 homology model (Model B) were joined using the Join Models function in Chimera (Pettersen et al. 2004). This created a final model (Model C) that consisted of the ctMGAM-N20 homology model with the structured 89-amino acid insertion. This final model (Model C) was then minimized using ModRefiner (Xu and Zhang 2011) and 3DRefine (Bhattacharya and Cheng 2012). To help identify each domain, comparisons were made to known structures and the tool InterPro Scan (Quevillon et al. 2005) was used.

3.3.3 Active Site Inspection

The homology models were aligned to known structural models (ntSI, PDB: 3LPO (Sim et al. 2010); ntMGAM, PDB: 2QLY (Sim et al. 2008); and ctMGAM, PDB: 3TON (Ren et al. 2011)) using structural alignment protocols in Chimera (Pettersen et al. 2004) and sequence-based alignment using MUSCLE (Edgar 2004) to determine the conserved residues that most likely act as the nucleophile, acid/base catalyst, and substrate-binding residues. All figures were created using PyMol (DeLano Scientific) and Chimera (Pettersen et al. 2004). Due to the significant structural differences in the ctMGAM-N20 homology model, the location of the active site was predicted by 3DLigandSite (Wass, Kelley, and Sternberg 2010) in addition to sequence and structural alignment.

3.3.4 ctMGAM Activity Assays

Activity assays of ctMGAM-N20 were done using the glucose oxidase activity assay as described in Chapter 2, Section 2.3.3, using isomaltose and sucrose as the substrate in place of maltose.

3.4 Results

3.4.1 ctSI

3.4.1.1 Alignment

The multiple sequence alignment calculated using ctSI, ntSI (PDB: 3LPO) (Sim et al. 2010), and ctMGAM (PDB: 3TON) (Ren et al. 2011) is illustrated in Figure 3-1. This alignment illustrates the high degree of similarity between the three enzymes, and also illustrates the 21-amino acid insertion found only in the C-terminal subunits.

Figure 3-1: Sequence alignment of ntSI (PDB: 3LPO) (Sim et al. 2010), ctMGAM (PDB: 3TON) (Ren et al. 2011), and ctSI used in homology modeling. 21-amino acid extension is indicated by the boxed region.

3.4.1.2 Homology Model

The homology model of ctSI (Figures 3-2, 3-3, and 3-4) is very similar to the crystallographic structure of ctMGAM (PDB: 3TON, 3TOP) (Ren et al. 2011) as well as the crystal structures of the other catalytic subunits (Figure 3-2). There is a trefoil type P domain (residues 1-52), an N-terminal β-sandwich domain (residues 53-272), a catalytic $(\beta/\alpha)_8$ barrel domain with two loop inserts (residues 273-678), a proximal C-terminal domain (residues 679-757) and distal C-

terminal domain (residues 758-897) in β-sandwich topologies. There is some disorder in the trefoil domain, which is likely an artifact of the modeling process.

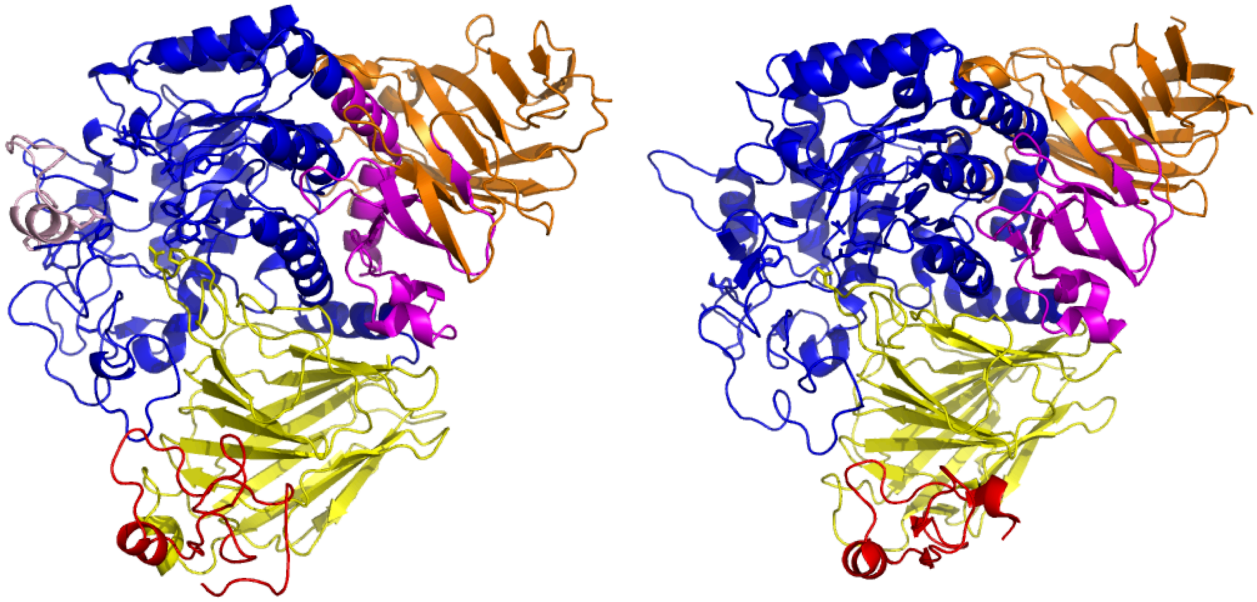

Figure 3-2: ctSI Homology Model (left) and ntMGAM Crystal Structure (PDB: 2QLY, right) (Sim et al. 2008). Red: trefoil type P domain; Yellow: N-terminal β-sandwich domain; Blue: catalytic $(\beta/\alpha)_8$ barrel domain; Magenta: proximal C-terminal domain; Orange: distal C-terminal domain.

The architecture of the active site is very similar to the known structures of similar function and there are few structural differences. The significant structural differences are summarized in Table 3-1, below. The 21-amino acid extension found in ctSI is located as a helical structure near the active site, as seen in the crystal structure of ctMGAM (PDB: 3TON) (Ren et al. 2011).

Table 3-1: The significant differences between the ctSI homology model and the other enzymatic subunits with known crystal structure.

ctSI Homology Model	ntSI (PDB: 3LPO)	ctMGAM (PDB: 3TON)	ntMGAM (PDB: 2QLY)
Glutamine Residue 420	Not Present	Tryptophan	Not Present
Tyrosine Residue 573	Threonine	Threonine	Threonine
Cysteine Residue 401	Isoleucine	Valine	Leucine
Tryptophan Residue 402	Isoleucine	Tryptophan	Isoleucine

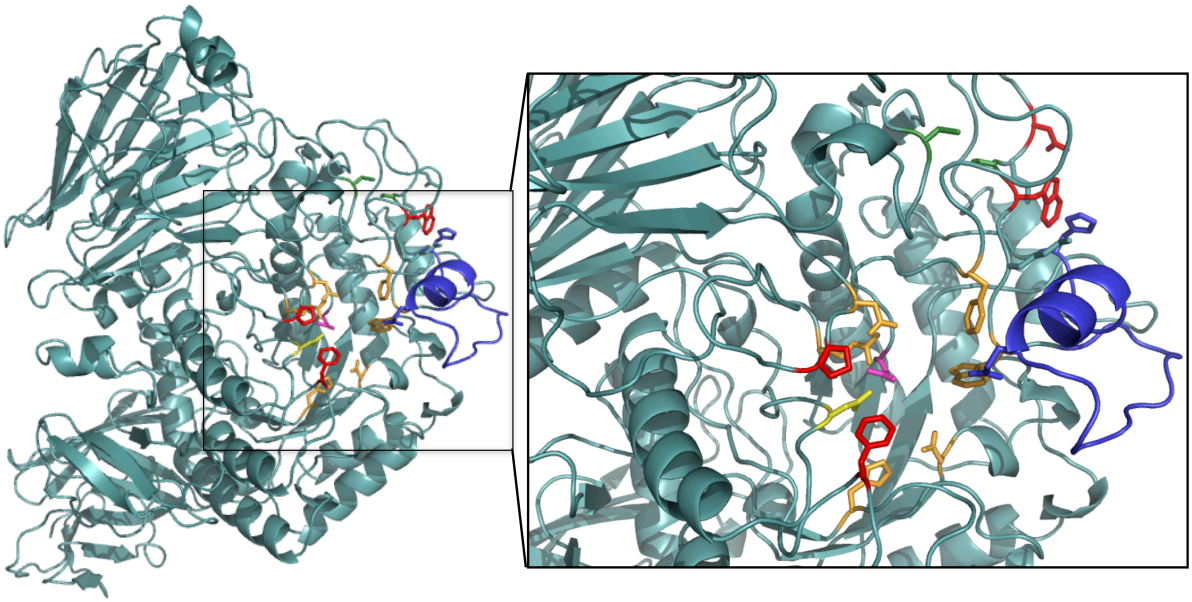

Figure 3-3: Full view of the ctSI homology model as well as a view of the active site. Blue: 21-amino acid extension; Orange: binding residues; Yellow: acid/base catalyst; Magenta: nucleophile; Red: residues that differ between ctSI and ntSI.

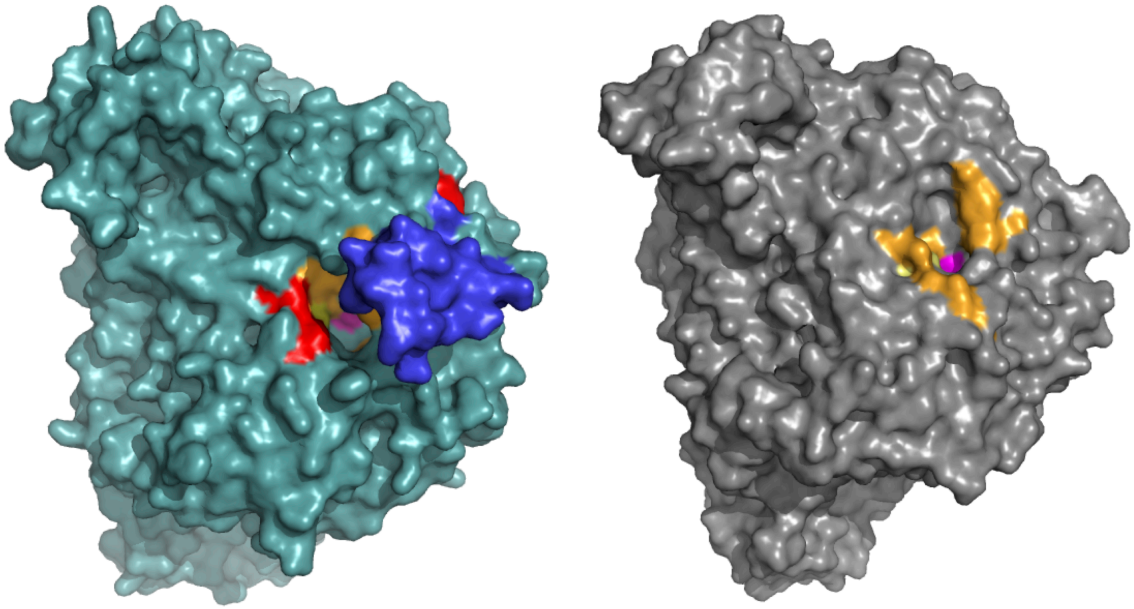

Figure 3-4: Surface representation of the ctSI homology model (left) and ntSI (Sim et al. 2010) (PDB: 3LPO, right). Blue: 21-amino acid extension; Orange: binding residues; Yellow: acid/base catalyst; Magenta: nucleophile; Red: residues that differ between ctSI and ntSI.

3.4.2 ctMGAM-N20

3.4.2.1 Alignment

The multiple sequence alignment calculated using ctMGAM-N20, ntMGAM (PDB: 2QLY) (Sim et al. 2008), and ctMGAM (PDB: 3TON) (Ren et al. 2011) is illustrated in Figure 3-5. As illustrated in the alignment, there is a high degree of similarity between the three catalytic subunits, but in addition to the 21-amino acid insertion found in the C-terminal subunits, there is an additional 89-amino acid extension found only in ctMGAM-N20.

Figure 3-5: Sequence alignment of ntMGAM (PDB: 2QLY) (Sim et al. 2008), ctMGAM (PDB: 3TON) (Ren et al. 2011), and ctMGAM-N20 for use in homology modeling. 21-amino acid extension and 89-amino acid extension are indicated by the boxed regions.

3.4.2.2 Homology Model

The homology model of ctMGAM-N20 (Figures 3-6, 3-7, and 3-8) is similar to the crystallographic structure of ctMGAM (PDB: 3TON, 3TOP) (Ren et al. 2011), which is the human equivalent of ctMGAM-N2. While the overall topology of the protein structure is similar, there are changes in the position of the domains as well as discrepancy in the architecture of the active site, in comparison to the known structures of the N- and C-terminal domains. Some of the domains, particularly the distal C-terminal domain, are slightly dissociated from the protein core, but this is likely an artifact of the modeling process. The protein is expected to have a more compact structure in its native form. All domains present in the known structures of catalytic subunits are also found in the homology model of ctMGAM-N20 (Figure 3-6). There is a trefoil domain that is disordered in the model (residue 1-49). Although there is disorder in the homology model, protein domain classification servers such as InterPro Scan (Quevillon et al. 2005) predict that this structural element is present. The putative trefoil domain is shifted and

is in close proximity to the 89-amino acid extension in the $(\beta/\alpha)_8$ catalytic domain. The N-terminal β-sandwich domain (residues 50-268) is also present. The $(\beta/\alpha)_8$ catalytic domain (residues 269-787) is also shifted so the catalytic centre is no longer in the central region of the protein. Further a few of the β-strands in this domain are unstructured. The proximal C-terminal domain (residues 788-849) and distal C-terminal domain (residues 850-970) are also shifted slightly and the β-sandwich topology of the distal C-terminal domain is slightly distorted. Despite these small structural changes, all domains are present, intact, and in the expected locations based on the known structures of the other catalytic subunits.

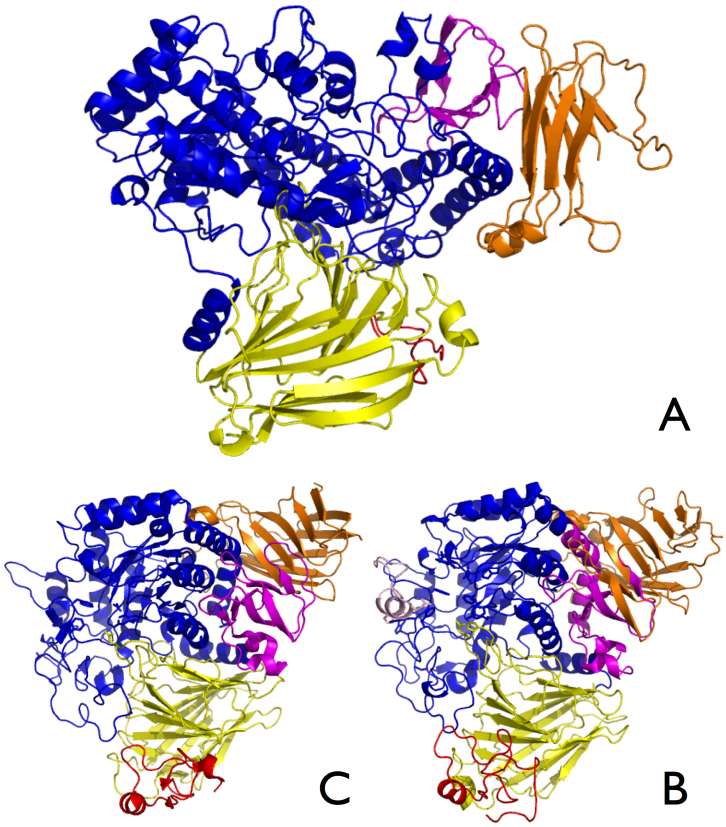

Figure 3-6: ctMGAM-N20 homology model (A) and homology model of ctSI (B) and ntMGAM crystal structure (PDB: 2QLY) (Sim et al. 2008) (C) for comparison. Red: trefoil type P domain; Yellow: N-terminal β-sandwich domain; Blue: catalytic $(\beta/\alpha)_8$ barrel domain; Magenta: proximal C-terminal domain; Orange: distal C-terminal domain.

The 21-amino acid extension in the C-terminal subunits is located as a helical structure near the active site, consistent with that seen in the crystal structure of ctMGAM (PDB: 3TON) (Ren et al. 2011), and a significant difference is seen in the presence and position of the 89-amino acid extension in close proximity to the opening of the active site (Figures 3-7 and 3-8).

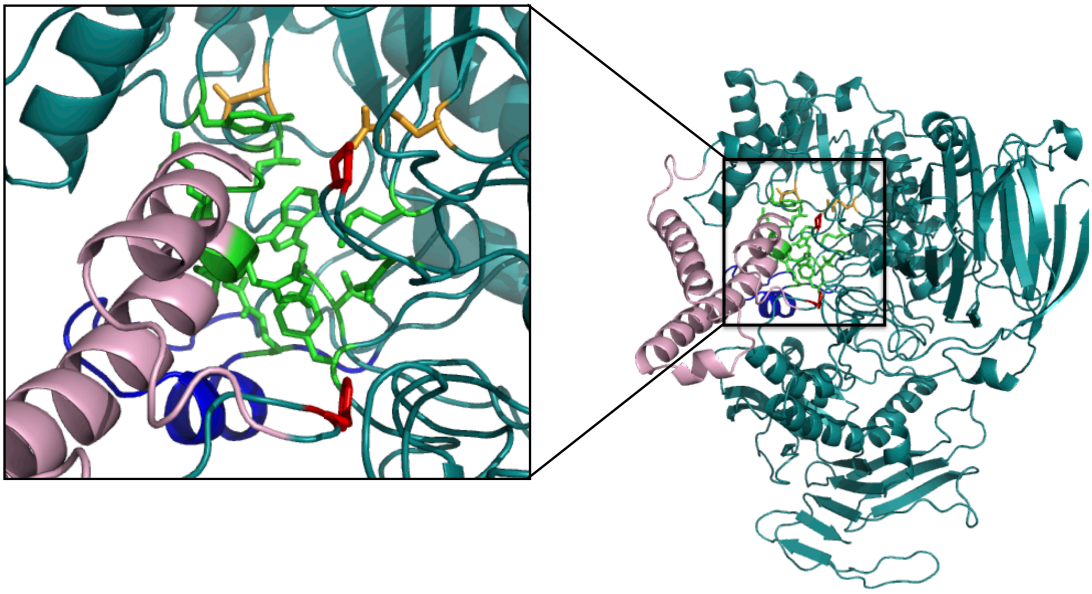

Figure 3-7: Full view of the ctMGAM-N20 homology model as well as a view of the active site. Blue: 21-amino acid extension; Green: predicted active site; Orange: binding residues consistent with other catalytic subunits; Red: residues found in the C-terminal subunits only; Pink: 89-amino acid extension.

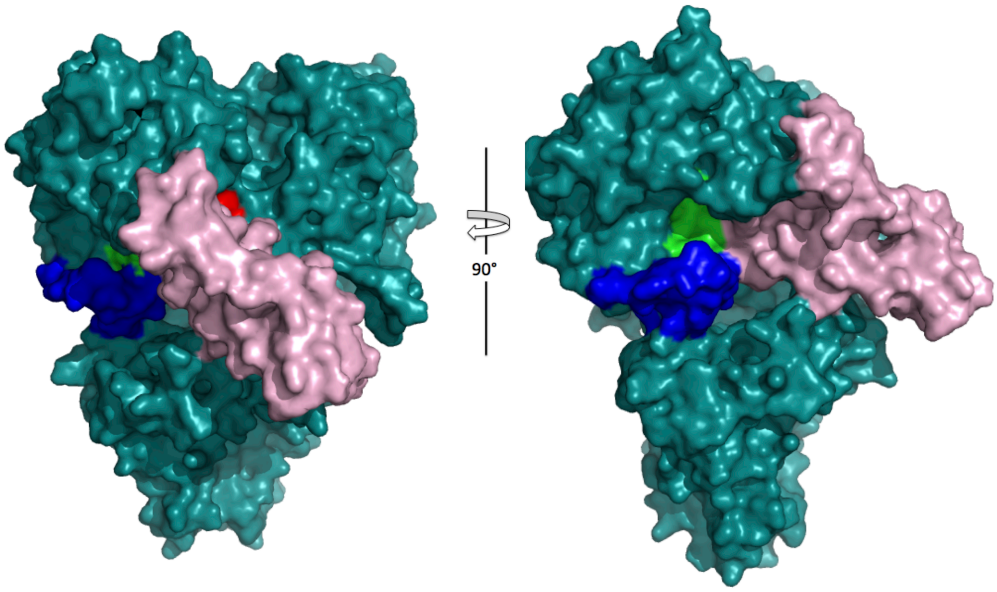

Figure 3-8: Surface representation of the ctMGAM-N20 homology model. Blue: 21-amino acid extension; Red: residues found in C-terminal subunits only; Pink: 89-amino acid extension; Green: predicted binding site.

Prediction of the active site in the homology model was done using 3DLigandSite (Wass, Kelley, and Sternberg 2010), which searches a structural library of structures with ligands in the active site. The highest scoring cluster (average MAMMOTH score of 11.2) had residues very similar to the active site of the other catalytic subunits of known structure and also included the acid/base catalyst that is conserved in this glycoside hydrolase family as part of the catalytic signature sequence (residue 467, aspartic acid). There were glucose ligands found in similar active site pockets in seven structures, supporting the notion that this active site likely binds glucose as well. Predicted binding site residues: Aspartic acid 204, Tyrosine 298, Isoleucine 327, Aspartic acid 364, Valine 401, Tryptophan 402, Arginine 424, Aspartic acid, 467, Metionine 468, Phenylalanine 747, Histidine 674, Methionine 675, Glycine 676, and Glycine 677. One of the residues predicted to be in the active site, Asp 364, is located approximately 5 Å from the acid/base catalyst and is the most likely candidate to act as the nucleophile during hydrolysis. There is also a possibility that aspartic acid 335 is the nucleophile, however in the homology model these side chains are approximately 8 Å from one another, too far for proper hydrolysis. The similarities to the active sites of the other catalytic subunits of known structure, the conservation of the acid/base catalyst, as well as and the location of the predicted active site in

the $(\beta/\alpha)_8$ catalytic domain provided the appropriate evidence that this is the likely binding site in ctMGAM-N20.

3.4.2.3 Hydrolytic Activity

After initially inspecting the model, a glucose oxidase assay was completed using ctMGAM-N20, providing isomaltase as the substrate, to identify whether or not this catalytic subunit is capable of hydrolyzing α-1,6 linkages. Although there was not enough substrate for a full kinetic analysis, it was confirmed that glucose is liberated from isomaltose in the presence of ctMGAM-N20. This demonstrates the capability of ctMGAM-N20 to hydrolyze α-1,6 linkages in addition to α-1,4 linkages. In a separate assay, sucrose (α-1,2) was also provided as a substrate, but no hydrolysis was detected.

