# TOWARDS UNIVERSAL BLOOD: MECHANISTIC STUDIES ON BLOOD GROUP CLEAVING GLYCOSIDASES

by

FATHIMA AIDHA SHAIKH

B.Sc., The University of British Columbia, 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

In

THE FACULTY OF GRADUATE STUDIES

(Chemistry)

**THE UNIVERSITY OF BRITISH COLUMBIA**

(Vancouver)

August 2012

# Abstract

The ABO blood groups are vitally important in blood transfusion and organ transplantation. Transfusion with an incorrect blood type results in destruction of the incompatible blood cells, which can result in death. In my thesis, the catalytic mechanisms of three enzymes, two of which can directly be used on red blood cells (RBCs), were investigated in detail as follows.

YesZ, a family GH 42 β-galactosidase (retaining), was used as a model system for the identification of catalytic residues. The mechanism-based inhibitor, 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-galactopyranoside was synthesized and used to inactivate YesZ via trapping of a reaction intermediate. Subsequent proteolytic digestion and comparative MS analysis identified the labeled peptide which, combined with, sequence alignments identified the catalytic nucleophile, a glutamate in the sequence ETSPSYAASL. Use of the acid/base mutant for trapping experiments provided support for its role thereby providing experimental verification of the identities of the catalytic residues in Family GH42.

EABase, a family GH98 endo-β-galactosidase, cleaves blood group A and B trisaccharides from glycoconjugates and RBCs. The mechanism of Family 98 glycosidases was unknown but inferred to be retaining. The DNP-β-A-trisaccharide substrate was synthesized by *in vivo* enzymatic and subsequent chemical methods and direct $^{1}$H NMR analysis of its hydrolysis by EABase revealed that EABase is an inverting glycosidase. Both activated and nonactivated substrates were used to kinetically characterize EABase and its mutants, revealing that D453 and/or E506 act as the base catalyst and that E354 is the acid catalyst. EABase was used, in collaboration with Dr. Kizhakkedathu's lab, to generate "universal blood cells" from type-B blood.

Several α-L-fucosidases from family GH29 (retaining), which cleave α(1,2)-fucose from glycoconjugates were kinetically characterized in the hope of identifying the acid/base residue which is not conserved by sequence. A combination of modeling, sequence comparisons and phylogenetic tree analysis was used to identify candidate acid/base residues and further

subgroup GH29 fucosidases based on these comparisons. The identity of the acid/base residue in four fucosidases is supported by kinetic characterization of a series of mutants of candidate residues and can now be predicted for all Family GH29 fucosidases.

# Preface

A version of chapter 2 has been published. Adapted with permission from Shaikh FA, Müllegger J, He S, Withers SG. *Identification of the catalytic nucleophile in Family 42 beta-galactosidases by intermediate trapping and peptide mapping: YesZ from Bacillus subtilis.* FEBS Lett. (2007); **581**(13):2441-6. Copyright (2007) Elsevier Ltd. The cloning and mutagenesis was performed by Dr. Johannes Müllegger and the LC/MS was run by Dr. Shouming He who also performed the analysis of the peptic digests. The 2FDNPGal used for the manuscript was previously synthesized in the lab—I synthesized a subsequent batch for collaborators. I performed all other experiments and wrote most of the manuscript, Dr. Müllegger and Dr. He helped write the experimental parts for their experiments and my supervisor, Dr. Withers, edited and helped me reword the manuscript.

A version of chapter 3 has been published. Adapted with permission from Shaikh FA, Randriantsoa M, Withers SG. *Mechanistic analysis of the blood group antigen-cleaving endo-beta-galactosidase from Clostridium perfringens.* Biochemistry. (2009); **48**(35):8396-404. Copyright (2009) American Chemical Society. Mialy Randriantsoa from Dr. Samain's lab synthesized the A-trisaccharide and A-pentasaccharide used for the studies (and in the publication). Ms. Maria Ezhova assisted in the acquisition of the time-course $^{1}$H NMR spectra on the AV600 for the stereochemical outcome of the EABase hydrolysis of DNP-A-tri. I performed all other experiments and wrote most of the manuscript. My supervisor, Dr. Withers edited and helped me reword the manuscript. Dr. Nick Rossi and Dr. Iren Constantinescu from Dr. Kizhakkedathu's lab performed the assays of EABase on red blood cells.