3.5 Discussion

3.5.1 ctSI

Both ctSI and ntSI are able to hydrolyze α-1,4-glycosidic linkages. The main difference in function between the two enzymatic subunits is the ability to hydrolyze α-1,2 and α-1,6-glycosidic bonds by ctSI and ntSI, respectively. In addition, all C-terminal catalytic subunits are able to accommodate longer substrates, such as acarbose, in comparision to the N-terminal subunits. Without a crystal structure of ctSI, the biochemical rationale for this difference in substrate specificity is unknown, but based on the homology model of ctSI, significant differences are seen in the active site in comparison to ntSI that can be postulated to be related to substrate specificity. The C-terminal subunits have been hypothesized to have an extended binding site in comparison to the N-terminal subunits based on inhibitor and substrate studies (Sim et al. 2008; Sim et al. 2010; Eskandari, Jones, et al. 2010b; Jones et al. 2011; Lee et al. 2012; Lee et al. 2013). The crystal structure of human ctMGAM (PDB: 3TON, 3TOP) (Ren et al. 2011) in which acarbose was bound in the active site, confirmed the extension of the binding site due to the 21-amino acid extension found only in the C-terminal subunits.

In the structure of human ctMGAM, there is a tryptophan in this 21-amino acid extension that interacts with the acarbose substrate, extending the binding cleft (Ren et al. 2011). The same 21-amino acid extension is found in ctSI, and in the homology model the active site architecture is very similar to that of ctMGAM (PDB: 3TON) (Ren et al. 2011). In the case of ctSI, the binding

cleft is also extended due to the presence of the 21-amino acid extension, however the tryptophan that interacts with acarbose in the ctMGAM crystal structure is replaced with a glutamic acid (residue 420, Figure 3-9, A and B). All C-terminal catalytic subunits have an increased affinity for longer substrates, but there is a small degree of selectivity for acarbose seen between ctMGAM-N2 and ctSI. ctMGAM-N2 (the murine equivalent of the human structure of ctMGAM (Ren et al. 2011)) has approximately 25 times higher affinity for acarbose than ctSI, with K_i values of 9 nM and 246 nM, respectively. The lack of this aromatic residue on the 21-amino acid extension in ctSI may explain this selectivity seen between ctSI and ctMGAM with acarbose as well as the high affinity of ctSI for small inhibitors, as the tryptophan residue was found to be important in stabilizing acarbose in the active site of ctMGAM (Ren et al. 2011). Thus, this small modification in which a tryptophan is replaced with a glutamic acid, losing the aromatic ring at this position to help coordinate longer substrates, may be the biochemical explanation for the slightly decreased affinity for acarbose seen with ctSI in comparison to ctMGAM-N2.

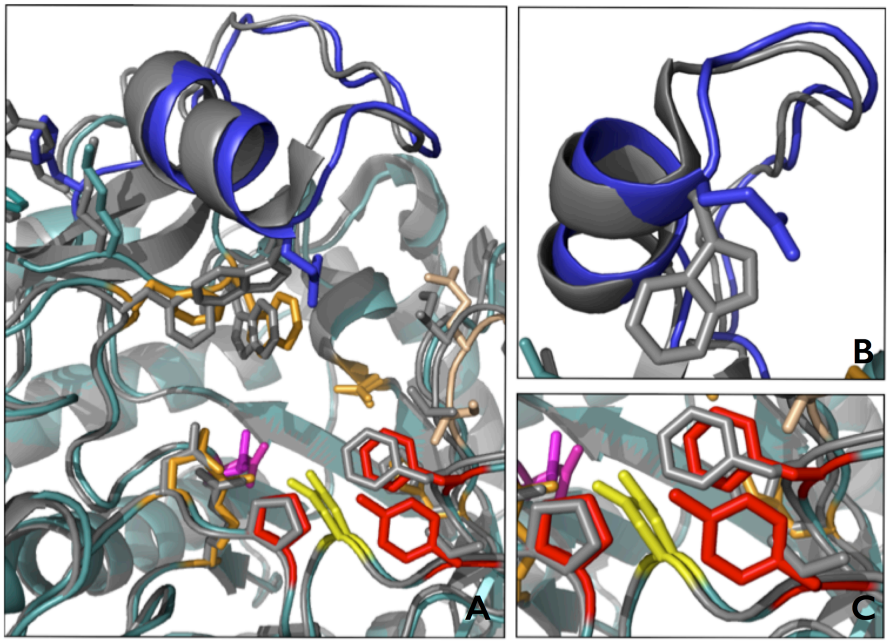

Figure 3-9: Comparison of ctSI and ctMGAM (PDB: 3TON) (Ren et al. 2011) active sites. A: ctSI homology model superposed with ctMGAM (PDB: 3TON) **(Ren et al. 2011)** Deep Teal: ctSI homology model; Gray: ctMGAM (PDB: 3TON) **(Ren et al. 2011)**; Blue: 21-amino acid extension; Yellow: acid/base catalyst; Magenta: nucleophile; Red: residues that differ between N- and C-terminal catalytic subunits. **B:** Close-up of the 21-amino acid extension illustrating tryptophan found in ctMGAM is a glutamic acid at the equivalent position in ctSI homology model. **C:** Close up of tyrosine (red, residue 573) in ctSI homology model illustrating that in ctMGAM (PDB: 3TON), the equivalent residue is threonine (gray). Also highlighted, proline (residue 210), and phenylalanine (residue 605) in ctSI homology model which may help coordinate the sucrose molecule in the active site during α-1,2 hydrolysis.

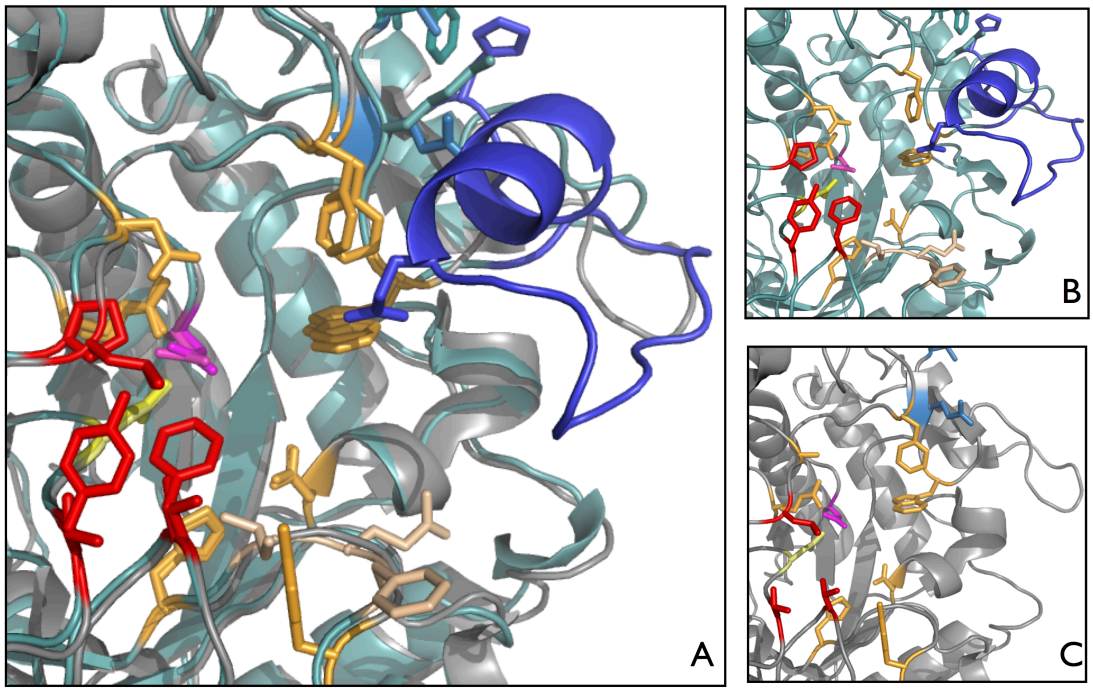

Figure 3-10: Comparison of ctSI and ntSI (PDB: 3LPO) (Sim et al. 2010) active sites. A: Superposition of ctSI (deep teal) and ntSI (PDB: 3LPO, gray) **(Sim et al. 2010)** active sites. **B:** ctSI homology model active site. **C:** ntSI (PDB: 3LPO) **(Sim et al. 2010)** active site. Dark Blue: 21-amino acid extension; Yellow: acid/base catalyst; Magenta: nucleophile; Red: residues that differ between N- and C-terminal catalytic subunits; Light Blue: residues that differ between ntSI and ctSI homology model that may be involved in the extended binding site.

Further, without the 21-amino acid extension, the binding site of ntSI is much more open and accessible, likely allowing for the hydrolysis of branch points (α-1,6). For debranching activity to be possible with ntSI, a relatively large, branched polymer of glucose must get close to the active site. If the 21-amino acid extension were present, this would likely prevent the active site from accommodating a branch point. This difference in structure between ctSI and ntSI is a possible explanation for the ability of ntSI to act as a debranching enzyme, whereas ctSI does not possess this ability.

One of the main functional differences between ctSI and the other catalytic subunits is the ability to hydrolyse α-1,2 linkages found in sucrose. With respect to the active site, there are some variances that may explain this difference in function (Figure 3-10). The most significant

difference is a threonine in the -1 subsite that is conserved in ntSI, ntMGAM, ctMGAM-N2, and ctMGAM-N20. ctSI is the only enzymatic subunit in which this threonine is not conserved, and is replaced with a tyrosine (residue 573) (Figure 3-9, A and C, Figure 3-10). This tyrosine is in close proximity to the acid/base catalyst (aspartic acid, residue 571). In the structures of other glycoside hydrolases (family 32 and 68) with sucrose in the active site, a tyrosine is found in the +1 subsite, specific for the glucosyl moiety, and is shown to be involved in hydrogen bonding with the acid/base catalyst. The pentose sits in the -1 subsite, which is specific to the binding of the fructose unit. Additionally, there is a hydrogen bond observed between the acid/base catalyst and this tyrosine residue (Lammens et al. 2009). In the case of ctSI, the tyrosine residue is found near the -1 subsite, only a few residues from the acid/base catalyst.

The presence of this tyrosine residue, which is only present in the ctSI active site, leads to the hypothesis that during the hydrolysis of sucrose, the glucosyl moiety sits in the -1 subsite while the fructose unit interacts with the +1 subsite. In addition, there is an aromatic phenylalanine residue in the -1 subsite (residue 605) and a proline in the +1 subsite (residue 210) (Figure 3-9, C, Figure 3-10) that are found only in the C-terminal catalytic subunits and likely form interaction with the sugar substrate. These residues are thought to assist in coordinating the sucrose molecule in the active site to allow for effective hydrolysis. The hypothesis that the glucosyl moiety sits in the -1 subsite is further supported by the overlapping substrate specificity seen in ctSI. This enzymatic subunit is capable of hydrolyzing α-1,4 linkages in addition to α-1,2 linkages (Nichols et al. 2003). Thus, a change in specificity in the -1 subsite for the fructosyl moiety of sucrose instead of the usual glucose molecule would be unlikely while still maintaining both α-1,2 and α-1,4 hydrolytic activity.

The coordination of the sucrose molecule into the active site is an important factor, and the residues surrounding the glucose unit in the -1 subsite likely stabilize the sucrose molecule to allow for cleavage. A unique characteristic of the α-1,2 linkage in sucrose is that it is able to attain an inter-saccharide angle more acute than that seen in α-1,4 linkages such as maltose, with bond angles of 108° and 120° respectively (Takusagawa and Jacobson 1978; Taga et al. 1993). This is illustrated in Barker and Rose 2013, in which the angle of the backbone of the kojibiose (α-1,2 linkage) is very acute, which is thought to be one of the main determinants of substrate specificity in α-Glucosidase I. Thus the sugar must be coordinated into the active site appropriately for effective cleavage, and the tyrosine residue found in the -1 subsite (residue

573) may play an important role in this process as well as the two residues found in the active site pocket, phenylalanine (residue 605) and proline (residue 210).

A final structural difference seen in the homology model of ctSI is the presence of a cysteine residue at position 401. ctSI is the only enzymatic subunit to have a cysteine in this position; the equivalent residue in ctMGAM is valine (position 399), in ntMGAM this residue is leucine (position 401), and in ntSI, this residue is an isoleucine (position 430). The cysteine found in ctSI at position 401 is in close proximity to a cysteine residue (position 508) conserved in all five enzymatic subunits (Figure 3-11). The distance between the Cα atoms of these two residues is 7.6 Å, and although this is a significant distance, each residue is found on a flexible loop region, which may move to allow for the creation of a disulfide bond. This disulfide bond may stabilize this region of the protein and ultimately, stabilize the extended binding cleft. Additionally, disulfide bonds have been shown to be involved in regulation of protein activity as well as conformational changes (Fass 2012), thus the creation and elimination of a disulfide bond at this position may explain how the activity of ctSI is regulated through conformational change.

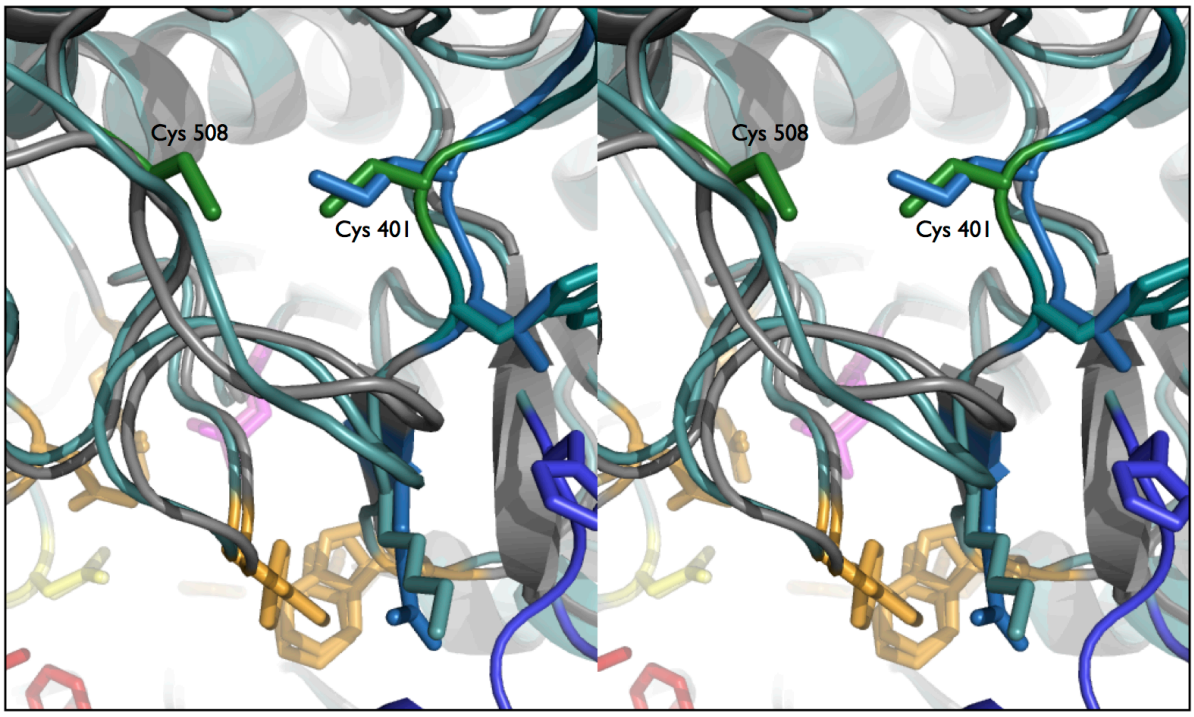

Figure 3-11: Stereo image of ctSI homology model (deep teal) superposed with ntSI (gray, PDB: 3LPO) (Sim et al. 2010). The cysteine residue found in position 401 in ctSI (green) and its potential partner in disufide bond formation, cysteine (position 508). The residue at the equivalent position in ntSI (Isoleucine, residue 430) is light blue.

There are also a number of amino acid variances found in the six amino acids preceding the 21-amino acid extension that are conserved in only the C-terminal subunits. Residue 402 is a tryptophan in the C-terminal catalytic subunits where as in the N-terminal enzymes, the amino acid at the equivalent position is isoleucine. The presence of this aromatic residue may allow for the extension of the binding site via CH-pi stacking interactions with the rings in sugar substrates.

Although these conclusions are based on small structural differences observed in the homology model of ctSI, there is evidence for the involvement of these residues and small structural changes in the differing activities of the catalytic subunits. The overall architecture of the homology model has a high degree of similarity to the known structures of ntSI (PDB: 3LPO) (Sim et al. 2010), ntMGAM (PDB: 2QLY) (Sim et al. 2008), and ctMGAM (PDB: 3TON) (Ren et al. 2011), even at the level of the active site. The concept that the structural changes observed in

the ctSI homology model are involved in substrate specificity and binding site extension is further supported by the fact that some of these structural differences are also seen in the crystal structure of ctMGAM (PDB: 3TON, 3TOP) (Ren et al. 2011).

Regulation of expression of the catalytic subunits likely also plays an important role in regulating the starch digesting activities in the small intestine. The presence of different substrates in the gut may influence the expression levels of each catalytic subunit. This is yet to be explored fully, and a study of expression profiles of each subunit under various conditions would be highly informative. Diet has been found to have a profound and rapid influence on gut microbiota (David et al. 2013), and further, diet has been found to influence the expression of maltase glucoamylase and sucrase isomaltase in multiple studies (Shimada, Mochizuki, and Goda 2009; Mochizuki, Hanai, et al. 2010a; Mochizuki, Honma, et al. 2010b; Mochizuki, Igawa-Tada, et al. 2010c). Thus, structural variance in addition to regulation of expression is the likely manner in which the activity of sucrase isomaltase and maltase glucoamylase in the mammalian gut are regulated to allow for digestion of a wide array of carbohydrate units.

3.5.2 ctMGAM-N20

The hydrolytic function of ctMGAM was initially thought to be constrained to α-1,4 linkages. Recently, it was discovered that similar to ntSI, ctMGAM-N20 demonstrates the ability to hydrolyze isomaltase, an α-1,6-glycosidic linkage. The length of substrate that can undergo this type of hydrolysis is unknown. As with ctSI, this C-terminal catalytic subunit has a higher affinity for longer inhibitors, such as acarbose, in comparison to the N-terminal substrates. Inhibitor studies have also shown additional differentiation between the ctMGAM-N2 and ctMGAM-N20 isoforms, with small inhibitors. De-*O*-sulfonated ponkoranol, its 5'stereoisomer, the selenium analogue of de-*O*-sulfonated ponkoranol, its 5'stereoisomer, and 3'-*O*-β-Maltosyl-de-*O*-sulfonated ponkoranol inhibit ctMGAM-N20, but have no inhibitory effect on ctMGAM-N2.

Although the model of ctMGAM-N20 is not as definitive as a crystal structure, looking at the model in the context of the known structure of other catalytic subunits and other similar enzymes provides some information that helps to better understand the functionality of the enzyme.

The shape and location of the binding site is different than the other catalytic subunits. Unlike the other catalytic subunits, the entrance to the binding site appears to be located adjacent to

the 89-amino acid extension, while the 21-amino acid extension appears to flank the side of the binding site. In the other two C-terminal subunits, the 21-amino acid extension appears to extend the binding cleft while maintaining an open entrance to the binding site. In the homology model of ctMGAM-N20, the 89-amino acid extension causes much of the active site to be buried within the protein structure. Additionally, this extension also creates a narrow entrance to the binding site.

The helical structure of the 89-amino acid extension does provide some clues as to how longer substrates may be stabilized. Residues lining the entrance to the active site, two of which reside on one of the α-helices of the 89-amino acid extension (histidine, residue 686 and proline, residue 683) and one on a loop near the active site entrance (proline, residue 207), may interact with the substrate (Figure 3-12). This may help to stabilize the longer substrates in the active site. ctMGAM-N2 and ctMGAM-N20 have the highest affinity for acarbose, with K_i values of 9 nM and 28 nM respectively, and both have residues at the entrance to the active site that may allow for interaction with the substrate. In ctMGAM-N2, this has been confirmed in the crystal structure of human ctMGAM (equivalent to murine ctMGAM-N2), with the substrate interacting with a tryptophan on the 21-amino acid extension. The presence of the 89-amino acid extension is also the main difference between the two isoforms ctMGAM-N2 and ctMGAM-N20 and may be a possible explanation for the selectivity seen with small inhibitors, such as de-*O*-sulfonated ponkoranol.

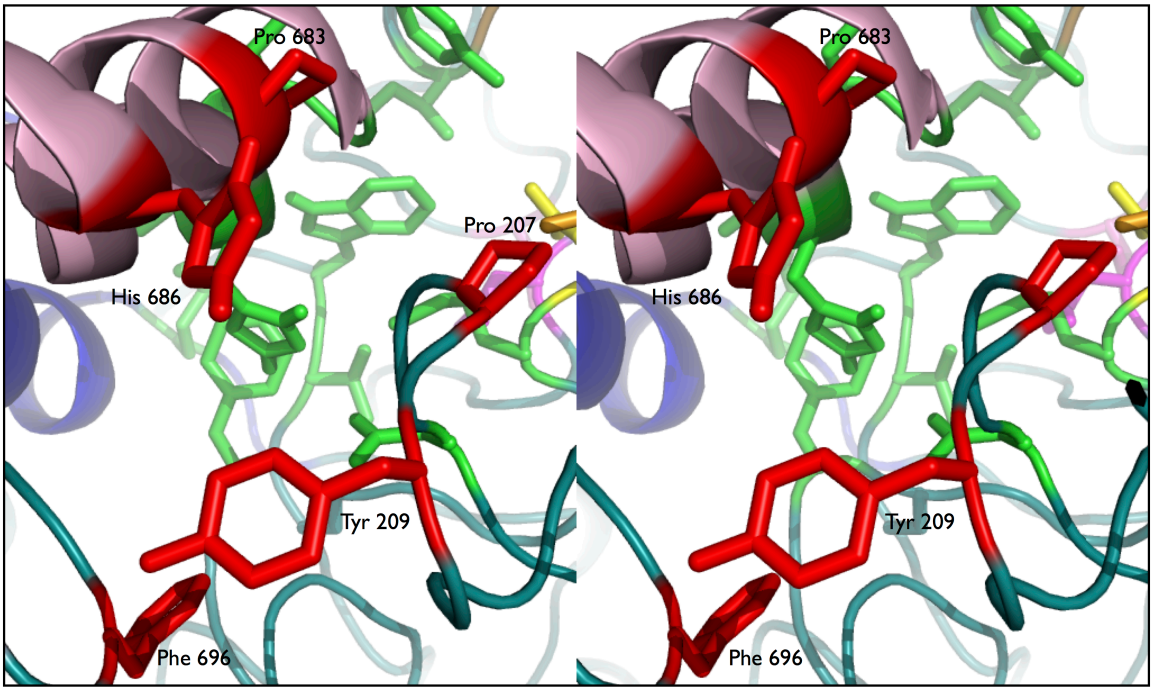

Figure 3-12: Stereo image of the proposed active site in ctMGAM-N20 homology model showing residues lining the entrance to the active site (red). Pink: 89-amino acid extension; Blue: 21-amino acid extension; Green: predicted active site residues; Orange: active site residues conserved in ntSI; Magenta: proposed nucleophile; Yellow: proposed acid/base catalyst.