The plasmids for the D229G and E288G mutant fucosidases were made by Dr. Alicia Lammerts Van Bueren at the York Structural Biology Laboratory (York, UK). Additionally Dr. Lammerts van Bueren supplied wild type plasmid for the BT4136 and BT1625 fucosidases (along with primers for glycine mutants which were not used). Dr. Lammerts van Bueren assigned E288 as the acid/base mutant for BT2970 based on crystal structure data she had obtained and thus this residue was kinetically characterized.

# Table of Contents

# List of Tables

# List of Figures

# List of Schemes

# List of Abbreviations

| | | |
|---|---|---|
| 2FDNPGal | - | 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-galactopyranoside |
| Abg | - | *Agrobacterium sp.* β-glucosidase |
| AcCl | - | acetyl chloride |
| αFucF | - | α-L-fucopyranosyl fluoride |
| A+PGM | - | A-trisaccharide containing porcine gastric mucin |
| A-tri | - | A-trisaccharide: 2,4-dinitrophenyl 2-acetamido-2-deoxy-α-D-galactopyranosyl-(1→3)-[α-L-fucopyranosyl-(1→2)]-β-D-galactopyranose |
| A-penta | - | A-pentasaccharide: 2-acetamido-2-deoxy-α-D-galactopyranosyl-(1→3)-[α-L-fucopyranosyl-(1→2)]-β-D-galactopyranosyl-(1→4)-2-acetamido-2-deoxy-β-D-glucopyranosyl-(1→3)-β-D-galactose |
| Bcx | - | *Bacillus circulans* xylanase |
| BT | - | *Bacteroides thetaiotaomicron* |
| CAZy | - | carbohydrate-active enzymes |
| Cex | - | *Cellulomonas fimi* endo-glycanase |
| Da | - | Dalton |
| DMF | - | dimethylformamide |
| DFJ | - | deoxyfuconojirimycin |
| DNP | - | dinitrophenol (or dinitrophenyl) |
| DNPGal | - | 2,4-dinitrophenyl β-D-galactose |

| | | |
|---|---|---|
| EABase | - | blood group A/B-antigen cleaving endo β-galactosidase |
| EtOAc | - | ethyl acetate |
| ESI | - | electrospray |
| FACS | - | fluorescence activated cell sorting |
| GH | - | glycoside hydrolase |
| Hfuc | - | human α-L-fucosidase |
| HPLC | - | high performance liquid chromatography |
| IPTG | - | isopropyl β-D-1-thiogalactopyranoside |
| $k_{cat}$ | - | catalytic rate constant |
| $K_M$ | - | Michaelis constant |
| $K_I$ | - | dissociation constant |
| LC/MS | - | liquid chromatography/ mass spectrometry |
| LIC | - | ligation independent cloning |
| MeOH | - | methanol |
| MS | - | mass spectrometry |
| NMR | - | nuclear magnetic resonance |
| PDB | - | protein data bank |
| PGM | - | porcine gastric mucin |
| pNP | - | para-nitrophenol (para-nitrophenyl) |
| pNPGal | - | para-nitrophenyl β-D-galactoside |

| | | |
|---|---|---|
| pNPαFuc | - | para-nitrophenyl α-L-fucoside |
| $R_f$ | - | retention factor |
| Sp3GH98 | - | blood group A/B cleaving endo-β-galactosidase from *Streptococcus pneumoniae* |
| Sp4GH98 | - | Lewis Y antigen cleaving endo-β-galactosidase from *Streptococcus pneumoniae* |
| Ssfuc | - | α-L-fucosidase from *Solfaribacter solfataricus* |
| TLC | - | thin layer chromatography |
| Tmfuc | - | α-L-fucosidase from *Thermatoga maritima* |
| UV | - | ultra-violet |
| YesZ | - | β-galactosidase from *Bacillus subtilis* |
| WT | - | wild type |

# List of Amino Acid Abbreviations

| | | |
|---|---|---|
| Ala | A | Alanine |
| Arg | R | Arginine |
| Asn | N | Asparagine |
| Asp | D | Aspartic acid (Aspartate) |
| Cys | C | Cysteine |
| Gln | Q | Glutamine |
| Glu | E | Glutamic acid (Glutamate) |
| Gly | G | Glycine |
| His | H | Histidine |
| Ile | I | Isoleucine |
| Leu | L | Leucine |
| Lys | K | Lysine |
| Met | M | Methionine |
| Phe | F | Phenylalanine |
| Pro | P | Proline |
| Ser | S | Serine |
| Thr | T | Threonine |
| Trp | W | Tryptophan |
| Tyr | Y | Tyrosine |
| Val | V | Valine |

# Acknowledgements

First and foremost, I am deeply grateful to my research supervisor, Dr. Stephen Withers, for his help, mentorship, encouragement and inspiration throughout the course of my degree. Without his hard work, guidance, insight, inspiration, and most importantly support and patience, this work would have not been possible. I feel very lucky that I have been a part of his research group and for all the help he has given me. The experiences I have gained in this lab have helped me grow significantly both as a person and as a researcher.

I'm very grateful to the many collaborators who have contributed to this work. In particular I'm grateful for the blood-group antigen synthesis by Mialy Riandriantsoa of Dr. Samain's lab and for the plasmids and primers from Dr. Alicia Lammerts van Bueren—their work was critical to the success of the EABase and fucosidase projects. I want to express my gratitude to Dr. Martin Tanner for his reading of this thesis. My thanks also go to members of the Chemistry Department who have helped in the projects including Ms. Zorana Danilovic for her help in NMR and Dr. Yun Ling for help with mass spectrometry.

I would like to sincerely thank the past and present members of the Withers group who have helped me immensely throughout the years. My special thanks go to Dr. Hongming Chen for all his help in synthesis, Dr. Shouming He for his help and work In ESI-MS, Dr. Ran Zhang for his help with the fluoride electrode and kinetics and enjoyable conversations, Dr. David, Kwan for

reading parts of my thesis and help in lab, Dr. Bojana Rakic for reading parts of my thesis, having wonderful discussions about our projects and lives, and making NCE324 an enjoyable place to be, Ms. Miranda Joyce for all her help with administrative materials and being a friendly face to talk to, Mr. Stephen Reid for teaching me synthetic techniques when I was an undergraduate, Dr. Luke Lairson for being an awesome Chemistry 449 graduate supervisor and teaching me my molecular biology techniques and Dr. Ian Greig for his help with kinetics and quantum chemistry. An extra special thank you to Ms. Emily Kwan for all her molecular biology help, support, and awesome company in lab.

I am very grateful to the Michael Smith Foundation for the MSFHR Senior Student Scholarship and to the UBC department of chemistry for the Gladys Estella Laird Research Fellowship.

I am eternally grateful for all of the support from my family and friends throughout the years—you've made my graduate years at UBC immensely memorable, enjoyable and full of new learning experiences—I would not be who I am without you.