There are also similarities present in the ctMGAM-N20 homology model seen in other glycoside hydrolases that demonstrate hydrolysis of α-1,6 bonds. Isomaltase (PDB: 3AXH) (Yamamoto et al. 2010), a family 13 glycoside hydrolase from *Saccharomyces cerevisiae*, is able to hydrolyze α-1,6 bonds, although there is one main difference between this protein and ctMGAM-N20: isomaltase is inhibited by maltose. Despite this difference, there is a structure of isomaltose with maltose in the active site, which can offer some insight into ctMGAM-N20. One of the main similarities is the narrow opening into the active site pocket. In the crystal structure of isomaltase, there is a short loop and two aromatic residues that constrain the entrance of the pocket. There is a similar constriction seen in the active site pocket of ctMGAM-N20, by a small loop (residues 571-578), aspartic acid (residue 602), proline (residue 207), and threonine (residue 679) (Figure 3-13). This proline residue is conserved in the C-terminal catalytic subunits. In the analysis of isomaltase, the authors concluded the tyrosine residue prevents the

binding of isomaltotriose (Yamamoto et al. 2010). In the case of ctMGAM-N20, the residues at the active site opening are not as bulky, but may assist in coordinating the binding of extended substrates. This loop may be flexible, as the ends of the loop are alanine and glycine residues, but this cannot be confirmed without a crystal structure.

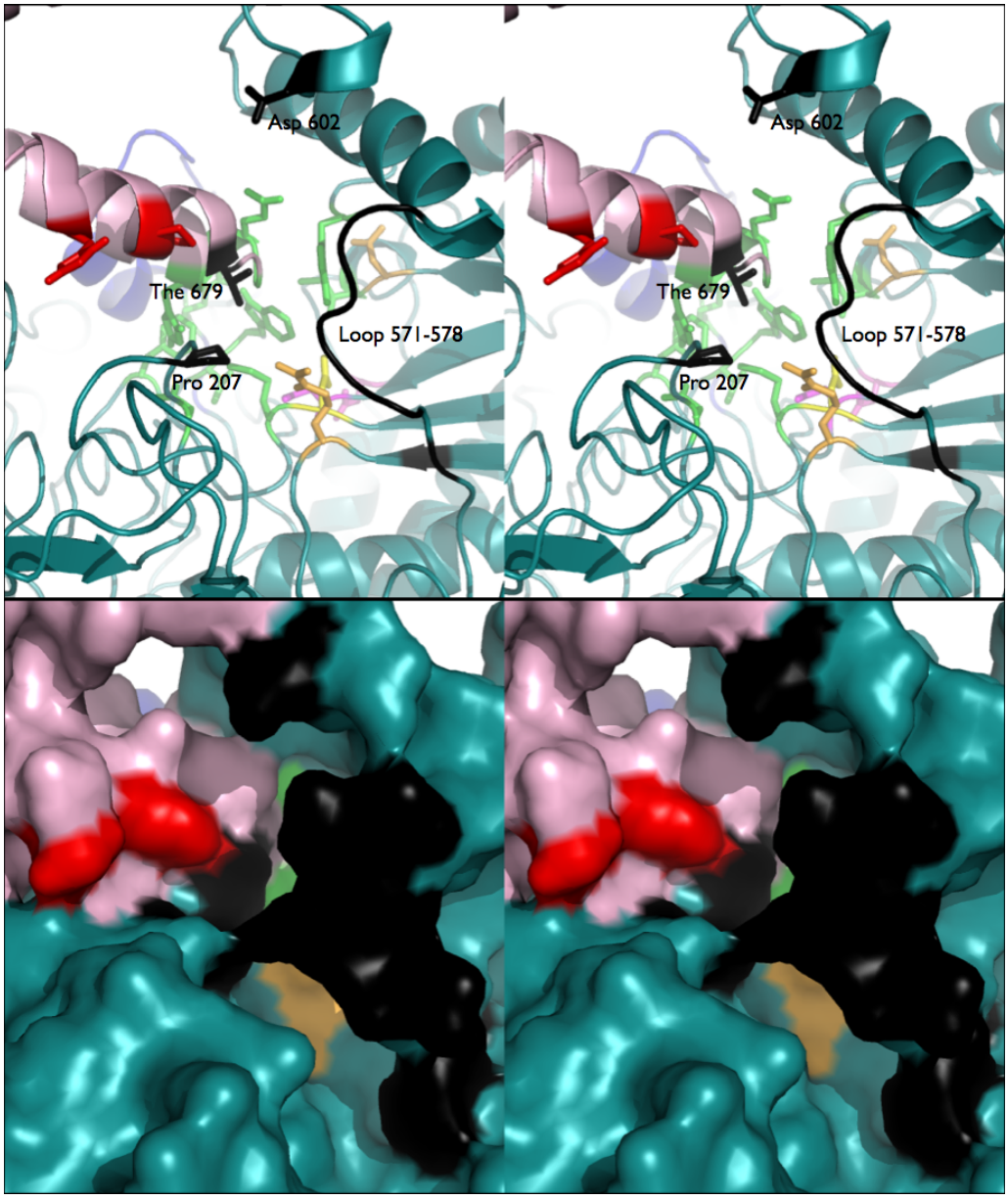

Figure 3-13: Stereo images of active site constriction in ctMGAM-N20. A: Stick representation of active site opening showing residues and loop that may restrict entrance to the active site (Black). **B:** Surface representation of ctMGAM-N20 showing the residues and loop that may restrict entrance to the active site (Black). Legend: Red: potential binding residues in 89-amino acid extension; Pink: 89-amino acid extension; Green: predicted active site residues; Yellow: proposed acid/base catalyst; Orange: conserved residues in ntSI and ctMGAM-N20. Blue: 21-amino acid extension.

There is also a tyrosine residue (residue 298) near the opening of the active site in ctMGAM-N20, similar to that seen in isomaltase, but instead of constricting the active site entrance, it is displaced from the opening and found between the +1 and +2 subsites. The displacement of the tyrosine residue and its location in the active site is somewhat similar to that of ctSI. In ctSI, the glucose moiety of the sucrose substrate is hypothesized to bind in the -1 subsite, coordinated by a tyrosine residue. In the case of ctMGAM-N20, the tyrosine residue found between the +1 and +2 subsites may hold a similar purpose, and in particular assist coordinating substrates with α-1,6 linkages. This is supported by the evidence seen in the isomaltase structure bound to its inhibitor, maltose. The maltose is positioned in and makes contact with the residues in -1 subsite in an almost identical manner to family 13 glycoside hydrolases capable of hydrolyzing the α-1,4 glycosidic linkage. The rationale for the inhibitory activity of maltose is that the contacts in the +1 subsite differ from other glycoside hydrolases, and the complex cannot reach the transition state as the dihedral angle of the α-1,4 glycosidic linkage does not have enough energy for hydrolysis (Yamamoto et al. 2010). Thus, the offset placement of the tyrosine in the active site may allow the substrates to be coordinated in the active site to allow for both α-1,4 and α-1,6 hydrolysis.

Although the most divergent with respect to sequence identity, the most similar enzymatic subunit in terms of hydrolytic activity is ntSI, as both ntSI and ctMGAM-N20 demonstrate dual hydrolytic activity on α-1,4 and α-1,6 glycosidic linkages. The specific differences in the pocket of the active site are difficult to pinpoint, as the catalytic domain in the ctMGAM-N20 homology model is not as well resolved as a crystal structure. In the case of the ctSI homology model, the active site was highly similar to that of the other catalytic subunits, and thus these known structures were excellent templates for modeling. The active site in the ctMGAM-N20 homology model is not as well modeled likely because there was no template for the 89-amino acid insertion. Further, the active site in ctMGAM-N20 is dissimilar enough from the other catalytic subunits of known structure that these small differences are difficult to identify. One marked difference between ntSI and ctMGAM-N20 is the opening to the active site. ntSI has neither the 21-amino acid nor 89-amino acid extensions, and thus the binding site is highly accessible for bulky substrates. In contrast, the ctMGAM-N20 homology model illustrates a tight entrance to the active site. It is unknown if there is movement of the 89-amino acid extension and other flexible loops to accommodate larger substrates, as this can only be answered through further

structural characterization. Finally, in contrast to other catalytic subunits, the spectrum of substrates used by ctMGAM-N20 has not been fully characterized. Thus, looking at the activity of ctMGAM-N20 with realistic substrates such as hydrolyzed waxy cornstarch and various length substrates such as isomaltotriose are a priority for future studies.

As ctMGAM-N2 and ctMGAM-N20 are differentially expressed alternatively spliced proteins originating from the same mRNA (Quezada-Calvillo, personal communication), regulation of glucosidase activities in the small intestine is likely done through regulation of expression. As with ctSI, the presence of starch substrates in the gut likely influences the expression levels of each catalytic subunit and the alternative splicing seen in ctMGAM.

3.6 Conclusion

Homology models are not as definitive as crystallographic models, however when examined in context of structural knowledge of other homologous catalytic subunits, known functionality, inhibitor studies, and other glycoside hydrolase structures, these models can provide valuable information. In the homology model of ctSI, the 21-amino acid extension seen in the C-terminal catalytic subunits has a variance in sequence. The tryptophan shown to interact with the acarbose substrate in the crystal structure of ctMGAM (PDB: 3TOP) (Ren et al. 2011) is replaced with a glutamic acid in ctSI, a possible explanation for the small degree of selectivity seen between ctMGAM and ctSI. In addition, two residues found only in ctSI have been identified that likely play a role in the hydrolysis of α-1,2 linkages as well as disulfide bond formation. Cysteine at position 401 may be involved in a disulfide bond with nearby residue cysteine 508. The formation and disruption of this bond may be involved in regulating hydrolytic activity and conformational change. Further, a tyrosine residue (residue 573) found in the ctSI active site may be a plausible explanation for the ability of ctSI to hydrolyze α-1,2 linkages. Finally, whereas the ntSI active site is open and highly accessible, the ctSI active site is partially occluded by the 21-amino acid extension. This may explain why ntSI is capable of debranching activity whereas ctSI is not.

In ctMGAM-N20, the homology model illustrates a narrowed entrance to the active site as well as an 89-amino acid extension seen only in ctMGAM-N20. This active site has some similarity to other glycoside hydrolases capable of hydrolyzing α-1,6-glycosidic linkages, which may explain the hydrolytic activity of ctMGAM-N20 on these types of glycosidic bonds. A

tyrosine residue located between the +1 and +2 subsites, may help coordinate the substrate in the active site, allowing for the duality of hydrolysis on both α-1,4 and α-1,6 bonds. Finally, the 89-amino acid extension is the main difference seen between ctMGAM-N20 and ctMGAM-N2 that may explain the selectivity demonstrated by small inhibitors such as de-*O*-sulfonated ponkoranol.

When structural determination proves difficult, homology models can be inspected in the context of current knowledge, providing valuable information to lead future studies. For example, in this case, hypotheses presented about the roles of specific residues can be tested by mutagenesis experiments.

Chapter 4
Deficiencies of Enzyme Function in Congenital Sucrase-Isomaltase Deficiency: Mutation Analysis

4.1 Overview

Mutations associated with patients with congenital sucrase isomaltase deficiency (CSID) have been identified in both the isomaltase (ntSI) and sucrase (ctSI) subunits of SI. These mutations have been shown to negatively impact enzyme activity, trafficking of the enzyme to the intestinal lumen, or both. In this study, the structure of ntSI (PDB: 3LPO) (Sim et al. 2010) was modified to reflect the mutations found in CSID patients and better understand the effect of the mutation on the activity of the catalytic subunit. Two mutations have been identified in which enzymatic activity is maintained: L340P and Q117R; and two mutations have been identified in which enzymatic activity is either completely or partially abolished: C635R and L620P (Jacob et al. 2000; Spodsberg 2001; Ritz et al. 2003; Keiser et al. 2006; Jones et al. 2012).

Mutations have been found in both the isomaltase and sucrase subunits of SI, however the mutations identified in the sucrase (ctSI) subunit mainly disrupt folding of the enzyme and thus, were not modeled in this study (Naim, Heine, and Zimmer 2012). In this chapter, the mutations found in the ntSI subunit are modeled onto the structure of ntSI and possible effects of these mutations on enzymatic function are postulated. In addition, the known missense SNPs in the exons of the SI gene will be mapped onto the model of ctSI discussed in Chapter 3, as well as the ntSI crystal structure (PDB: 3LPO). The SNPs will be compared to known mutations in the N- and C-terminal subunits of SI associated with CSID.

4.2 Contributions

Collaborator Hassan Y. Naim and his colleagues reported the mutations found in CSID patients. Sections of this chapter are modified from or taken verbatim from a report prepared by the candidate for BIOL 614: Bioinformatics – Tools and Techniques as well as the following publication:

K. Jones, R. Eskandari, H.Y. Naim, B.M. Pinto, D.R. Rose. Journal of Pediatric Gastroenterology and Nutrition. 55:Suppl2 (2012) S20-S24.

4.3 Materials and Methods

4.3.1 Mutation Models

Mutations were made to the deposited structure of ntSI (PDB: 3LPO) (Sim et al. 2010) using COOT (Emsley and Cowtan 2004) and energy minimization was completed on the mutated PDB files using ModRefiner (Xu and Zhang 2011). Graphical visualization and all figures were created using PyMol (DeLano Scientific). The method used to create the model of ctSI is discussed in section 3.3.2.1. SNP Mapping and Analysis

SNPs present in the exons of ntSI and ctSI were obtained from NCBI dbSNP (Sherry et al. 2001). All visualizations of ntSI and ctSI with SNPs indicated on the structures were created using PyMol (DeLano Scientific). SNPs present in the exons of the SI gene were queried in the Chromosome 3 data in the 1000 Genomes Database (Consortium et al. 2012) to determine the prevalence of the SNP in the sequencing data available as well as attempt to correlate SNPs related to CSID to ancestry and diet. Graphical representations of SNP population data were created using R (R Core Team 2013).

4.4 Results

4.4.1 L340 and Q117R

Enzymatic subunits with mutations L340P and Q117R have been shown to maintain enzymatic activity, but are unfortunately not trafficked to the small intestinal lumen. Figure 4-1 illustrates the catalytic site of ntSI including the acid/base catalyst and catalytic nucleophile, substrate binding residues, and the respective mutations.

4.4.2 C635R and L620P

Enzymatic subunits with mutations C635R and L620P abolish and partially abolish enzymatic activity, respectively. Figure 4-2 below illustrates the catalytic site of ntSI including the acid/base catalyst and catalytic nucleophile, substrate binding residues, and the respective mutations.

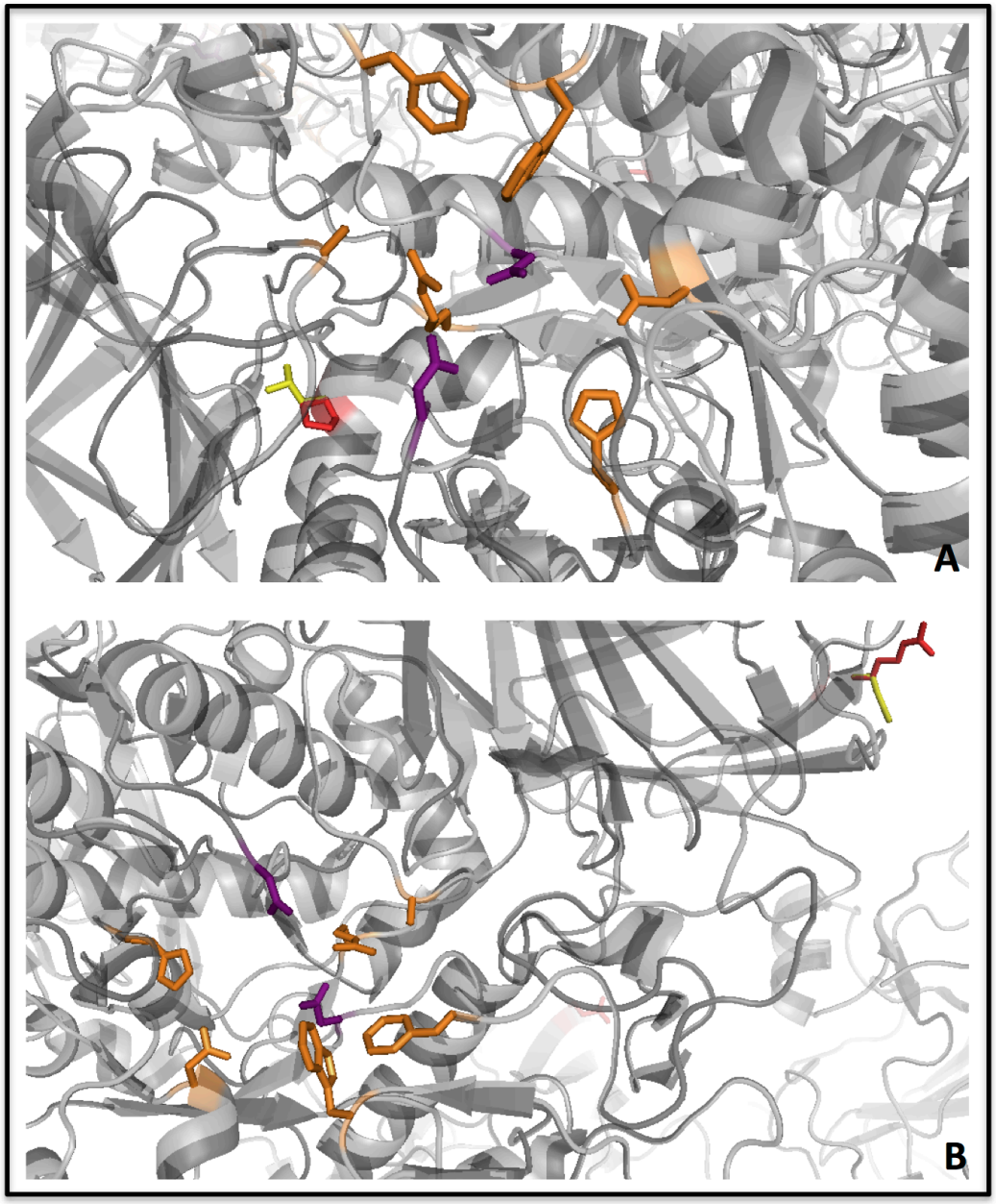

Figure 4-1: L340P and Q117R mutations. A, L340P mutation and ntSI active site. **B**, Q117R mutation and ntSI active site. **Purple**: catalytic nucleophile and acid/base catalyst; **orange**: substrate-binding residues; **yellow**: original residue; **red**: replaced residue (Jones et al. 2012).

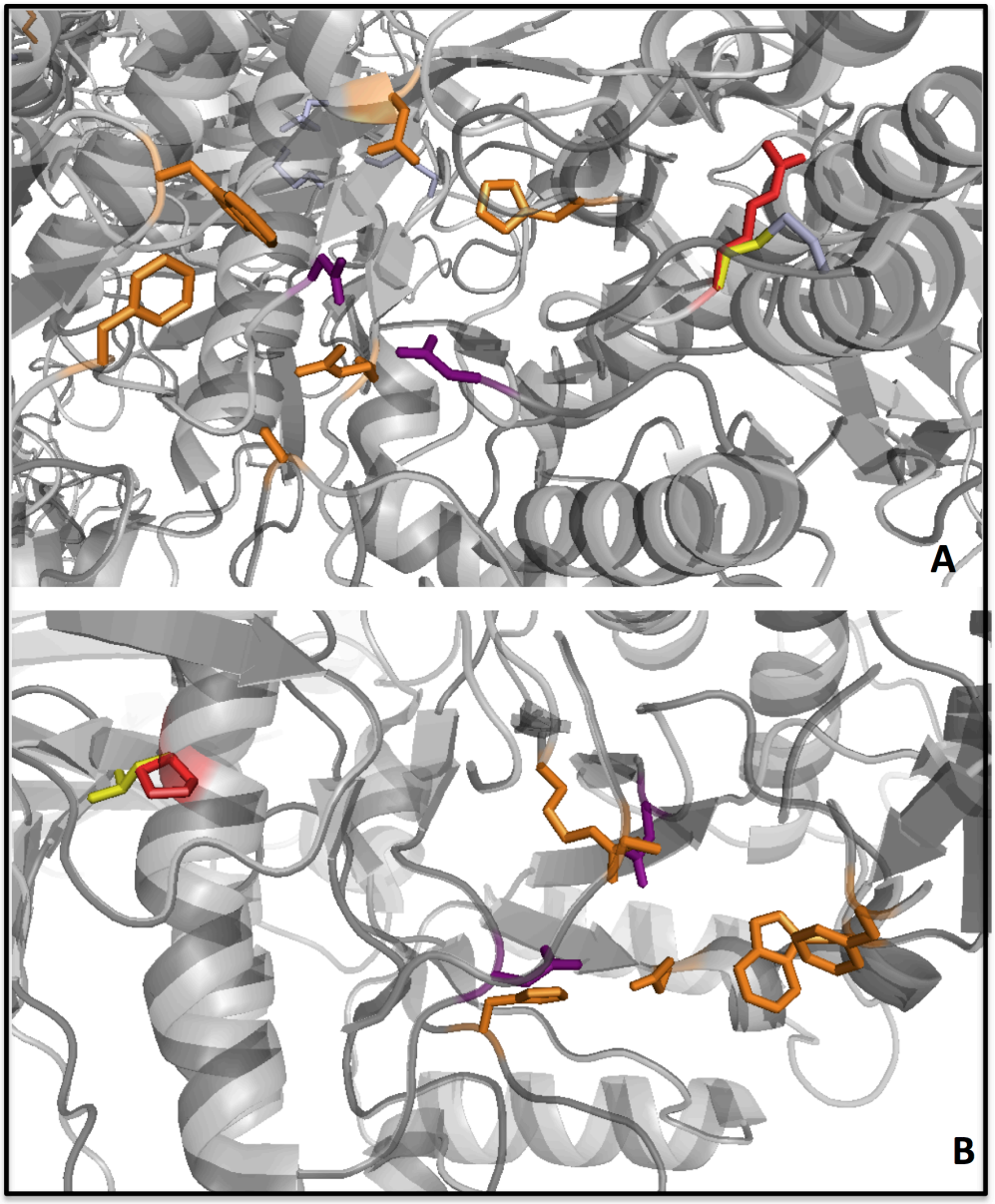

Figure 4-2: C635R and L620P mutations. A, C635R mutation and ntSI active site. **B**, L620P mutation and ntSI active site. **Purple**: catalytic nucleophile and acid/base catalyst; **orange**: substrate-binding residues; **yellow**: original residue; **red**: replaced residue (Jones et al. 2012).