# Chapter 1: Introduction

## 1.1 General introduction

Poly- and oligosaccharides are some of the most abundant organic molecules in nature.[1] They are found as structural biopolymers, such as cellulose and chitin, as storage polymers and sources of metabolic energy, such as starch and glycogen, and as information-bearing oligosaccharides, such as blood group determinants, cell surface antigens, glycolipids and glycoproteins.[2] Since these glycan structures play important roles in biological processes, such as cell signaling, ligand-receptor binding and cell-cell recognition, the clinical administration of these molecules or their analogues, free or as glycoconjugates, has been suggested as a therapeutic approach to alleviate symptoms of cancer, diabetes, Alzheimer's and Parkinson's diseases and autoimmunity arising from transplantation.[3–8] As a result of their biological importance and role in medical applications and treatments, carbohydrates and the enzymes that act on them are the subjects of extensive research efforts.[9] One group of enzymes which acts on carbohydrates is the glycosidases which transfer glycosyl moieties from a glycoside to water, i.e. hydrolyze them, under mild conditions. These enzymes are used extensively in industry, in clinical medicine, transfusion medicine, and screening and synthetic applications.[9] Understanding the mechanism by which these enzymes hydrolyze glycosidic bonds, and the knowledge of specific residues involved in catalysis are fundamental to their use in medicine and industry and for the development of these enzymes as synthetic tools or as more efficient hydrolytic tools.

## 1.2 The role of carbohydrates in blood types, transfusion and transplantation

Blood types are classifications of blood based on the presence or absence of antigenic substances on the surfaces of red blood cells. These antigens are protein, carbohydrate, glycoprotein or glycolipid in nature.[10,11] The two major blood types are the ABO type and the RhD factor, the first of which is carbohydrate in nature and the most important blood group system to consider in transfusion and the second of which is protein in nature.[12,13] The ABO blood groups differ through the identities of the sugar antigens on the surface of the red blood cells: a terminal α(1, 3)-linked N-acetylgalactosamine within the sequence GalNAcα(1,3)-(Fucα1,2)-Galβ- for the A-antigen or an α(1,3)-linked galactose within the sequence Galα(1,3)-(Fucα(1,2)-Galβ- for the B-antigen, both of which are absent in the O-blood type (Fuc-α 1,2-Galβ-) (Figure 1.1).[11,14]

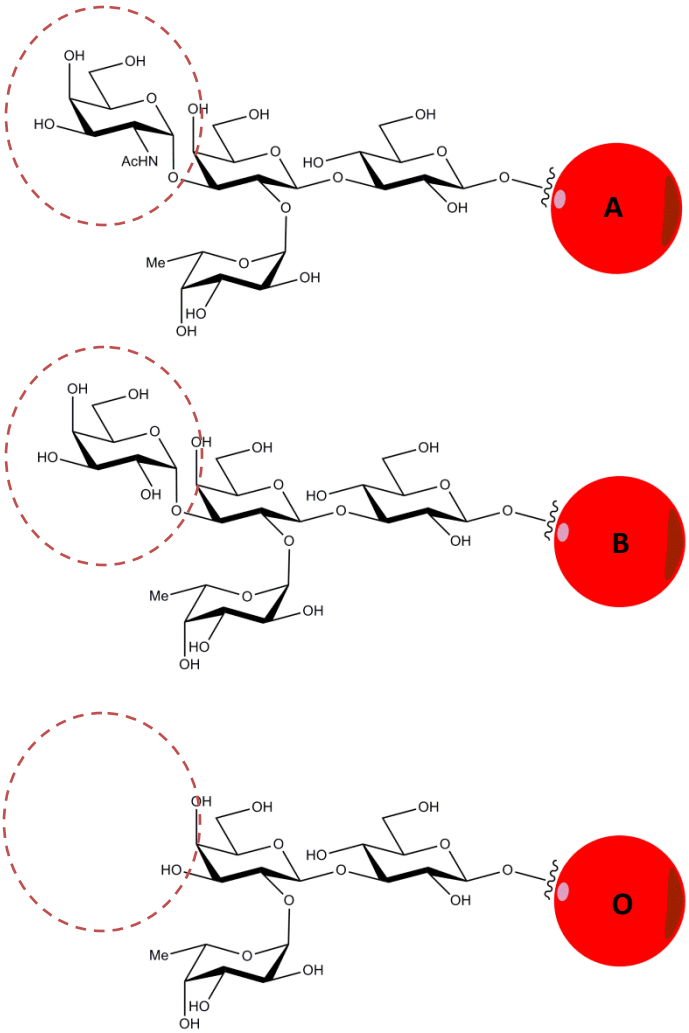



***Figure 1.1*** *Blood group trisaccharides A, B and H (type O) illustrating the single carbohydrate difference between the three blood types A, B and O. Type AB consists of both A and B antigens.*

Since all individuals have antibodies in their blood plasma to the antigen that they lack, transfusing the incorrect blood type or an organ from a donor with the incorrect blood type results in the lysis of the incompatible blood cells and/or an immune response attacking the implanted organ, which can be fatal.[15] Thus, these antigens are of vital importance in blood transfusion and organ transplantation. The major ABO blood types are A, B, AB, and O; the A blood type has two subtypes $A_1$ and $A_2$ (which are discussed in Chapter 3).[16] Currently blood

type O donors are in high demand as their cells are suitable for transfusion to individuals of any ABO blood type. In times of blood shortages, there is often no O-type blood available which makes it difficult to find suitable blood for certain individuals.[17] The development of an effective and efficient enzyme-conversion technology to convert any type of blood into antigen-null blood would have a significant positive impact on transfusion medicine. This thesis aims to understand the mechanism(s) of enzymes that are relevant to blood group antigens and assist in the development of processes to generate universal blood cells.

The A and B antigens are biosynthesized by glycosyltransferases that transfer GalNAc or Gal from nucleotide donors (UDP-GalNAc or UDP-Gal) to an oligosaccharide chain present on the red blood cells. In particular they are transferred by specific α-(1,3)-galactosyl or N-acetylgalactosaminyltransferases.[11,13] These enzymes are encoded by the A or B alleles of the ABO gene. The O allele contains a frameshift mutation and therefore produces no active enzyme.[7,8] O-blood types therefore lack the terminal galactose or N-acetylgalactosamine residue. The H (the H-disaccharide is blood type O) locus encodes a fucosyltransferase which transfer fucose onto galactose. There are generally no antibodies to the H-antigen; however people with the rare Bombay phenotype, in which they do not express the H antigen, can only receive blood which has no H antigen as they produce anti-H antibodies.[7,11,18,19] Once again the development of antigen-null cells would overcome transfusion issues and the use of an enzyme such as EABase or of a collection of enzymes—α-galactosidase/N-acetylgalactosidaminidase, α-L-fucosidase, and β-galactosidase—which cleaves the entire antigen would allow for transfusion into people with rare blood types as well.

## 1.3 Glycosidases

Glycosidases, or glycoside hydrolases, comprise a large superfamily of enzymes which catalyze the hydrolytic cleavage of the C-O bond between the anomeric carbon of the donor sugar, known as the glycone, and the aglycone. While aglycone technically refers to a non-sugar leaving group (i.e. phenolates, lipids), the term aglycone in this thesis refers to all leaving groups including other sugar moieties (Figure 1.2). This glycosidic linkage is extremely stable under neutral aqueous conditions, with a half-life of five million years for a common glycosidic linkage.[20,21] While glycosidases are typically involved in the catabolism of complex oligo- or polysaccharides (where the aglycone is another carbohydrate moiety), the flexibility in aglycone specificity allows for the use of synthetic glycosides which contain aryl or halide groups thereby making characterization and kinetic evaluation of these enzymes more feasible. In the case of hydrolysis, the leaving group (aglycone) is displaced by a water molecule.