4.4.3 SNP Mapping

The following table summarizes the missense SNPs found in the coding region of the SI gene mapped to the SI protein NP_001032.

Table 4-1: Missense SNPs in SI protein (NP_001032)

SNP ID	Protein Location (NP_001032)	Amino Acid Change	Corresponds to CSID Mutation?	SNP ID	Protein Location (NP_001032)	Amino Acid Change	Corresponds to CSID Mutation?	SNP ID	Protein Location (NP_001032)	Amino Acid Change	Corresponds to CSID Mutation?
rs141620268	71	V to A		rs193921060	698	P to L		rs149783130	1261	D to N	
rs200552633	99	N to D		rs188320908	717	V to D		rs200047372	1270	T to A	
rs138564183	105	W to C		rs143885457	774	R to K		rs138283071	1330	C to F	
rs146006371	108	F to V		rs147207752	774	R to G		rs141495440	1343	I to T	
rs149498200	109	V to I		rs138275610	781	Y to F		rs192640245	1348	T to M	
rs121912612	117	Q to R	Yes	rs149382090	790	H to R		rs143388292	1367	R to G	
rs138092600	144	N to S		rs138778317	794	G to S		rs148831941	1378	I to S	
rs150373520	154	T to A		rs150246328	799	I to V		rs145734588	1414	E to K	
rs139012673	157	R to G		rs200972419	801	E to STOP		rs142090504	1417	Y to STOP	
rs201105530	159	R to W		rs201944977	839	D to N		rs145246112	1484	T to H	
rs146960446	173	P to A		rs200498454	864	H to Q		rs201282506	1484	R to C	
rs193172680	184	T to I		rs140230726	867	Y to H		rs4855271	1523	M to I	
rs142447888	186	S to P		rs80339263	883	L to STOP		rs139908762	1574	T to S	
rs34213113	191	D to A		rs141329406	884	T to K		rs144187804	1598	H to Q	
rs142923909	198	P to L		rs146647167	887	V to A		rs200328403	1609	R to STOP	
rs113903766	200	S to G		rs145893756	893	A to V		rs139621295	1614	E to A	
rs199559251	229	I to F		rs200768342	905	F to C		rs202225928	1617	D to H	
rs9283633	231	T to A		rs200328496	913	V to F		rs146899814	1623	D to A	
rs79468450	231	T to S		rs199706219	914	L to I		rs149414344	1625	F to V	
rs200745562	250	R to C		rs150927256	930	Q to R		rs139876383	1651	V to I	
rs146187373	262	T to I		rs148434280	963	R to K		rs150702232	1654	A to G	
rs143135955	263	R to Q		rs144419320	965	G to D		rs113739230	1660	H to R	
rs201140811	266	L to F		rs146785675	975	Y to H		rs202004001	1661	T to I	
rs148454534	311	I to V		rs147226975	1000	Q to P		rs142018224	1667	V to I	
rs201980143	331	P to S		rs143631771	1001	L to I		rs142018224	1667	V to L	
rs77546399	348	P to L		rs111316957	1005	N to T		rs191598487	1686	R to H	
rs143782842	366	D to E		rs139183302	1021	E to Q		rs138524473	1694	Q to K	
rs138434001	371	V to M		rs201018248	1022	V to M		rs145773793	1700	T to K	
rs150428844	389	I to N		rs121912616	1073	G to D	Yes	rs113023426	1715	D to G	
rs150428844	389	I to T		rs121912611	1098	Q to P	Yes	rs143378401	1717	N to D	
rs75188774	399	T to I		rs138397431	1105	R to C		rs146956807	1737	R to G	
rs202151060	404	A to G		rs145913658	1111	I to V		rs79717168	1745	F to C	Yes
rs148511215	422	Y to D		rs140725570	1120	T to I		rs79717168	1745	F to S	
rs112053533	432	I to V		rs200451408	1124	R to STOP		rs79717168	1745	F to Y	
rs141450837	460	S to C		rs201739571	1124	R to Q		rs187532711	1749	Q to K	
rs143176463	491	S to I		rs190796723	1132	G to R		rs145556619	1760	G to D	
rs202203549	495	Q to H		rs138927818	1136	R to S		rs145556619	1760	G to V	
rs144972103	515	G to V		rs202077921	1140	P to A		rs199881454	1767	T to M	
rs200607628	543	T to K		rs140049686	1143	K to N		rs140894992	1778	G to R	
rs149898910	565	S to T		rs151040229	1155	A to S		rs146501096	1788	T to M	
rs199996048	573	A to V		rs142205728	1158	E to Q		rs139504152	1795	S to L	
rs121912615	577	V to G		rs185479515	1187	T to I		rs9917722	1802	T to S	
rs201583834	588	R to C		rs147619761	1194	F to Y		rs149654947	1808	R to H	
rs201055347	613	E to STOP		rs78013297	1200	P to S		rs201372251	1808	R to C	
rs121912613	620	L to P	Yes	rs121912614	1229	C to Y	Yes				
rs142789249	640	E to G		rs41273567	1239	V to L					
rs140611221	657	P to S		rs144579834	1258	T to A					

After identification, the sequence of NP_001032 with SNPs annotated was aligned with the murine ctSI sequence (model discussed in Chapter 3) to determine missense SNP locations in the murine homology model. These were then represented on the ctSI homology model. Similarly, the missense SNPs found within the exons of the ntSI subunit were represented visually on the ntSI crystallographic model (PDB: 3LPO) (Sim et al. 2010). The SNPs that have already been correlated to the occurrence of CSID are indicated in the column "Corresponds to CSID Mutation?". There is no evidence that the other SNPs found in the SI gene are correlated to

the occurrence of CSID, however at the present time there is also no evidence to the contrary and thus the possibility remains that they may play a role in the disease state of CSID.

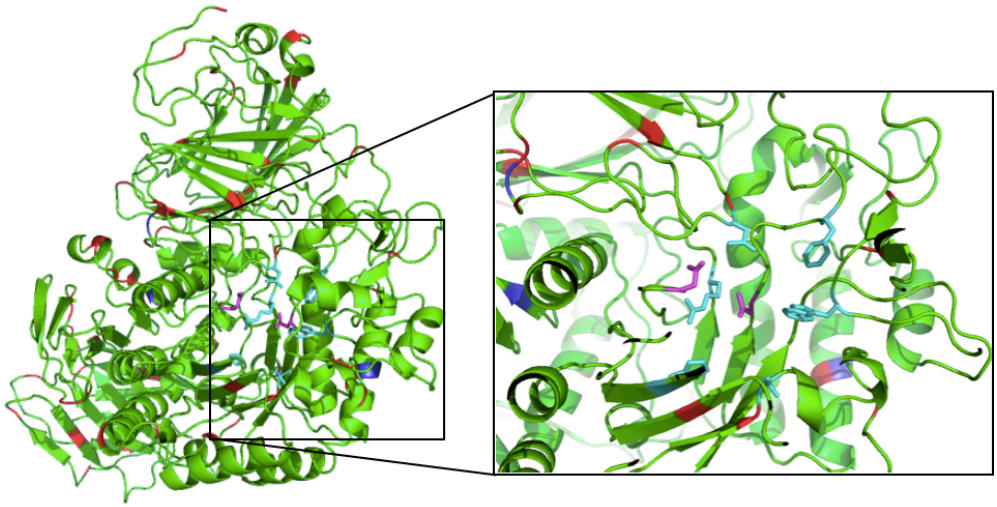

Figure 4-3: ctSI homology model with missense SNPs indicated and a view of the active-site pocket. Legend: nucleophile and acid/base catalyst (magenta); binding residues (cyan); missense SNPs (red) SNPs resulting in a stop codon (blue).

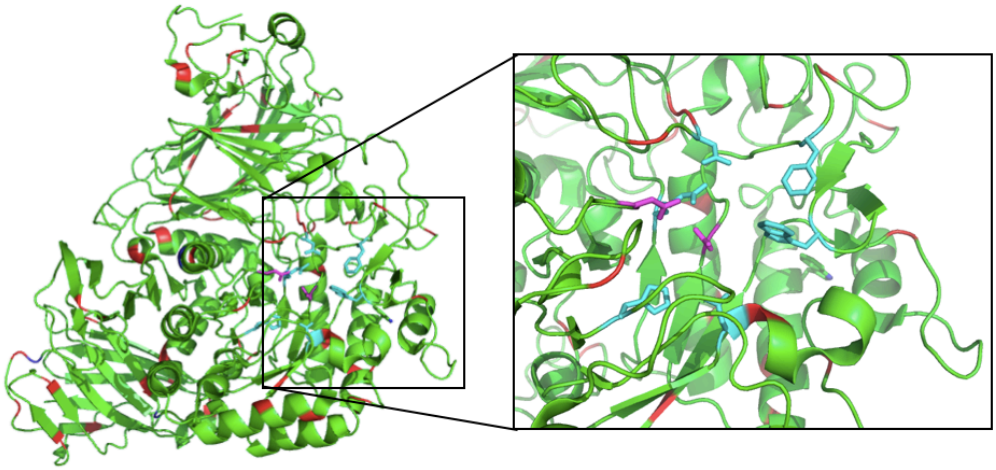

Figure 4-4: ntSI (PDB: 3LPO) with missense SNPs indicated and a view of the active-site pocket. Legend: nucleophile and acid/base catalyst (magenta); binding residues (cyan); missense SNPs (red) SNPs resulting in a stop codon (blue).

4.5 Discussion

During translocation across the endoplasmic reticulum (ER) membrane as well as in the ER lumen, events associated with protein maturation take place including co-translational glycosylation (Kornfeld and Kornfeld 2002), intermolecular or intramolecular disulfide bond formation (Gething and Sambrook 1992), and subunit assembly. Additionally, in the Golgi apparatus, posttranslational modification such as N- and O-linked glycosylation (Kornfeld and Kornfeld 2002) occurs and errors in this process can impact protein trafficking, secretion, and protein sorting in epithelial cells. These processes are negatively impacted by two of the mutations found in the ntSI subunit in CSID patients (Naim, Heine, and Zimmer 2012).

4.5.1 L340P

The first mutation, in which a leucine is replaced with a proline at position 340 (corresponding to residue number 308 in PDB file 3LPO), results in an active protein. The protein is cleaved in the endoplasmic reticulum and secreted, but it is not accessible in the intestinal lumen to cleave sugars, presumably due to a trafficking defect (Jacob et al. 2000; Spodsberg 2001; Ritz et al. 2003; Keiser et al. 2006; Naim, Heine, and Zimmer 2012). The mutation is 28.5 Å from the catalytic nucleophile. Proline is added in place of leucine in the middle of a structural element, an α-helix at the surface of the protein monomer. The insertion of a rigid amino acid in the middle of this structural element likely disrupts the secondary structure in this region of the protein (Figure 4-1, A). We propose the failure of the protein to be properly trafficked to the intestinal lumen stems from this disruption in secondary structure. Although the structural change is not in close proximity to the active site of the catalytic subunit, it is near a predicted glycosylation site (ASN 119 in PDB file 3LPO, Figure 4-5, A), which is 12.6 Å from the mutation. The distortion of the mutated α-helix likely disrupts the neighbouring β-ladder containing asparagine 119. An additional consideration is that the mutation is on the surface of the protein located in a region in which two monomers of the protein interact in the crystal form. For proper trafficking, the N- and C-terminal enzymatic subunits of SI may interact with one another, and this mutation may prevent this interaction from occurring. Disruption of any of these properties may lead to the trafficking defects (Jones et al. 2012).

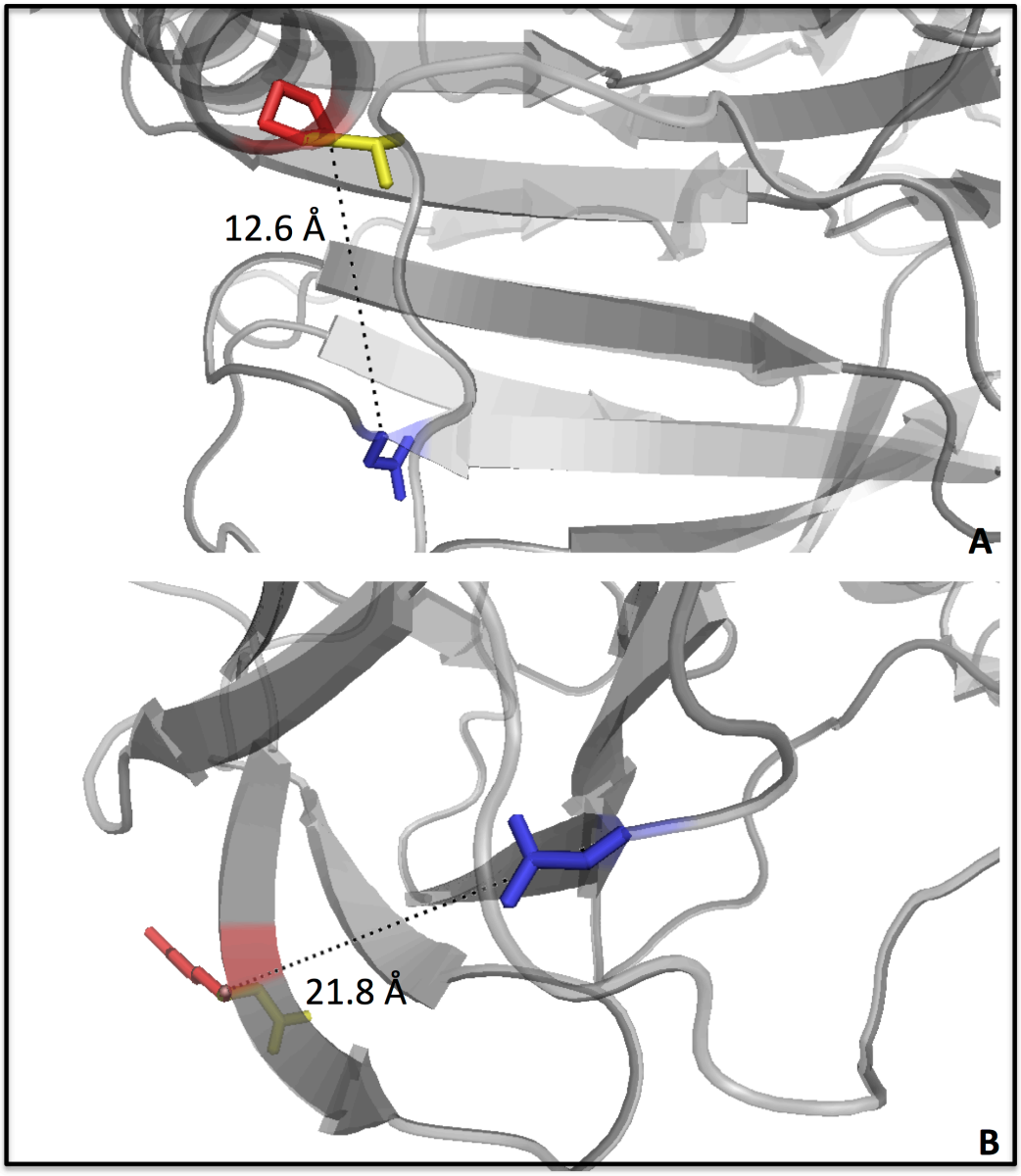

Figure 4-5: Glycosylation sites in vicinity of L340P and Q117R. A, L340P and predicted glycosylation site ASN 119. B, Q117R and predicted glycosylation site ASN 119. Yellow: original residue; red: mutated residue; blue: ASN 119 (Jones et al. 2012).

4.5.2 Q117R

A second mutation, in which a glutamine is replaced with an arginine at position 117 (corresponding to residue 84 in PDB 3LPO; Fig. 4-5, B), leads to a protein that is active from an

enzymatic standpoint, but no isomaltase activity is seen in patients with CSID. The catalytic subunit is not transported to the small intestinal lumen and therefore there is no functionality of ntSI in this location (Naim, Heine, and Zimmer 2012). The mutation is on the surface of the protein 52.2 Å from the catalytic nucleophile in the active site. Because of the distance from the active site, there is likely no direct effect on conformation, explaining the enzymatic functionality of the protein. We propose that similar to L340P, the mutation has a negative effect on trafficking of the catalytic subunit to the small intestinal lumen. This mutation is also in close proximity to the predicted glycosylation site ASN 119, with both residues found on the same β-ladder (Fig. 4-5, B). The trafficking defect is likely caused by a change in glycosylation at the surface of the protein as a result of the disruption to the secondary structure in this region, causing the signal pathway in which the enzyme is transported to the intestinal lumen to fail. The involvement of ASN 119 in two possible mutations involving a trafficking defect gives credence to the possibility that this residue is important in the trafficking pathway of ntSI and mutagenesis studies involving this residue should be further explored (Jones et al. 2012).

4.5.3 C635R

The third mutation, in which a cysteine is replaced with an arginine at position 635 (corresponding to residue 602 in the PDB file 3LPO, Fig. 4-2, A), results in a protein with only partial α-glucosidase activity. The distance of the mutation from the catalytic centre is significant, at 17.4 Å. Despite being so far removed from the active site, the removal of the cysteine disrupts the disulfide bond between the mutated residue, 635, and residue 613. The likely result is a significant disruption of secondary structure, disordering the active site and affecting function (Jones et al. 2012), particularly if the result of this disruption is the movement of the acid/base catalyst and catalytic nucleophile more than 5.5 Å away from one another. If these two side chains are too distant from one another, the enzyme will be unable to cleave glycoside linkages byway of the classic Koshland retaining mechanism and the enzyme will not be active (Cantarel et al. 2009; Jones et al. 2012).

4.5.4 L640P

The final mutation, in which a leucine is replaced by a proline at position 620 (corresponding to residue 587 in the PDB file 3LPO; Fig. 4-2, B), results in an inactive protein. Leucine is replaced by proline, a rigid amino acid, in the middle of a buried α-helix, which likely disrupts the helical

structure and associated secondary structure. The mutation is 21.2 Å from the catalytic nucleophile, but it is in the interior of the protein where proline is found only infrequently, as this residue typically lies on the surface of globular proteins. In addition, the α-helix borders the active site pocket. The disorder of the active site caused by this mutation likely affects function and causes the inactivity, especially if the distance between the acid/base catalyst and catalytic nucleophile is more than 5.5 Å away from one another, similar to the case of C635R.

4.5.5 SNPs

There are also many other mutations in the form of missense SNPs in the exons of both SI subunits, located in the binding pocket of the enzymes near the nucleophile, acid/base catalyst, and binding residues. There are many SNPs that one can propose will have an effect on enzymatic activity due to the proximity to the active site; thus, this analysis can help to direct future mutagenesis studies to explore the effect of the different SNPs on enzyme function. Figure 4-3 and 4-4 illustrate the active sites of the ctSI homology model and ntSI crystallographic model (PDB: 3LPO) (Sim et al. 2010) and the proximity of the SNPs to the nucleophile, acid/base catalyst, and substrate binding residues. The SNPs in the active site pocket that correlate to newly identified mutations involved in CSID have been prioritized for mutagenesis studies.

Two of the mutations identified in CSID patients, Q117R and L620P, are SNPs rs121912612 and rs121912613 respectively. Further, in the ctSI homology model, all four of the mutations seen in CSID patients correlate to SNPs rs121912616, rs121912611, rs121912614, and rs79717168 (See Table 4-1 for summary of SNPs found in SI gene exons). Correlating these mutations seen in CSID patients to SNPs present in the NCBI dbSNP database opens the possibility of mining data from SNP databases to investigate the link between ancestry, diet, and occurrence of CSID. To begin this investigation, the prevalence of the missense SNPs found in the exons of the SI gene (Chromosome 3:164,696,686..164796283) in various populations was explored using the 1000 Genomes Database (Consortium et al. 2012). Figure 4-6 illustrates the number of SNPs found in the SI gene that are also found in the 1000 Genomes Database. The number of SNPs is grouped by general ancestry (East Asian Ancestry, South Asian Ancestry, African Ancestry, European Ancestry, and Americas Ancestry). Despite having fewer samples in

the database, the highest numbers of SNPs in the exons of the SI gene were found in those of African ancestry whereas the least number of SNPs were found in those of Asian ancestry.

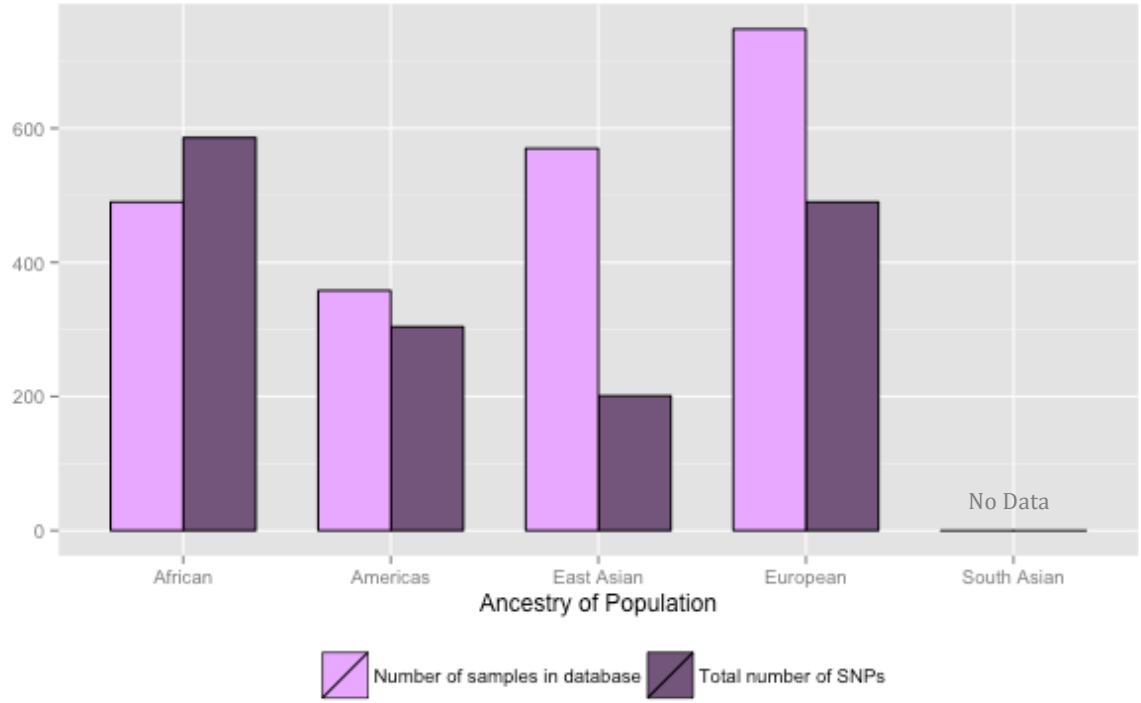

Figure 4-6: Barplot illustrating the size of the population database and total number of SNPs found in the exons of the SI gene in that population. Some populations have more SNPs than number of individuals because some individuals have multiple SNPs in the SI gene.

A representation of the number SNPs in a more detailed breakdown of individual populations is also shown in Figure 4-7. A relatively large number of SNPs are seen in populations of African ancestry, particularly Luhya in Webuye, Kenya (LWS) and Yoruba in Ibadan, Nigeria (YRI). The sample size and diversity of the available data is not nearly large enough to make conclusions about the prevalence of SNPs that are potentially related to CSID in relation to ancestry, but the data act as a starting point for future studies as more SNP data become available.

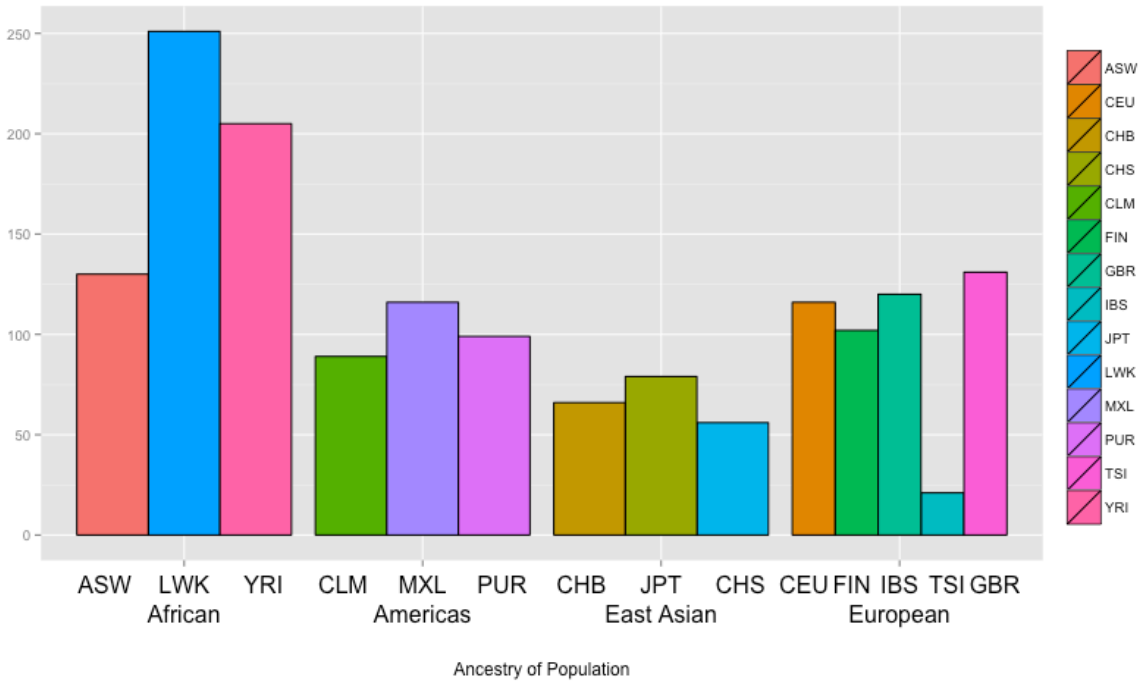

Figure 4-7: Number of SNPs found in each sample population. African Ancestry: ASW – African Ancestry in Southwest US, **LWK -** Luhya in Webuye, Kenya, **YRI –** Yoruba in Ibadan, Nigeria; **Americas: CLM –** Colombian in Medellin, Colombia, **MXL –** Mexican Ancestry in Los Angeles, California, **PUR -** Puerto Rican in Puerto Rico; **East Asian: CHB –** Han Chinese in Beijing, China, **JPT –** Japanese in Tokyo, Japan, **CHS –** Southern Han Chinese, China; **European: CEU –** Utah residents with Northern and Western European ancestry, **FIN –** Finnish in Finland, **IBS –** Iberian populations in Spain, **TSI –** Toscani in Italy, **GRB –** British in England and Scotland.

Unfortunately only one of the mutations found in previous studies (Naim, Heine, and Zimmer 2012) in CSID patients is also found in the 1000 Genomes Database, and the SNP occurs in only two individual samples of the 2178 samples in the database, indicating it is relatively rare. Both of these individuals are of European ancestry. This exercise highlights the need for genetic studies of populations with relatively high incidence of CSID. The understanding of the genetic elements involved in CSID would be further enhanced through the study of the population of Greenland. Approximately 5% of the population of Greenland suffers from CSID (Bell, Draper, and Bergan 1973; Ellestad-Sayed, Haworth, and Hildes 1978), thus determining whether this is caused by mutations found only in this population would be of interest (Naim, Heine, and

Zimmer 2012). We are only at the onset of being able to understand the role of SNPs, if any, in the occurrence of CSID and how multiple mutations found within the enzymatic subunits may compound, producing the wide range of clinical symptoms presented in patients.

4.6 Conclusion

The postulated effects of the mutations found in CSID patients on the functionality and trafficking of the ntSI subunit increases our understanding of the molecular basis of this dietary disorder. The four mutations in the ntSI subunit studied were found in four different individuals diagnosed with CSID, which indicates a wide range of mutations may lead to the clinical symptoms of CSID. Residues that potentially play important roles in the trafficking of the enzyme and protein-protein interaction, such as ASN 119, have been identified and are priority targets for future mutagenesis studies. Relating SNPs to enzymatic function must be done with caution, as seen in Eccleston et al. 2012, in which a homozygous deletion was found through SNP arrays in the maltase-glucoamylase gene, however from a clinical standpoint, maltase-glucoamylase was indeed functional. The four mutations explored in this chapter were validated clinically and the postulated effects on enzymatic function and/or trafficking correlate to the physiology seen in the patient.

The prevalence of CSID in different populations does not correlate to SNP data available at the present time. Because the symptoms of CSID are also seen in a variety of other digestive disorders and there is a lack of knowledge of this condition in the medical field, CSID is often under-diagnosed. Due to these factors, there is a lack of documentation of this condition and the epidemiological data is not yet complete enough for proper analysis. However, the number of SNPs seen in different ancestral populations does provoke interest and provides a starting point for future genetic studies. This highlights the need for genetic studies of populations with relatively high incidence of CSID, and the study of the African populations identified in this preliminary study and their diet will likely be informative. Thus, these results are incomplete in their current form, however they provide a viable starting point to explore the genetic component of CSID. As more data become publically available from genetic studies in different populations and through the exploration of ancestry and common dietary components, we hope to explore the role diet played in the evolution of starch digesting enzymes and the observed

overlapping substrate specificity allowing for the accommodation of diverse starches in the human diet.

Chapter 5
Structure of Family 31 Glycoside Hydrolase CfXyl31

5.1 Overview

In addition to producing homology models of ctSI and ctMGAM-N20, a 3 Å structure of another Family 31 glycoside hydrolase, CfXyl31, an α-xylosidase isolated from *Cellulomonas fimi*, has been determined. This is the first Family 31 glycoside hydrolase to be studied in *C. fimi*. In preliminary enzymatic characterization, this enzyme was found to have weak α-xylosidase activity.

Structural similarities have been found between CfXyl31 and mammalian intestinal α-glucosidases both at the level of global structure and active site architecture. Ultimately, these similarities can be exploited to use CfXyl31 as a model for MGAM and SI hydrolytic activity as this bacterial enzyme is expressed in a bacterial expression system as is much more amenable to mutation to facilitate the study of the roles of individual residues in substrate hydrolysis. Thus, in this chapter, the structure of CfXyl31 will be presented in addition to potential mutations that can be made to the structure in hopes of mimicking the α-glucosidase activity of ntMGAM. Following a proof of concept that mimicking ntMGAM is attainable, mutagenesis studies to mimic the remaining enzymatic subunits can be completed with the ultimate goal of engineering a "starchosome", in which each monomer in the hexameric structure of CfXyl31 can be modified to possess the activity of each individual mammalian catalytic subunit.

5.2 Contributions

The CfXyl31 protein purification and preliminary enzymatic characterization was done by Warren Wakarchuk's research group. The candidate completed all crystallization and structural work as well as a preliminary enzymatic assay with maltose as the substrate.

5.3 Materials and Methods

5.3.1 Crystallization

Prior to crystallization trials, the Hampton Pre-crystallization Test (Catalogue number: HR2-140) was used to estimate the appropriate protein concentration for crystallization. This was found to be approximately 15 mg/mL. Crystal screening was initially done in-house, generating a crystallization condition producing crystals with a thin, plate morphology (Figure 5-1). A sample of protein (concentration: 15 mg/mL) was sent to the Hauptman-Woodward Medical Research Institute (HWI) for high-throughput screening with the standard screen, testing 1536 conditions using the batch crystallization technique(Luft et al. 2003). Over the course of the screening, three conditions were found to produce crystals (Figure 5-1). These conditions were then reproduced using the vapour diffusion crystallization technique. For optimization, protein concentration was modified in the drops through modifying the ratio of protein and crystallization solution. Protein concentrations above and below that used in the HWI high-throughput screen were tested, producing crystals of rectangular prism morphology (Figure 5-1). The condition is the same as that found through the HWI high-throughput screen, but the drop ratio is 1 part reservoir to 3 parts protein with concentration 15 mg/mL.

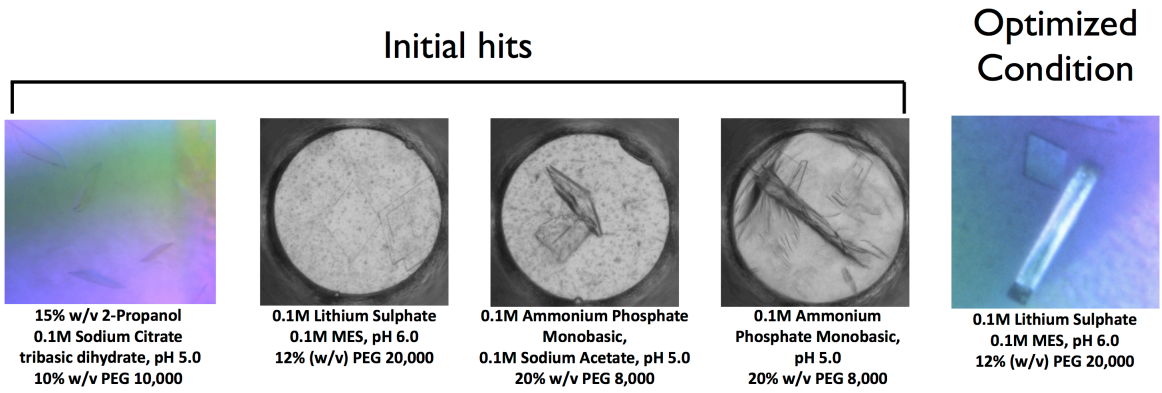

Figure 5-1: Evolution of the crystallization hits, CfXyl31.

5.3.2 Data Collection and Analysis

Data for CfXyl31 were collected on a home source (Rigaku) and processed with StructureStudio (D*TREK) (Pflugrath 1999). A diffraction pattern of a CfXyl31 crystal is shown (Figure 5-2) and statistics are summarized in Table 5-1.

Table 5-1: Collection and refinement statistics

Crystal Parameters	
Space Group	$P2_1$
Unit Cell	a: 142.0245 b: 96.7916 c: 194.3230
Data Collection	
Beamline	Rigaku
Wavelength (Å)	1.54
Resolution	73.32-2.90 (3.00-2.90)
# Unique Reflections	107384
Average Redundancy	2.40 (2.44)
Completeness (%)	91.7
R_{merge}	0.173 (0.506)
R_{meas}	0.216 (0.630)
I/σI	3.6 (1.3)
Refinement	
R_{Work}	0.2792
R_{Free}	0.3393
Ramachandran Favoured (%)	84.1
Ramachandran Outliers (%)	4.6
Rotamer Outliers (%)	11.1
RMS bonds (Å)	0.005
RMS angles (°)	1.297
# Residues	798
Average Protein B-factor (Å)	71.70

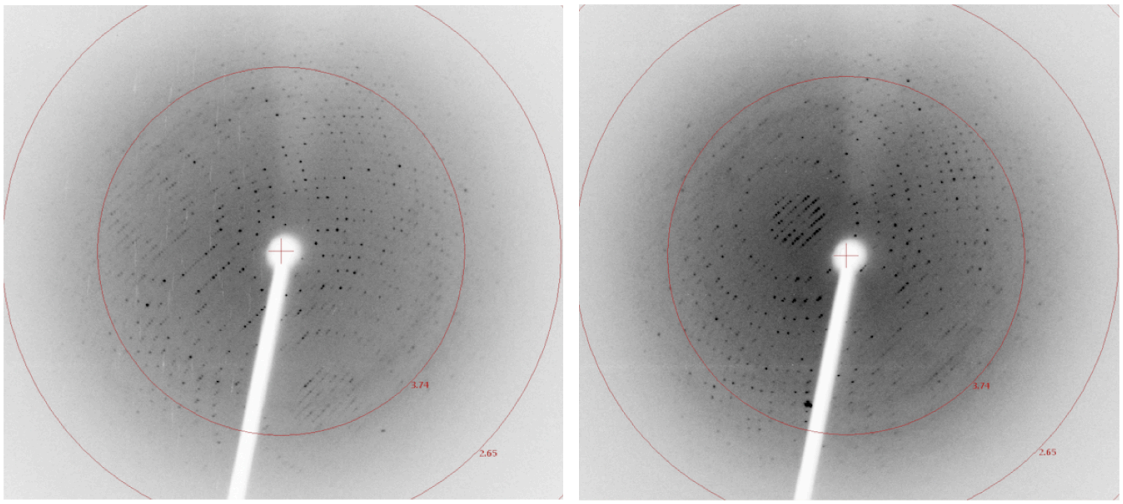

Figure 5-2: CfXyl31 Diffraction Patterns

5.3.3 Structure Phasing and Refinement

Structure phasing was completed with molecular replacement using homologous protein, Family 31 Glycoside Hydrolase YicI from *Escherichia coli* (PDB: 1XSI), as a search model. Phaser-MR in the Phenix (Adams et al. 2010) suite was used for molecular replacement, followed by AutoBuild. Phenix.refine was used for refinement. The crystal was found to be twinned (pseudo-merohedral), and twin law h, -k, -l was used in AutoBuild and subsequent refinement.

5.3.4 Active Site Inspection

The active site of CfXyl31 and residues hypothesized to be involved in substrate binding were predicted using structural and sequence alignments, comparing CfXyl31 to ntMGAM (PDB: 2QLY) (Sim et al. 2008) and YicI (PDB: 1XSI) (Lovering 2004).

5.3.5 Enzymatic Activity

As of the current time, the activity of this enzyme on various substrates has not been fully characterized. There is a small degree of activity in the presence of xylan (Personal communication, Warren Wakarchuk), leading to the conclusion this is likely an α-xylosidase. Maltose has also been tested as a substrate and a small degree of hydrolytic activity was

observed, however activity was so low that kinetics cannot be determined due to substrate solubility limits.

5.4 Results

5.4.1 Structure of CfXyl31

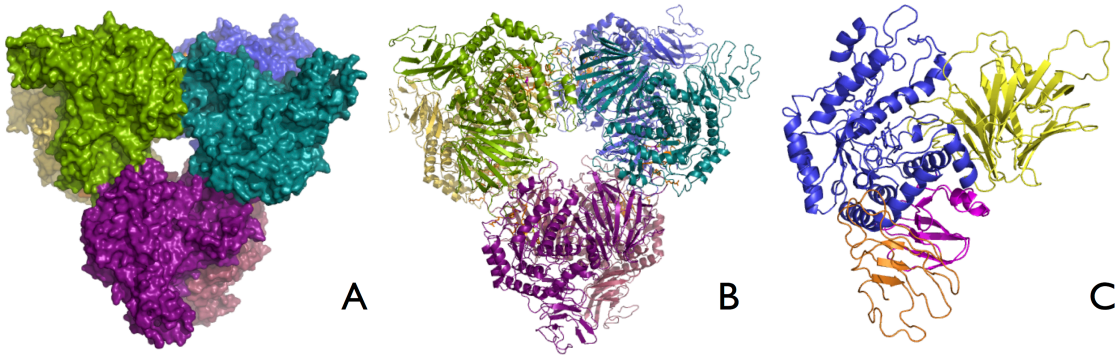

Figure 5-3: Structure of CfXyl31. A: Surface representation of the hexameric structure; **B:** Ribbon representation of hexameric structure; **C:** CfXyl31 monomer with structural domains indicated: Yellow: N-terminal β-sandwich domain; Blue: catalytic (β/α) barrel domain; Magenta: proximal C-terminal domain; Orange: distal C-terminal domain.

Analysis of the crystal packing of CfXyl31 indicates there are six molecules in asymmetric unit. The hexamer is composed of a dimer of trimers (Figure 5-3). Each monomer consists of five structural domains: N-terminal β-sandwich domain (residues 1-247); catalytic (β/α) domain (residues 248-584); proximal C-terminal domain (residues 585-664); and a distal C-terminal domain (residues 666-740) (Figure 5-3, C). There are six active sites in the hexamer, each at the point at which three monomers converge. The active site consists of the predicted acid/base catalyst (aspartic acid, residue 482) and nucleophile (aspartic acid, residue 416). Both the predicted acid/base catalyst and predicted nucleophile are located on Chain A. Residues from three monomers within the hexamer converge to create the active site (Figure 5-4). Most of the binding residues are on one monomer (Chain A) with the acid/base catalyst and nucleophile: Phe277, Trp315, Trp345, Trp380, Phe417, Trp479, and Phe515. Three additional residues from two different monomers also contribute to substrate binding: Asp49 (Chain E)

and Tyr7 and Trp8 (Chain D) (Figure 5-5).

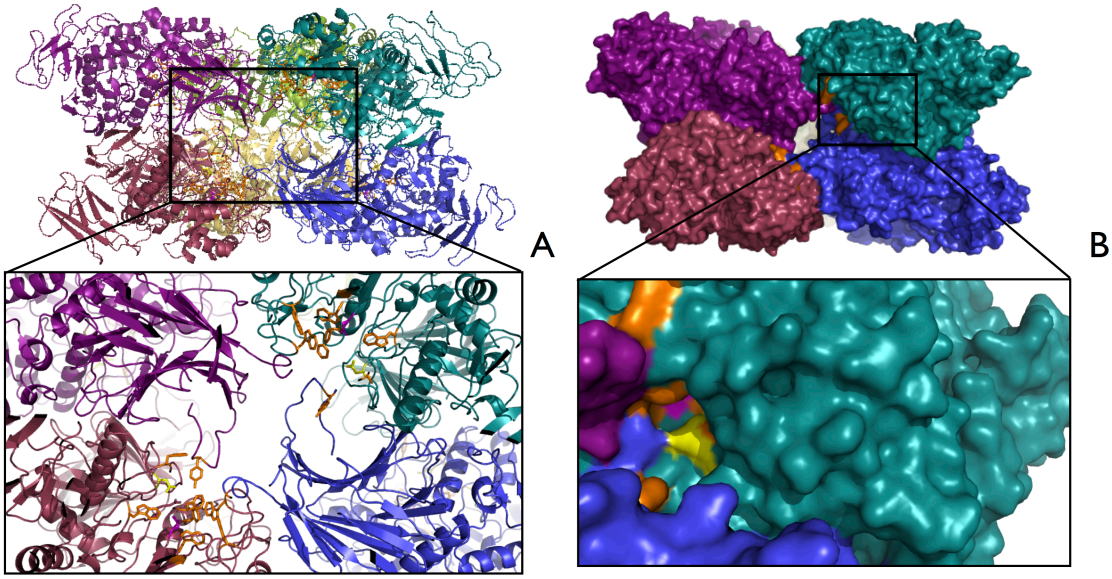

Figure 5-4: Top view of hexamer with close up of active site pocket. A: Ribbon representation showing two active site pockets (orange residues); **B:** Surface representation showing one active site pocket (**Legend:** Orange: Binding residues; Yellow: acid/base catalyst; Magenta: nucleophile; Deep Teal: Chain A; Blue: Chain D; Purple: Chain E).

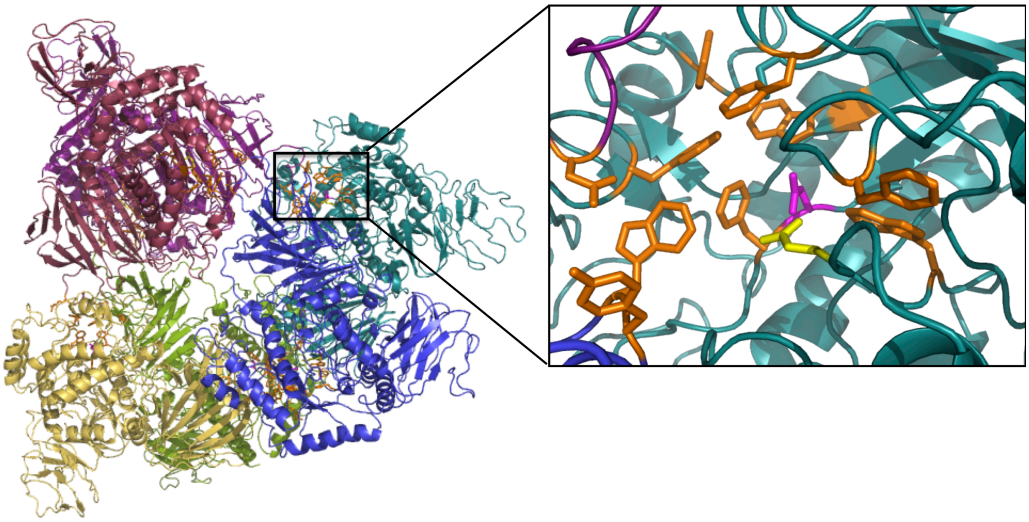

Figure 5-5: CfXyl31 hexamer and close up view of the active site. (**Legend:** Orange: Binding residues; Yellow: acid/base catalyst; Magenta: nucleophile; Deep Teal: Chain A; Blue: Chain D; Purple: Chain E)

5.5 Discussion

Currently, CfXyl31 has not been fully characterized enzymatically; however, from previous enzymatic work, there is detectable xylosidase activity (Personal Communication, Warren Wakarchuk). With respect to sequence similarity, CfXyl31 is 55% identical to YicI (PDB: 1SXI) (Lovering 2004) (Figure 5-6), a Family 31 glycoside hydrolase found in *Escherichia coli*. YicI is a xylosidase, hydrolyzing *p*NPαXyl and isoprimeverose (Xyl-α-1,6-Glc) (Lovering 2004; Okuyama et al. 2004). Structurally, CfXyl31 is highly similar to YicI. Both are hexamers and many of the active site residues, including the acid/base catalyst and nucleophile, are structurally conserved (Figure 5-7). Unlike mammalian Family 31 glycoside hydrolases, in which the catalytic signature sequence is WiDMNE, in bacterial Family 31 glycoside hydrolases, the catalytic signature sequence is KTDFGE (Lovering 2004). This suggests early separation of bacterial family 31 glycoside hydrolases and those of plants, fungi, and mammals during the course of evolution (Sampedro et al. 2001; Monroe et al. 2003).