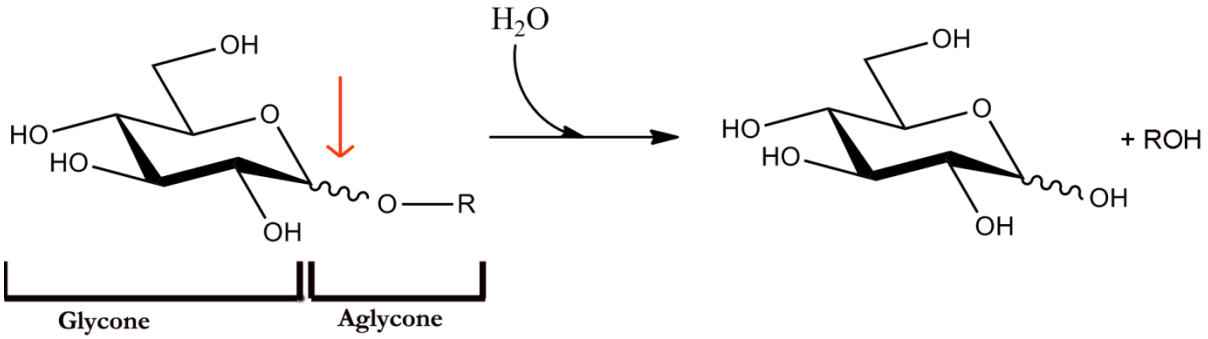
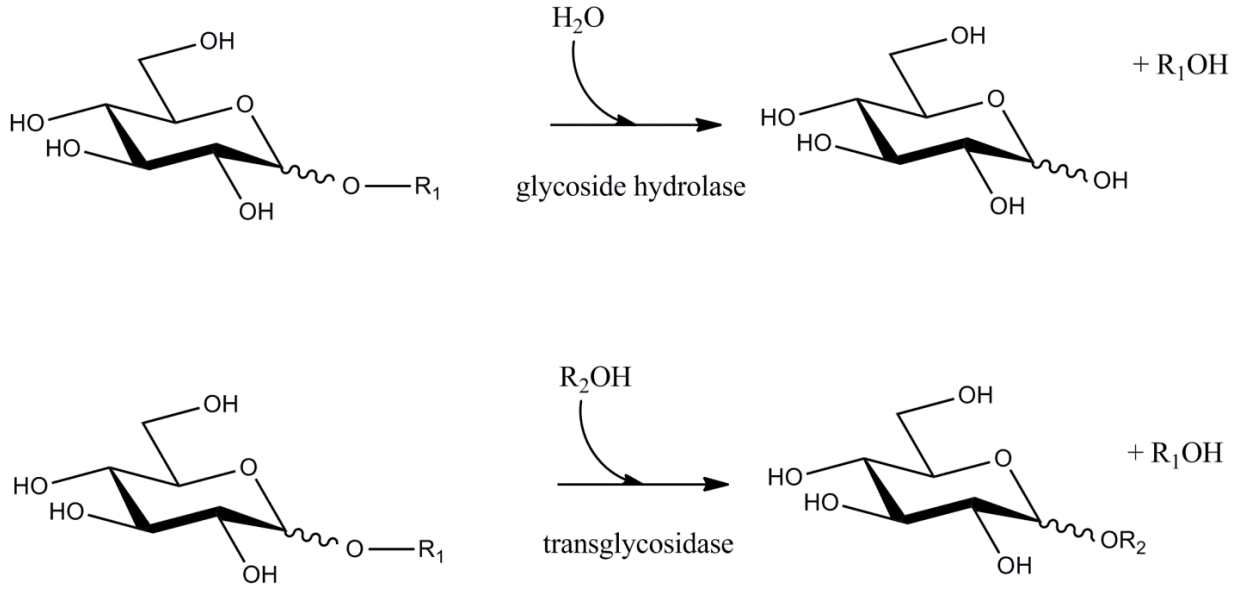



***Figure 1.2:*** *The glycone and aglycone moieties of a β-glucoside are shown and the position of the glycosidic bond which is cleaved is indicated with an arrow.*

Glycosidases can also catalyze transglycosylation reactions where the acceptor is a free hydroxyl group on another carbohydrate moiety or other entity (Scheme 1.1).[22] In contrast to their aglycone promiscuity, glycosidases are fairly specific towards their glycone portion and are commonly named or classified based on their glycone specificity.