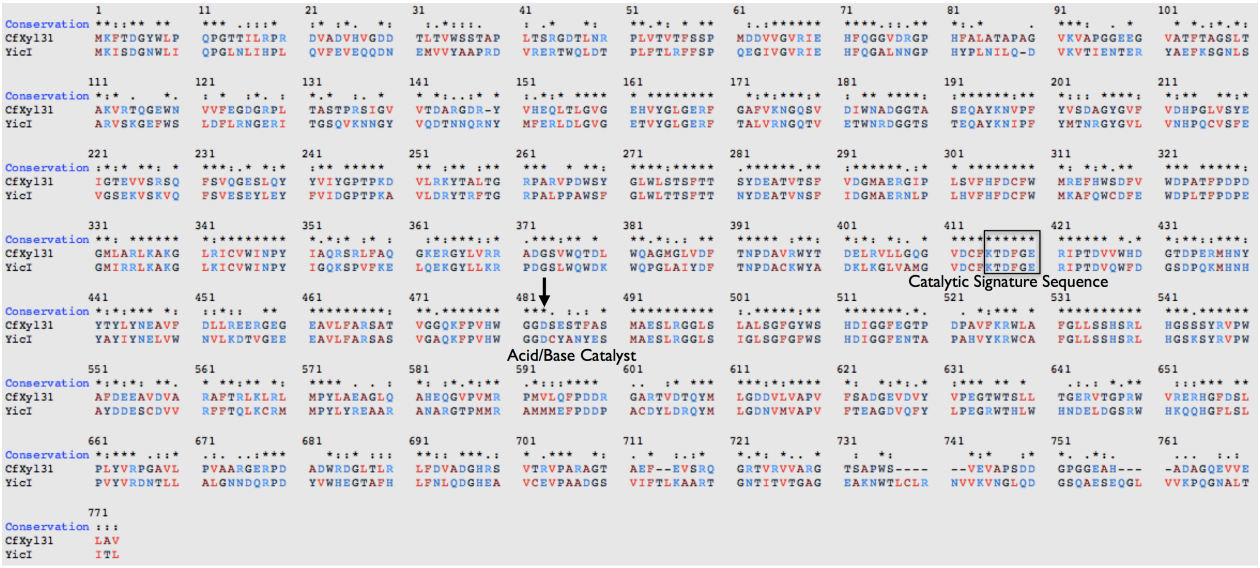

Figure 5-6: Alignment of CfXyl31 sequence and YicI (PDB: 1XSI) (Lovering 2004) .

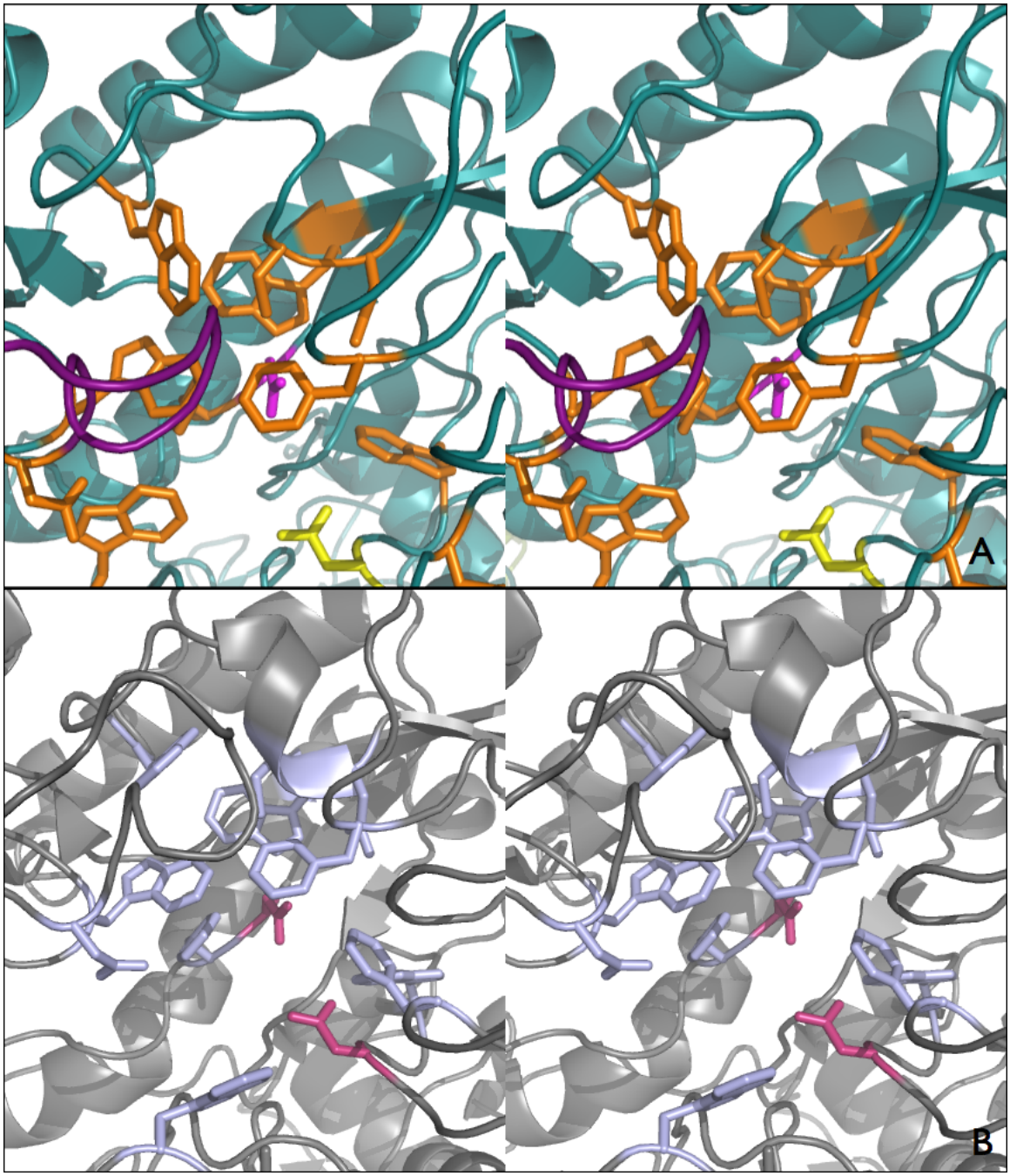

Figure 5-7: Stereo view comparing the CfXyl31 active site pocket (A: Orange: binding residues; Yellow: acid/base catalyst; Magenta: nucleophile**) and YicI (PDB: 1XSI) active site pocket (B:** Light blue: Binding residues; Pink: acid/base catalyst and nucleophile**).**

When comparing the CfXyl31 active site to that of YicI (PDB: 1XSI) (Lovering 2004), the differences seen in the active site are difficult to account for, due to the current quality of the structural model. For instance, in CfXyl31, the Phe515 side chain is orientated away from the active site pocket, whereas in YicI, this side chain extends into the -1 subsite. Although there is density to support the position of the phenylalanine side chain in the CfXyl31 crystal structure, it is important to keep in mind the model quality when making inferences of functionality. The orientation may be of importance with respect to substrate specificity, but with a low quality model and without full enzymatic characterization, these hypotheses are difficult to support.

Despite these challenges, there are observations that can be made and supported based on previous studies involving YicI and other α-glucosidases such as ntMGAM. Firstly, previous gel filtration analysis of YicI has indicated that the protein exists as a hexamer (Okuyama et al. 2004). Due to the structural similarity of CfXyl31 to YicI in addition to the fact that three monomers converge to create an active site pocket, CfXyl31 is thought to exist as a hexamer as well. Both enzymes exhibit α-xylosidase activity as well as a very small degree of hydrolytic activity on maltose (Okuyama et al. 2004) and demonstrate similar active site architecture (Figure 5-7). Further, similarity is seen in both the global structure and active site architecture of the CfXyl31 when compared to ntMGAM (Figure 5-8). When the CfXyl31 monomer is superposed with an ntMGAM monomer (PDB: 2QLY) (Sim et al. 2008), the overall structure is well conserved (RMSD value: 3.719 Å) (Figure 5-7, A). Further, the nucleophile and acid/base catalyst are structurally conserved as the catalytic residues in ntMGAM (the nucleophile: residue 443, and acid/base catalyst: residue 542) align with the proposed catalytic residues in CfXyl31 (residues 416 and 482 respectively) (Figure 5-8, B & C).

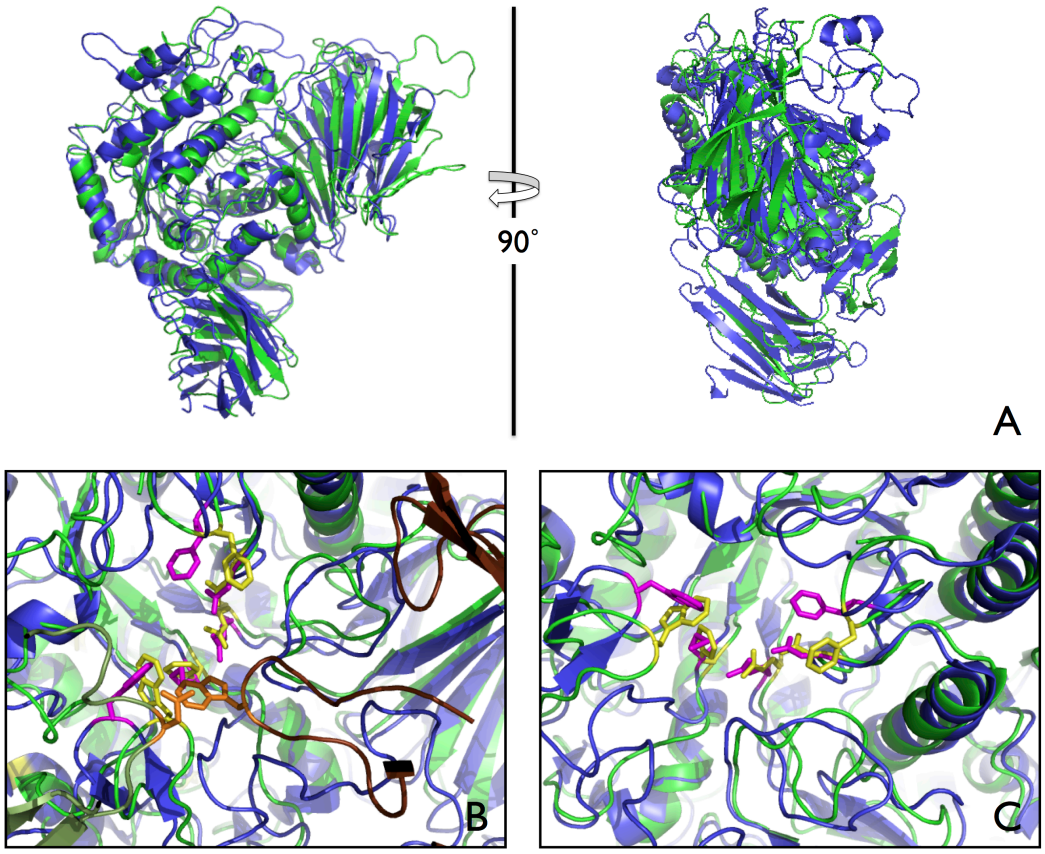

Figure 5-8: Comparison of CfXyl31 and ntMGAM globally (A) and a close up view of the active sites with CfXyl31 as a hexamer (B) and monomer (C). Legend: Green: CfXyl31; Blue: ntMGAM (PDB: 2QLY)**(Sim et al. 2008)**; Magenta: active site residues, ntMGAM; Yellow: active site residues, CfXyl31.

Interestingly, previous work has been done by Okuyama et al. 2006 in which mutations were made to YicI to convert the enzyme from an α-xylosidase to an α-glucosidase. With this precedent in mind, it is proposed that with a high degree of structural similarity, CfXyl31 is a good candidate for mutagenesis in an attempt to mimic the α-glucosidase activity of ntMGAM. A stereo view of both the ntMGAM active site and CfXyl31 active site with conserved residues can be seen in Figure 5-9. The first noticeable difference is the number of aromatic residues lining the CfXyl31 active site, as the majority of binding residues that line the active site in CfXyl31 are bulky, aromatic residues. To mimic the activity of ntMGAM and allow for binding of small oligosaccharides, some of these bulky aromatic residues may need to be replaced with those

that better mimic the active site pocket in ntMGAM. Thus, it is proposed that mutating aromatic residues at the entrance to the active site pocket such as Phe417 and Trp345.

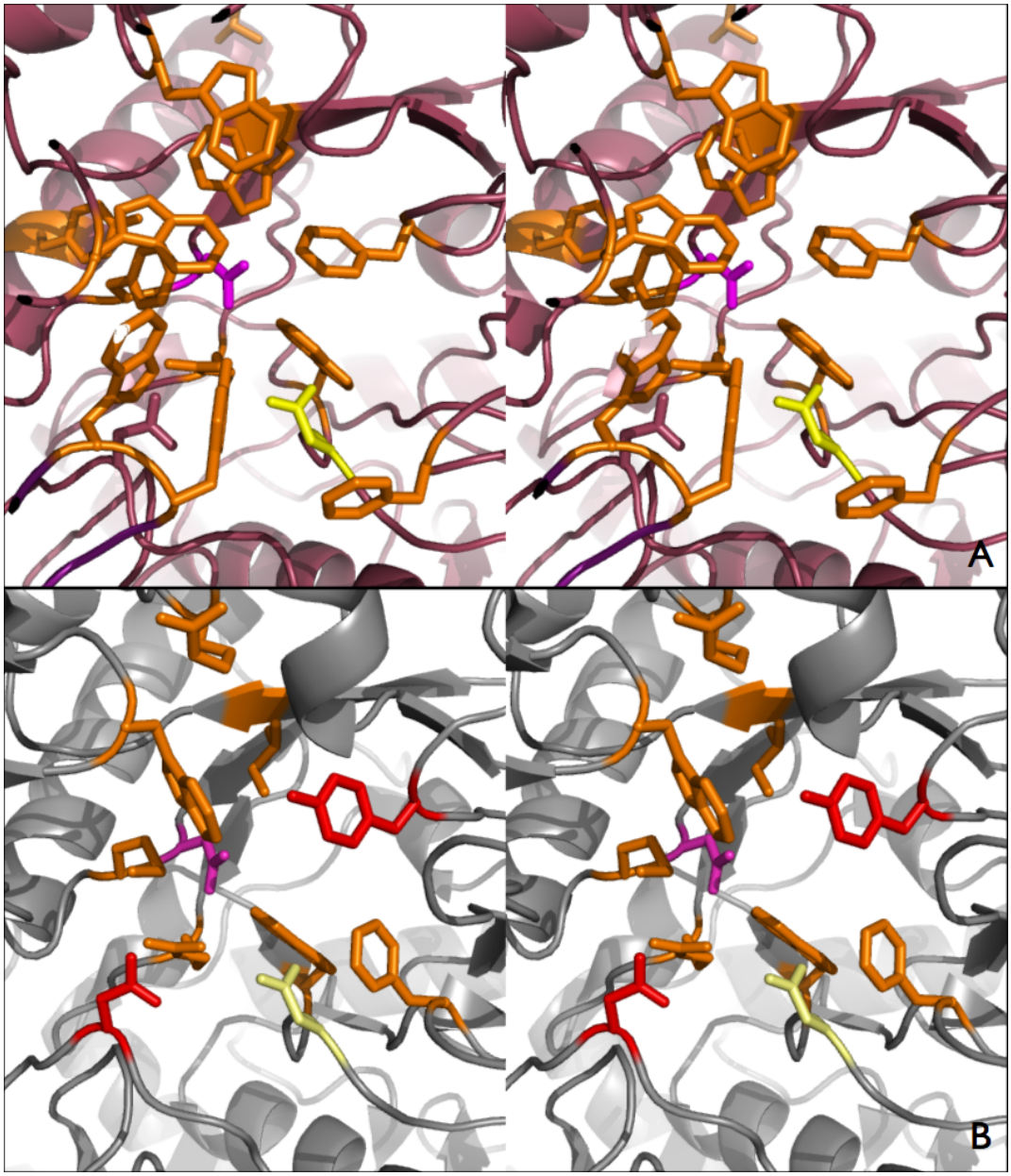

Figure 5-9: Stereo view comparing CfXyl31 (A) and ntMGAM (PDB: 2QLY, B) (Sim et al. 2008) active sites. Tyrosine 299 and Aspartic acid 185 are shown in red.

The hypothesis that the main difference between CfXyl31 and ntMGAM is the entrance to the active site pocket is further supported by a close look at the -1 and +1 subsites (Figure 5-9),

which are very similar. A comparison of the active site residues that have been shown to be involved in substrate binding in ntMGAM and CfXyl31 are summarized in Table 5-2.

Table 5-2: Comparison of residues involved in hydrolysis in CfXyl31 and ntMGAM (PDB: 2QLY) (Sim et al. 2008).

Active Site Residues implicated in substrate binding in ntMGAM only and corresponding residue in CfXyl31		Active Site residues implicated in substrate binding in CfXyl31 and ntMGAM		Active Site Residues implicated in substrate binding in CfXyl31 only and corresponding residue in ntMGAM	
ntMGAM	**CfXyl31**	**ntMGAM**	**CfXyl31**	**ntMGAM**	**CfXyl31**
Asp203	Asp185	Phe575	Phe515	Trp539	Trp479
Thr205	Thr188	His 600	His 540	Trp441	Lys414
		Tyr299	Phe277	Ile364	Trp345
		Asp327	Asp306	Met444	Phe417
		Trp406	Trp380	Ile328	Cys307
		Asp443	Asp416		
		Arg526	Arg466		
		Asp482	Asp542		

As can be seen in the summary, many of the proposed active site residues in CfXyl31 have identical corresponding residues in the ntMGAM active site. Thus if binding residues Trp345 and Phe417 are mutated to be isoleucine and methionine, respectively, this will mimic the active site structure of ntMGAM more closely. The difference between an α-xyloside and α-glucoside moiety is the presence of a hydroxymethyl group in the α-glucoside (C6) that is not present in the α-xyloside. In the mutagenesis work completed by Okuyama et al. 2006 in YicI, the phenylalanine (residue 308) was mutated to an aspartic acid and cysteine (residue 307) was mutated to an isoleucine to facilitate interaction with the C6 in the glucose moiety based on structural alignment. The resulting mutant enzyme lost α-xylosidase activity and gained α-

glucosidase activity (Okuyama et al. 2006). These two residues are conserved in CfXyl31, and interestingly, in ntMGAM, the residues equivalent to Cys307 and Phe308 are isoleucine (residue 328) aspartic acid (residue 329), identical to that of the YicI mutant. The crystal structure of ntMGAM had not been determined at the time this work was completed, but subsequent study of the ntMGAM structure in addition to the activity of the YicI C307I/F308D mutant supports the notion that these residues are involved in determining substrate specificity. Thus, it follows these two mutations will likely confer the same activity in CfXyl31.

The other key difference between the ntMGAM and CfXyl31 active sites is the presence of two loops from two other YicI monomers, in particular loop 1 (between β-strand 1 and α-helix 1), in the active site pocket. In the same mutagenesis study of YicI (Okuyama et al. 2006), loop 1 in was found to be three amino acids longer than equivalent loops in other α-glucosidases, and was hypothesized to prevent interaction between the C6' OH in the α-glucose moiety due to steric hindrance. Thus, this loop was shortened, and the resultant YicI mutant demonstrated both α-glucosidase and α-xylosidase activity. This loop is also present in CfXyl31, and is one amino acid longer than the loop seen in YicI (Figure 5-10). Further, the loop is farther inside the active site pocket. The residues on loop 1 nearest to the active site are bulky, aromatic residues (Tyr7 & Trp8) whereas in YicI, the comparable residues are Asn7 and Trp8. Due to the small degree of hydrolysis of maltose seen in both YicI and CfXyl31, this loop is likely flexible, but shortening it may help ensure the interaction between the active site and glucose C6' OH can occur efficiently. In the crystal structure of ntMGAM, an aspartic acid residue (residue 203) interacts with the C6' hydroxyl group and the side chain of residue 203 (aspartic acid). There is an equivalent residue in CfXyl31, Asp185, however due to the presence of loop 1 in the active site, this aspartic acid is not accessible to the substrate. By shortening this loop in CfXyl31, the interaction between aspartic acid (residue 185) and the substrate can occur, conferring α-glucosidase activity. Figure 5-11 illustrates the structurally conserved aspartic acid residue in CfXyl31 and ntMGAM.

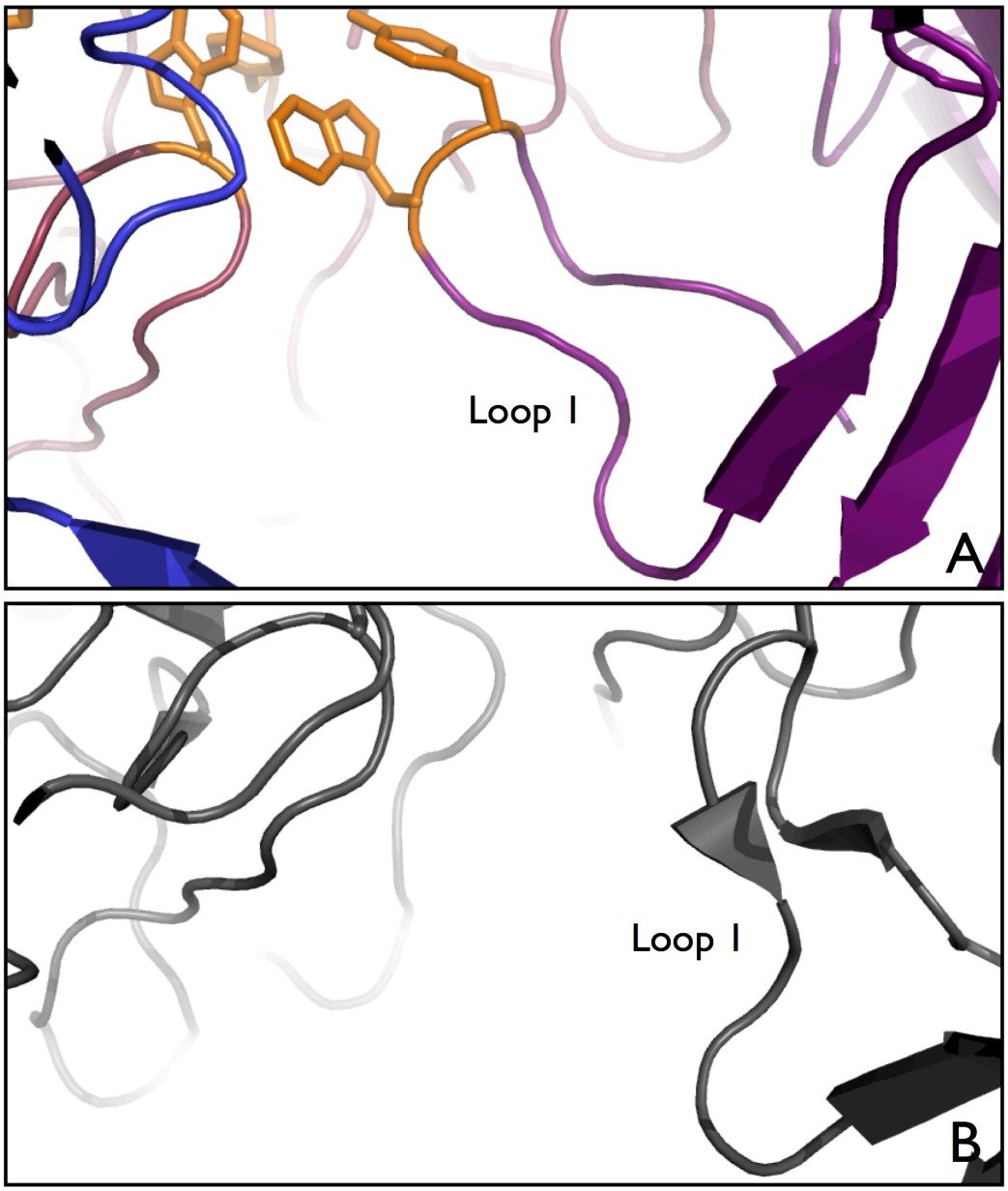

Figure 5-10: Comparision of Loop 1 in CfXyl31 (A) and YicI (B).

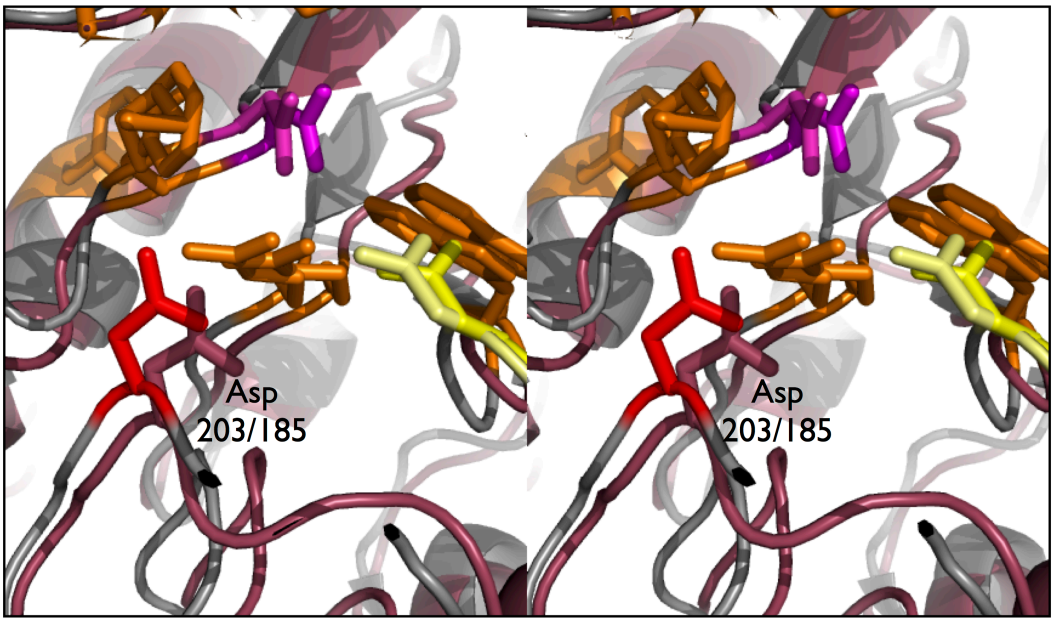

Figure 5-11: Structurally conserved aspartic acid residue in ntMGAM (backbone: gray; aspartic acid residue 203: red) and CfXyl31 (backbone and aspartic acid residue 185: deep red).

Ultimately, to mimic ntMGAM α-glucosidase activity, both shortening the loop and mutating residues 307 (from a cysteine to an isoleucine) and 308 (from a phenylalanine to an aspartic acid) may be the optimal approach. Solely shortening loop 1 likely will not abolish α-xylosidase activity, whereas additional C307I/F308D mutations are proposed to abolish α-xylosidase activity while conferring α-glucosidase activity. Thus, the combination of both modifications may accomplish the goal of mimicking the activity of ntMGAM. The hexameric structure of CfXyl31 and its similarity to ntMGAM prompted the concept that this protein could one day be engineered to act as a "starchosome", with each individual monomer in the hexameric structure mimicking the activity of ctSI, ctMGAM-N20, ctMGAM-N2, ntMGAM, and ntSI.

5.6 Conclusion

Presently, the CfXyl31 crystallographic model is not optimal. Due to poor resolution of approximately 3 Å, there are rotamer and Ramachandrian outliers that are unable to be rectified. However, the crystals are reproducible, thus personnel in the laboratory will be able to reproduce the crystals and collect data at a more powerful beamline, such as the Canadian

Light Source. Further, full enzymatic characterization must be completed to support hypotheses related to substrate specificity. Despite this, comparison to other related structures ntMGAM and YicI, has provided some insight into possible mutations that can be made to CfXyl31 to mimic ntMGAM α-glucosidase activity. This can be accomplished with two specific mutations: C307I and F308D, as well as shortening the length of loop 1 by 3-4 amino acid residues. Further, once more structural information with respect to the C-terminal enzymes and an improved crystallographic model of CfXyl31 are available, this knowledge can be extended to create a "starchosome": a CfXyl31 hexameric structure engineered such that each monomer has α-glucosidase activity that mimics each mammalian glycoside hydrolase catalytic subunit: ctSI, ctMGAM, ntSI, and ntMGAM. This starchosome can be used to study different starch structures as well as investigate the plausibility of use of a similar approach in enzyme replacement therapy for metabolic disorders such as CSID.

Chapter 6
Conclusions and Future Work

6.1 Concluding Remarks

The work in this thesis helps to unravel the questions surrounding the seemed duplicity of mammalian intestinal α-glucosidases as well as the differing substrate specificities of the four catalytic subunits as well as two alternatively spliced forms of ctMGAM. First, the inhibition of the four catalytic subunits is explored, the results of which highlight the structural and biochemical differences between the enzymatic subunits, despite overall similarity. Each individual catalytic subunit can be selectively inhibited, demonstrated with linear malto-oligosacharides and α-limit dextrins (Lee et al. 2012) in addition to maltose. This study also supports the notion that sulfonium ion-based inhibitors will be valuable tools in independently toggling the activities of each subunit *in vivo*, controlling the rate of starch digestion. Slowing the digestion of starch will minimize the initial spike in blood glucose levels following carbohydrate ingestion, which may be used to treat non-insulin dependent Type II diabetes.

Structural differences between known structures (ntMGAM, ntSI, and ctMGAM) and homology models (ctSI and ctMGAM-N20) are explored in Chapter 3, providing valuable insight with respect to both structure and substrate specificity. A small difference in the 21-amino acid extension present in the C-terminal catalytic subunits is seen in the ctSI homology model. A tryptophan that interacts with acarbose in the ctMGAM crystal structure (PDB: 3TOP) (Ren et al. 2011) is a glutamic acid in ctSI. This may explain the small degree of selectivity seen between ctMGAM and ctSI for longer substrates. Two residues in ctSI that likely play a role in the hydrolysis of α-1,2 linkages as well as disulfide bond formation have also been identified: Cysteine at position 401 and tyrosine 573. Finally, the ctSI active site is partially occluded by the 21-amino acid extension seen in C-terminal catalytic subunits. In comparison, the ntSI active site is open and highly accessible, a plausible explanation for the debranching activity of ntSI that is not seen in ctSI. The ctMGAM-N20 homology model highlights a narrow entrance to the active site in addition to partial occlusion by the 89-amino acid extension seen only in the N20 isoform. The active site of ctMGAM-N20 exhibits some similarity to active sites of enzymes with α-1,6 hydrolytic activity. This activity may be due to a tyrosine residue located between the +1

and +2 subsite that helps to coordinate substrates in the active site pocket. Lastly, selectivity seen in Chapter 2 between ctMGAM-N20 and ctMGAM-N2 with small inhibitors such as de-*O*-sulfonated ponkoranol may be explained by the presence of the 89-amino acid extension seen in the N20 isoform but absent from the N2 isoform.

In Chapter 4, mutations seen in ntSI in CSID patients and the potential functional and structural implications are presented. Four mutations were studied that were found in four different individuals diagnosed with CSID. Residues involved in trafficking and protein-protein interaction have been identified. Further, the prevalence of CSID in the general population needs to be investigated further, however exploration of the SI gene in chromosome 3 data from 1000 Genomes (Consortium et al. 2012) highlights two tribal populations of African descent with a high number of SNPs in the SI gene. Further genetic studies of populations with high incidence of CSID such as Nordic and Aboriginal populations will help to examine the role of diet in the evolution of starch digesting enzymes.

Finally, the 3 Å structure of CfXyl31, a Family 31 glycoside hydrolase from *C. fimi* was presented. Although not of optimal quality, comparison of this structure to other Family 31 glycoside hydrolases such as ntMGAM and YicI lead to possible mutations that can be made to CfXyl31 to mimic ntMGAM α-glucosidase activity. Two specific mutations can likely accomplish this in addition to shortening the length of loop 1 by 3-4 residues. In addition, the concept of the "starchozyme" is introduced: an engineered hexameric structure of CfXyl31 in which each CfXyl31 monomer can independently mimic each mammalian glycoside hydrolase catalytic subunit: ctSI, ctMGAM, ntSI, and ntMGAM.

Despite these advances, there is more work to be done to further understand duplicity of these enzymes in the mammalian small intestine as well as the intricacies of the active sites of each catalytic subunit. Glucogenesis is an important process that needs to be better understood with respect to how different types of starch are processed by mammalian α-glucosidases. Carbohydrates are a substantial component of the human diet and postprandial blood glucose control significantly impacts human health. Thus, it is important to understand how our own digestive enzymes process the carbohydrates we eat and the consequences associated with the food choices we make.

6.2 Future Studies

6.2.1 N-terminal Structural Studies

In Chapter 2, inhibition studies were conducted in which potent inhibitors of both N-terminal enzymes were identified. With efficient expression protocols in place, an increased amount of protein is now available to use in crystallographic studies. Co-crystallization of both ntSI and ntMGAM with potent inhibitors will increase our understanding of the active site requirements of these catalytic subunits by examining the interaction of the inhibitor with the active site. Inhibitors of a high degree of potency against both ntMGAM and ntSI such as 3'- and 5'-*O*-β-maltosyl de-*O*-sulfonated ponkoranol would be excellent candidates as well as de-*O*-sulfonated ponkoranol and its selenium analogues.

In addition to inhibitor studies, structural studies of ntSI mutants representative of those seen in CSID patients discussed in Chapter 4 would help support or refute the current hypotheses of the implications of these mutations on protein transport and function. In particular, the predicted glycosylation site thought to impact trafficking of ntSI to the small intestinal lumen, ASN 119, would be an interesting target for mutagenesis studies.

6.2.2 C-terminal Catalytic Subunits

6.2.2.1 ctSI

A small number of crystal forms of ctSI have been produced, but each investigated to date does not diffract well enough for data collection. Limited proteolysis was found to help crystal formation, leading to the hypothesis that there may be a flexible portion of the protein that does not promote crystallization. Upon investigation of the sequence, a portion of the N-terminal construct is seen at the beginning of the ctSI sequence in the expression construct. At the time these constructs were initially created, a cutoff point between the N- and C-terminal subunits was not as clear as it is presently. Thus, re-examining the C-terminal clones, in particular ctSI, to determine a proper cutoff point between the N- and C-terminal subunits is highly recommended. Eliminating the flexible region of protein that is thought to be a component of the N-terminal subunit, will likely lead to increased packing efficiency and hopefully, the discovery of more crystal forms of ctSI. A new crystal form of ctSI has also recently been discovered and is currently being optimized.

If no structure is determined with the newfound crystal form, mutagenesis experiments would be beneficial in testing the hypotheses in Chapter 3 based on the ctSI homology model. Two residues found only in ctSI were hypothesized to play an important role in substrate specificity: Cys401 and Tyr573. These residues are hypothesized to be responsible for the ability of ctSI to hydrolyze α-1,2 linkages. Mutating these two residues and characterizing the activity of the mutant enzyme will help support or refute this hypothesis. Finally, to facilitate mutagenesis experiments, it would be beneficial to attempt expression in a bacterial expression system, making mutagenesis less expensive, fast, and convenient. If this is not possible, mutagenesis can be done in the baculovirus system.

6.2.2.2 ctMGAM-N2

Although ctMGAM-N2 has been structurally characterized by Ren et al. 2011, crystallization studies with inhibitors co-crystallized or soaked into the crystal would further explore the active site requirements of this isoform. The duplicity of the mammalian intestinal α-glucosidases as well as their alternate splice forms is not fully understood. Thus, studying inhibitors that differentiate between N- and C-terminal subunits (such as acarbose) as well as differentiate between ctMGAM-N2 and ctMGAM-N20 (such as de-*O*-sulfonated ponkoranol) would provide insight into this research question.

6.2.2.3 ctMGAM-N20

Although the crystallographic structural model is not currently available, as seen in Chapter 3, there are interesting structural characteristics that differentiate the two isoforms ctMGAM-N2 and ctMGAM-N20. In the proposed homology model of ctMGAM-N20, the 89-amino acid extension is near the opening of the active site and potentially plays a role in substrate specificity. Structural characterization of ctMGAM-N20 would help determine the key residues that allow for hydrolysis of both α-1,4 and α-1,6-linked substrates. Additionally, further enzymatic characterization needs to be completed as hydrolytic activity on substrates with α-1,6 linkages has only recently been discovered. Full enzymatic characterization of ctMGAM-N20 with longer substrates such as isomaltotriose and α-limit dextrins would increase our understanding of the role this isoform plays in starch digestion in the small intestinal lumen. In addition, docking studies of various substrates such as isomaltose would be helpful in analyzing the active site.

6.2.3 Interaction

In addition to studying the individual catalytic subunits, it would be interesting to study the interaction between all four catalytic subunits: ntSI, ctSI, ntMGAM, and ctMGAM. Although the N- and C-terminal subunits are thought to associate with one another in the small intestinal lumen, it would be interesting to explore these interactions further. It is hypothesized there may also be interaction between C-terminal subunits, thus it is suggested that protein interaction studies take place testing all possible protein combinations and characterizing these interactions.

6.2.4 Population Studies

The study attempting to relate SNPs in the SI gene to ancestry has pinpointed two tribes in Africa with a high occurrence of SNPs in the SI gene. The diet of these two tribes consists of a corn meal preparation, plantains, as well as starchy tubers and grains native to that region. It is proposed that the complexity of the starchy tubers and grains be explored in more detail to better understand the potential impact of diet on CSID. Further, it is presently unknown if the incidence of CSID is higher in this population compared to European and North American populations, and this could be explored through data mining or a genetic population study. Finally, more databases, especially those with individuals of Nordic descent, (such as NordiCDB) need to be explored and compared to those of other ancestral origins to make a link between diet and incidence of CSID. Perhaps even looking at the SI gene over evolutionary time based on recent genome sequence results of ancient human genomes, such as the 4000 year old human from Greenland (Shapiro and Hofreiter 2010) would be of interest.

6.2.5 CfXyl31 Mutagenesis

The current structural model of CfXyl31 is not optimal; however, the crystals are reproducible. Personnel in the laboratory will be able to reproduce the crystals and collect higher resolution data at a more powerful beamline, such as the Canadian Light Source. Secondly, the mutagenesis studies to mimic ntMGAM α-glucosidase activity can be completed by mutating cysteine 307 to isoleucine and phenylalanine 308 to aspartic acid. Loop 1 can also be shortened by removing 3-4 residues. This protein is expressed in a bacterial expression system, making mutagenesis relatively simple and inexpensive. Further, this enzyme must be more fully characterized from an enzymatic standpoint, in particular isoprimeverose (Xyl-α-1,6-Glc) is a

substrate that may be hydrolyzed by CfXyl31 based on analysis of YicI (Lovering 2004; Okuyama et al. 2004; Okuyama et al. 2006).

6.2.6 Application to Disease

In addition to tools to investigate the role of each enzymatic subunit in starch digestion, inhibitors have potential use as an antidiabetic treatment. By slowing the release of glucose in the small intestine, the rapid release and absorption of glucose can be moderated, resulting in a more gradual spike in postprandial blood glucose levels. Current α-glucosidase inhibitors have a wide range of specificity, and a more targeted approach would help alleviate side effects such as gastrointestinal upset. Recently, ctMGAM has been identified as a target for this type of individual inhibition, as it was found to be involved in the early stages of starch digestion, in conjunction with α-amylase (Lee et al. 2014). The physiological effects of poorly controlled blood glucose levels impact human health significantly, leading to ailments such as poor cardiovascular health. By further understanding the role of each enzymatic subunit, targeted inhibitors can be developed to modulate the release of glucose in the small intestine, taking into account the physiological needs of different metabolic diseases such as diabetes and obesity. Further, the specific function of each catalytic subunit can be more fully understood by toggling the enzymatic activity on and off in animal models. The full understanding of the individual roles of the enzymes will lead to the development of novel, resistant starch structures engineered for a gradual release of glucose during starch digestion.

Akin to lactose intolerance, enzyme replacement and diet modification are the current treatments for CSID. Deciphering the unique role of each catalytic subunit will allow for modifications of the enzymes currently used in replacement therapy. Thus, specific enzymatic activities absent in CSID patients can be corrected for individually in a targeted fashion, personalizing the treatment based on specific symptoms.

A complete understanding of how MGAM and SI function in the small intestine will give insight into the response of our bodies and digestive enzymes upon the ingestion of carbohydrates. The future research in this area is an exciting opportunity to understand the interplay between diet, physiology, and nutrition.

Bibliography

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D-Biological Crystallography (2010). D66, 213-221:1–9.

Alfalah M, Jacob R, Preuss U, Zimmer K-P, Naim HY. 1999. O-linked glycans mediate apical sorting of human intestinal sucrase-isomaltase through association with lipid rafts. Current Biology 9:593–596.

Alfalah M, Keiser M, Leeb T, Zimmer K-P, Naim HY. 2009. Compound heterozygous mutations affect protein folding and function in patients with congenital sucrase-isomaltase deficiency. Gastroenterology 136:883–892.

Altschul SF, Madde TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25:3389–3402.

Balfour JA, McTavish D. 1993. Acarbose: An update of its pharmacology and therapeutic use in diabetes mellitus. Drugs 46:1025–1054.

Barker MK, Rose, DR. 2013. Specificity of Processing α-Glucosidase I Is Guided by the Substrate Conformation: Crystallographic and *in silico* studies. The Journal of Biological Chemistry 288:13563–13574.

Bell RR, Draper HH, Bergan JG. 1973. Sucrose, lactose, and glucose: Eskimos. The American Journal of Clinical Nutrition 26:1185–1190.

Bhattacharya D, Cheng J. 2012. 3Drefine: Consistent protein structure refinement by optimizing hydrogen bonding network and atomic-level energy minimization. Proteins 81:119–131.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Research 37:D233–D238.

Chandrasena JPC. 1935. The Chemistry and Pharmacology of Ceylon and Indian Medicinal Plants.

Chiasson J, Josse R, Gomis R, Hanefeld M, Karasik A, Laakso M. 2002. Acarbose for prevention of type 2 diabetes mellitus: the STOP-NIDDM randomised trial. The Lancet 359:2072–2077.

Core R Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing: URL http:––www.R–project.org–.

The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature 467:1061–1073.

Dahlqvist A. 1964. Method for Assay of Intestinal Disaccharidases. Analytical Biochemistry 7:18–25.

David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, et al. 2013. Diet rapidly and reproducibly alters the human gut microbiome. Nature:1–18.

Eccleston JL, Koh C, Markello TC, Gahl WA, Heller T. 2012. An apparent homologous deletion in maltase-glucoamylase, a lesson in the evolution of SNP arrays. Molecular Genetics and Metabolism 107:674–678.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792–1797.

Ellestad-Sayed JJ, Haworth JC, Hildes JA. 1978. Disaccharide malabsorption patterns in two Canadian communities14. International Nutrition 31:1473–1478.

Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. Acta Crystallographica Section D-Biological Crystallography 60:2126–2132.

Englyst H, Kingman S. 1992. Classification and measurement of nutritionally important starch fractions. European Journal of Clinical Nutrition 46:S33-50.

Ernst H, Leggio Lo L, Willemoës M, Leonard G, Blum P, Larsen S. 2006. Structure of the Sulfolobus solfataricus α-glucosidase: implications for domain conservation and substrate recognition in GH31. Journal of Molecular Biology 358:1106–1124.

Eskandari R, Jayakanthan K, Kuntz DA, Rose DR, Pinto BM. 2010a. Bioorganic & Medicinal Chemistry. Bioorganic & Medicinal Chemistry 18:2829–2835.

Eskandari R, Jones K, Ravinder Reddy K, Jayakanthan K, Chaudet M, Rose DR, Pinto BM. 2011a. Probing the Intestinal α-Glucosidase Enzyme Specificities of Starch-Digesting Maltase-Glucoamylase and Sucrase-Isomaltase: Synthesis and Inhibitory Properties of 3′- and 5′-Maltose-Extended De-O-sulfonated Ponkoranol. Chemistry - A European Journal 17:14817–14825.

Eskandari R, Jones K, Rose DR, Pinto BM. 2010b. Probing the active-site requirements of human intestinal N-terminal maltase glucoamylase: the effect of replacing the sulfate moiety by a methyl ether in ponkoranol, a naturally occurring α-glucosidase inhibitor. Bioorganic & Medicinal Chemistry Letters 20:5686–5689.

Eskandari R, Jones K, Rose DR, Pinto BM. 2011b. The effect of heteroatom substitution of sulfur for selenium in glucosidase inhibitors on intestinal α-glucosidase activities. Chemical Communications 47:9134-9136.

Eskandari R, Jones K, Rose DR, Pinto BM. 2011c. Selectivity of 3â€²-O-methylponkoranol for inhibition of N- and C-terminal maltase glucoamylase and sucrase isomaltase, potential therapeutics for digestive disorders or their sequelae. Bioorganic & Medicinal Chemistry Letters 21:6491–6494.

Eskandari R, Kuntz DA, Rose DR, Pinto BM. 2010c. Potent Glucosidase Inhibitors: De- O-sulfonated Ponkoranol and Its Stereoisomer. Organic Letters 12:1632–1635.

Fass D. 2012. Disulfide Bonding in Protein Biophysics. Annual Review of Biophysics 41:63–79.

Freiburghaus AU, Dubs R, Hadorn B, Gaze H, Hauri HP, Gitzelmann R. 1977. The brush border membrane in hereditary sucrase-isomaltase deficiency: abnormal protein pattern and presence of immunoreactive enzyme. European Journal of Clinical Investigation 7:455–459.

Gething M-J, Sambrook J. 1992. Protein folding in the cell. Nature 355:33–45.

Ghavami A, Johnston B, Pinto B. 2001. A New Class of Glycosidase Inhibitor: Synthesis of Salacinol and Its Stereoisomers†. Journal of Organic Chemistry 66:2312-3217.

Goda T, Takase S. 1994. Dietary Carbohydrate and Fat Independently Modulate Disaccharidase Activities in Rat Jejunum. 124:2233–2239.

Goda T, Yasutake H, Takase S. 1994. Dietary fat regulates cellular retinol-binding protein II gene expression in rat jejunum. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1200:34–40.

Godbout A, Chiasson J. 2007. Who should benefit from the use of alpha-glucosidase inhibitors? Current Diabetes Reports 7:333–339.

Hauri H, Sterchi E, Bienz D, Fransen J. 1985. Expression and intracellular transport of microvillus membrane hydrolases in human intestinal epithelial cells. Journal of Cell Biology 101:838-851.

Hauri H-P, Quaroni A, Isselbacher KJ. 1979. Biogenesis of intestinal plasma membrane: posttranslational route and cleavage of sucrase-isomaltase. Proceedings of the National Academy of Sciences 76:5183–5186.

Heijmans B, Tobi E, Stein A, Putter H, Blauw G, Susser E, Slagboom P, Lumey L. 2008. Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proceedings of the National Academy of Sciences 105:1–4.

Heymann H, Breitmeier D, Günther S. 1995. Human small intestinal sucrase-isomaltase: different binding patterns for malto-and isomaltooligosaccharides. Biological Chemistry Hoppe-Seyler 376:249.

Heymann H, Günther S. 1994. Calculation of Subsite Affinities of Human Small Intestinal Glucoamylase-Maltase. Biological Chemistry Hoppe-Seyler 375:451–456.

Holman R, Cull C, Turner R. 1999. A randomized double-blind trial of acarbose in type 2 diabetes shows improved glycemic control over 3 years (UK Prospective Diabetes Study 44). Diabetes Care 22:960.

Honma K, Mochizuki K, Goda T. 2007. Carbohydrate/fat ratio in the diet alters histone acetylation on the sucrase–isomaltase gene and its expression in mouse small intestine. Biochemical and Biophysical Research Communications 357:1124–1129.

Jacob G. 1995. Glycosylation inhibitors in biology and medicine. Current Opinion in Structural Biology 5:605–611.

Jacob R, Zimmer K, Schmitz J, Naim HY. 2000. Congenital sucrase-isomaltase deficiency arising from cleavage and secretion of a mutant form of the enzyme. Journal of Clinical Investigation 102:281–287.

Jayakanthan K, Mohan S, Pinto BM. 2009. Structure Proof and Synthesis of Kotalanol and De- O-sulfonated Kotalanol, Glycosidase Inhibitors Isolated from an Herbal Remedy for the Treatment of Type-2 Diabetes. Journal of the American Chemical Society 131:5621–5626.

Jayawardena MHS, de Alwis NMW, Hettigoda V, Fernando DJS. 2005. A double blind randomised placebo controlled cross over study of a herbal preparation containing Salacia reticulata in the treatment of type 2 diabetes. Journal of Ethnopharmacology 97:215–218.

Jayaweera DMA. 1981. Medicinal Plants Used in Ceylon. National Science Council of Sri Lanka, Colombo.

Johnston BD, Jensen HH, Pinto BM. 2006. Synthesis of Sulfonium Sulfate Analogues of Disaccharides and Their Conversion to Chain-Extended Homologues of Salacinol: New Glycosidase Inhibitors. Journal of Organic Chemistry 71:1111–1118.

Johnston PS, Coniff RF, Hoogwerf BJ, Santiago JV, Pi-Sunyer FX, Krol A. 1994. Effects of the carbohydrase inhibitor miglitol in sulfonylurea-treated NIDDM patients. Diabetes Care 17:20.

Jones K, Eskandari R, Naim HY, Pinto BM, Rose DR. 2012. Investigations of the Structures and Inhibitory Properties of Intestinal Maltase Glucoamylase and Sucrase Isomaltase. Journal of Pediatric Gastroenterology and Nutrition 55:S20–S24.

Jones K, Sim L, Mohan S, Kumarasamy J, Liu H, Avery S, Naim HY, Quezada-Calvillo R, Nichols BL, Mario Pinto B, et al. 2011. Mapping the intestinal alpha-glucogenic enzyme specificities of starch digesting maltase-glucoamylase and sucrase-isomaltase. Bioorganic & Medicinal Chemistry 19:3929–3934.

Keiser M, Alfalah M, Propsting MJ, Casrelletti D, Naim HY. 2006. Altered Folding, Turnover, and Polarized Sorting Act in Concert to Define a Novel Pathomechanism of Congenital Sucrase-Isomaltase Deficiency. The Journal of Biological Chemistry 281:14393–14399.

Kornfeld R, Kornfeld S. 2002. Assembly of Asparagine-linked Oligosaccharides. Annual Review of Biochemistry 54:631–634.

Krasilnikoff PA, Gudmand-Hoyer E, Moltke HH. 1975. Diagnostic Value of Disaccharide Tolerance Tests in Children. Acta Paediatrica Scandinavica 64:693–698.

Lammens W, Le Roy K, Schroeven L, Van Laere A, Rabijns A, Van den Ende W. 2009. Structural insights into glycoside hydrolase family 32 and 68 enzymes: functional implications. Journal of Experimental Botany 60:727–740.

Lee B-H, Lin AH-M, Nichols BL, Jones K, Rose DR, Quezada-Calvillo R, Hamaker BR. 2014. Mucosal C-terminal maltase-glucoamylase hydrolyzes large size starch digestion products that may contribute to rapid postprandial glucose generation. Molecular Nutrition & Food Research 1:1-11.

Lee B-H, Eskandari R, Jones K, Reddy KR, Quezada-Calvillo R, Nichols BL, Rose DR, Hamaker BR, Pinto BM. 2012. Modulation of Starch Digestion for Slow Glucose Release through "Toggling" of Activities of Mucosal α-Glucosidases. Journal of Biological Chemistry 287:31929–31938.

Lee B-H, Yan L, Phillips RJ, Reuhs BL, Jones K, Rose D, Nichols BL, Quezada-Calvillo R, Yoo S, Hamaker BR. 2013. Enzyme-synthesized highly branched maltodextrins have slow glucose generation at the mucosal a-glucosidase level and are slowly digestible in vivo. PLoS ONE 8:e59745.

Lillycrop KA, Slater-Jefferies JL, Hanson MA, Godfrey KM, Jackson AA, Burdge GC. 2007. Induction of altered epigenetic regulation of the hepatic glucocorticoid receptor in the offspring of rats fed a protein-restricted diet during pregnancy suggests that reduced DNA

methyltransferase-1 expression is involved in impaired DNA methylation and changes in histone modifications. British Journal of Nutrition 97:1064-1073.

Liu H, Pinto BM. 2005. Efficient Synthesis of the Glucosidase Inhibitor Blintol, the Selenium Analogue of the Naturally Occurring Glycosidase Inhibitor Salacinol. Journal of Organic Chemistry 70:753–755.

Lovering AL. 2004. Mechanistic and Structural Analysis of a Family 31 -Glycosidase and Its Glycosyl-enzyme Intermediate. Journal of Biological Chemistry 280:2105–2115.

Lucke T, Keiser M, Illsinger S, Lentze MJ, Naim HY, Das AM. 2009. Congenital and Putatively Aquired Forms of Sucrase-isomaltase Deficiency in Infancy: Effects of Sacrosidase Therapy. Journal of Pediatric Gastroenterology and Nutrition 49:485–487.

Luft JR, Collins RJ, Fehrman NA, Lauricella AM, Veatch CK, DeTitta GT. 2003. A deliberate approach to screening for initial crystallization conditions of biological macromolecules. Journal of Structural Biology 142:170–179.

Marsden RL, McGuffin LJ, Jones DT. 2002. Rapid protein domain assignment from amino acid sequence using predicted secondary structure. Protein Science 11:2814–2824.

Matsuda H, Murakami T, Yamahara J, Yoshikawa M. 1998. Protective effects of oleanolic acid oligoglycosides on ethanol or indomethacin-induced gastric mucosal lesions in rats. Pharmacology Letters 63:245–250.

Minami Y, Kuriyama C, Ikeda K, Kato A, Takebayashi K, Adachi I, Fleet GWJ, Kettawan A, Okamoto T, Asano N. 2008. Effect of five-membered sugar mimics on mammalian glycogen-degrading enzymes and various glucosidases. Bioorganic & Medicinal Chemistry 16:2734–2740.

Mochizuki K, Hanai E, Suruga K, Kuranuki S, Goda T. 2010a. Changes in α-glucosidase activities along the jejunal-ileal axis of normal rats by the α-glucosidase inhibitor miglitol. Metabolism 59:1442–1447.

Mochizuki K, Honma K, Shimada M, Goda T. 2010b. The regulation of jejunal induction of the maltase-glucoamylase gene by a high-starch/low-fat diet in mice. 54:1445–1451.

Mochizuki K, Igawa-Tada M, Takase S, Goda T. 2010c. Feeding rats a high fat/carbohydrate ratio diet reduces jejunal S/I activity ratio and unsialylated galactose on glycosylated chain of S–I complex. Life Sciences 86:524–531.

Mohan S, Pinto BM. 2007. Zwitterionic glycosidase inhibitors: salacinol and related analogues. Carbohydrate Research 342:1551–1580.

Mohan S, Pinto BM. 2010. Towards the elusive structure of kotalanol, a naturally occurring glucosidase inhibitor. Natural Products Report 27:481-488.

Mohan S, Sim L, Rose DR, Pinto BM. 2010. Probing the active-site requirements of human intestinal maltase glucoamylase: Synthesis and enzyme inhibitory activities of a six-membered ring nitrogen analogue of kotolanol and its de-O-sulfonated derivative. Bioorganic & Medicinal Chemistry 18:7794–7798.

Monroe JD, Garcia-Cazarin ML, Poliquin KA, Aivano SK. 2003. Antisense Arabidopsis plants indicate that an apoplastic α-xylosidase and α-glucosidase are encoded by the same gene. Plant Physiology and Biochemistry 41:877–885.

Muraoka O, Xie W, Tanabe G, Amer MFA, Minematsu T, Yoshikawa M. 2008. Tetrahedron Letters. Tetrahedron Letters 49:7315–7317.

Naim H, Sterchi E, Lentze M. 1988. Structure, biosynthesis, and glycosylation of human small intestinal maltase-glucoamylase. Journal of Biological Chemistry 263:19709-19717.

Naim HY, Heine M, Zimmer K-P. 2012. Congenital Sucrase-Isomaltase Deficiency: Heterogeneity of Inheritance, Trafficking, and Function of an Intestinal Enzyme Complex. Journal of Pediatric Gastroenterology and Nutrition 55:S13–S19.

Nasi R, Sim L, Rose D, Pinto B. 2007a. New chain-extended analogues of salacinol and blintol and their inhibitory activities. Mapping the active-site requirements of human maltase glucoamylase. Journal of Organic Chemistry 72:180-186.

Nichols BL, Avery S, Sen P, Swallow D. 2003. The maltase-glucoamylase gene: common ancestry to sucrase-isomaltase with complementary starch digestion activities. Proceedings of the National Academy of Sciences 100:1432-1437.

Nichols BL, Eldering J, Avery S, Hahn D. 1998. Human small intestinal maltase-glucoamylase cDNA Cloning. Journal of Biological Chemistry 273:3076–3081.

Nichols BL, Auricchio S. 2012. 50 Years of Progress Since Congenital Sucrase-Isomaltase Deficiency Recognition. Journal of Pediatric Gastroenterology and Nutrition 55:S2–S6.

Nichols BL, Nichols VN, Putman M, Avery SE, Fraley JK, Quaroni A, Shiner M, Sterchi EE, Carrazza FR. 2000. Contribution of Villous Atrophy to Reduced Intestinal Maltase in Infants with Malnutrition. Journal of Pediatric Gastroenterology and Nutrition 30:494–502.

Oe H, Ozaki S. 2008. Hypoglycemic Effect of 13-Membered Ring Thiocyclitol, a Novel α-Glucosidase Inhibitor from Kothala-himbutu (Salacia reticulata). Bioscience, Biotechnology, and Biochemistry 72:1962–1964.

Okuyama M, Kaneko A, Mori H, Chiba S, Kimura A. 2006. Structural elements to convert Escherichia coli α-xylosidase (YicI) into α-glucosidase. FEBS Letters 580:2707–2711.

Okuyama M, Mori H, Chiba S, Kimura A. 2004. Overexpression and characterization of two unknown proteins, YicI and YihQ, originated from Escherichia coli. Protein Expression and Purification 37:170–179.

Ozaki S, Oe H, Kitamura S. 2008. Alpha-glucosidase inhibitor from Kothala-himbutu (Salacia reticulata WIGHT). Journal of Natural Products 71:981–984.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods 8:785–786.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera - a visualisation system for exploratory research and analysis. Journal of Computational Chemistry 25:1605–1612.

Pflugrath JW. 1999. The finer things in X-ray diffraction data collection. Acta Crystallographica Section D-Biological Crystallography D55:1718-1725.

PyMol. The PYMOL Molecular Graphics System (DeLano Scientific) (2000).

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. InterProScan: protein domains identifier. Nucleic Acids Research 33:W116–W120.

Quezada-Calvillo R, Robayo-Torres C. 2007. Contribution of mucosal maltase-glucoamylase activities to mouse small intestinal starch alpha-glucogenesis. Journal of Nutrition 137:1725-1733.

Quezada-Calvillo R, Sim L, Ao Z, Hamaker BR, Quaroni A, Brayer GD, Sterchi EE, Robayo-Torres CC, Rose DR, Nichols BL. 2008. Luminal starch substrate “brake” on maltase-glucoamylase activity is located within the glucoamylase subunit. Journal of Nutrition 138:685–692.

Ren L, Qin X, Cao X, Wang L, Bai F, Bai G, Shen Y. 2011. Structural insight into substrate specificity of human intestinal maltase-glucoamylase. Protein Cell 2:827–836.

Ritz V, Alfalah M, Zimmer K-P, Schmitz J, Jacob R, Naim HY. 2003. Congenital sucrase-isomaltase deficiency because of an accumulation of the mutant enzyme in the endoplasmic reticulum. Gastroenterology 125:1678–1685.

Robayo Torres C, Quezada Calvillo R, Nichols B. 2006. Disaccharide digestion: clinical and molecular aspects. Clinical Gastroenterology and Hepatology 4:276–287.

Robayo-Torres CC, Opekun AR, Quezada-Calvillo R, Villa X, Smith EO, Navarrete M, Baker SS, Nichols BL. 2009. ^{13}C-Breath tests for sucrose digestion in congenital sucrase isomaltase–deficient and sacrosidase-supplemented patients. Journal of Pediatric Gastroenterology and Nutrition 48:412–418.

Robyt J, French D. 1970. The action pattern of porcine pancreatic agr-Amylase in relationship to the substrate binding site of the enzyme. Journal of Biological Chemistry 245:3917.

Rossi EJ, Sim L, Kuntz DA, Hahn D, Johnston BD, Ghavami A, Szczepina MG, Kumar NS, Sterchi EE, Nichols BL, et al. 2006. Inhibition of recombinant human maltase glucoamylase by salacinol and derivatives. FEBS Journal 273:2673–2683.

Sali A, Blundell TL. 1999. Comparative protein modelling by satisfaction of spatial restraints. Journal of Molecular Biology 234:779–815.

Sampedro J, Sieiro C, Revilla G, Gonzalez-Villa T, Zarra I. 2001. Cloning and expression pattern of a gene encoding an α-xylosidase active against xyloglucan oligosaccharides from Arabidopsis. Plant Physiology 126:910–920.

Samulitis-Dos Santos BK, Goda T, Koldovsky O. 1992. Dietary-induced increases of disaccharidase activities in rat jejunum. British Journal of Nutrition 67:267–278.

Scheen A. 2003. Is there a role for alpha-glucosidase inhibitors in the prevention of type 2 diabetes mellitus? Drugs 63:933–951.

Scotter AJ, Kuntz DA, Saul M, Graham LA, Davies PL, Rose DR. 2006. Expression and purification of sea raven type II antifreeze protein from Drosophila melanogaster S2 cells. Protein Expression and Purification 47:374–383.

Shapiro B, Hofreiter M. 2010. Analysis of ancient human genomes. Bioessays 32:388–391.

Shimada M, Mochizuki K, Goda T. 2009. Feeding rats dietary resistant starch shifts the peak of SGLT1 gene expression and histone H3 acetylation on the gene from the upper jejunum toward the ileum. Journal of Agricultural and Food Chemistry 57:8049–8055.

Sim L, Jayakanthan K, Mohan S, Nasi R, Johnston BD, Pinto BM, Rose DR. 2009. New glucosidase inhibitors from an Ayurvedic herbal treatment for type 2 diabetes: Structures and inhibition of human intestinal maltase-glucoamylase with compounds from Salacia reticulata. Biochemistry 49:443-451.

Sim L, Quezada-Calvillo R, Sterchi EE, Nichols BL, Rose DR. 2008. Human intestinal maltase-glucoamylase: crystal structure of the N-terminal catalytic subunit and basis of inhibition and substrate specificity. Journal of Molecular Biology 375:782–792.

Sim L, Willemsma C, Mohan S, Naim HY, Pinto BM, Rose DR. 2010. Structural basis for substrate selectivity in human maltase-glucoamylase and sucrase-isomaltase N-terminal domains. Journal of Biological Chemistry 285:17763–17770.

Simpson R, Shaw J, Zimmet P. 2003. The prevention of type 2 diabetes—lifestyle change or pharmacotherapy? A challenge for the 21st century. Diabetes Research and Clinical Practice 59:165–180.

Spodsberg N. 2001. Molecular basis of aberrant apical protein transport in an intestinal enzyme disorder. Journal of Biological Chemistry 276:23506–23510.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Research 29:308–311.

Sterchi EE, Lentze MJ, Naim HY. 2004. Molecular aspects of disaccharidase deficiencies. Baillikre's Clinical Gastroenterology 4:79–96.

Taga T, Inagaki E, Fujimori Y, Nakamura S. 1993. The crystal and molecular structure of the trisaccharide erlose as the monohydrate. Carbohydrate Research 240:39–45.

Takusagawa F, Jacobson RA. 1978. The crystal and molecular structure of α-maltose. Acta Crystallographica Section B-Structural Science B34:213–218.

Tanabe G, Xie W, Ogawa A, Cao C, Minematsu T, Yoshikawa M, Muraoka O. 2009.Facile synthesis of de-O-sulfated salacinols: Revision of the structure of neosalacinol, a potent α-glucosidase inhibitor. Bioorganic & Medicinal Chemistry Letters 19:2195–2198.

Vaidyartanam PS. 1993. Indian Medicinal Plants. (Warrier PK, Nambiar VPK, Ramankutty C, editors.). Madras, India: Orient Longman.

Wass MN, Kelley LA, Sternberg MJE. 2010. 3DLigandSite: predicting ligand-binding sites using similar structures. Nucleic Acids Research 38:W469–W473.

Withers, S. and Williams, S. "Glycoside Hydrolases" in CAZypedia, available at URL www.cazypedia.org/index.php/Glycoside_Hydrolases, accessed 14 April 2014.

Xie W, Tanabe G, Akaki J, Morikawa T, Ninomiya K, Minematsu T, Yoshikawa M, Wu X, Muraoka O. 2011. Isolation, structure identification and SAR studies on thiosugar sulfonium salts,

neosalaprinol and neoponkoranol, as potent a-glucosidase inhibitors. Bioorganic & Medicinal Chemistry 19:2015–2022.

Xu D, Zhang Y. 2011. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophysical Journal 101:2525–2534.

Xu D, Zhang Y. 2012. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins 80:1715-1735.

Yamamoto K, Miyake H, Kusunoki M, Osaki S. 2010. Crystal structures of isomaltase from Saccharomyces cerevisiae and in complex with its competitive inhibitor maltose. FEBS Journal 277:4205–4214.

Yoshikawa M, Murakami T, Shimada H, Yamahara J, Tanabe G, Muraoka O. 1997. Salacinol, potent antidiabetic principle with unique thiosugar sulfonium sulfater structure from the ayurvedic traditional medicine Salcia reticulata in Sri Lanka and India. Tetrahedron Letters 38:8367–8370.

Yoshikawa M, Xu FM, Nakamura S, Wang T, Matsuda H, Tanabe G, Muraoka O. 2008. Salaprionol and ponkoranol with thiosugar sulfonium sulfate structure from Salacia prinoides and α-glucosidase inhibitory activity of ponkoranol and kotalanol desulfate. Heterocycles 75:1397–1405.

Yuasa H, Takada J, Hashimoto H. 2000. Synthesis of salacinol. Tetrahedron Letters 41:6615–6618.

Zhang G, Hamaker BR. 2009. Slowly digestible starch: Concept, mechanism, and proposed extended glycemic index. Critical Reviews in Food Science and Nutrition 49:852–867.