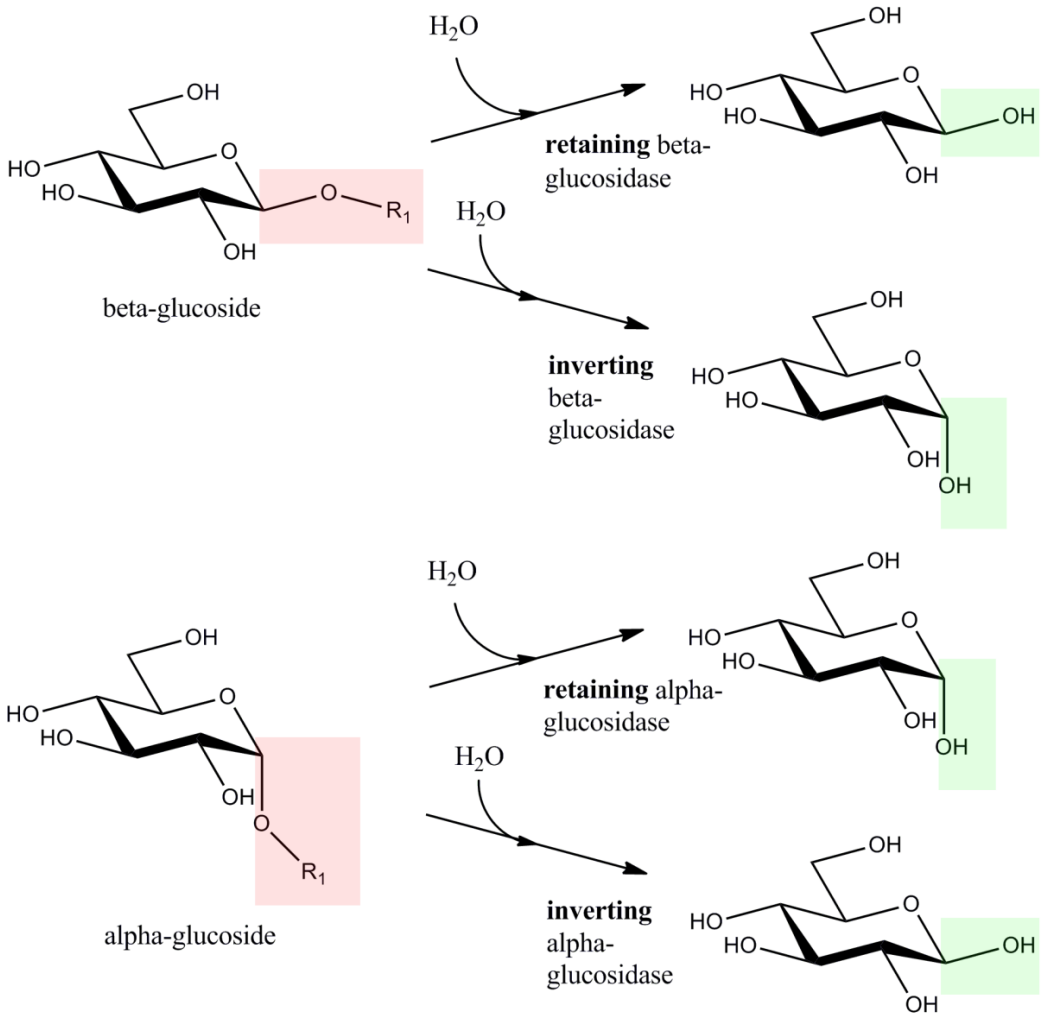

***Scheme 1.1:*** *Hydrolysis versus transglycosylation by a β-glucosidase*

### 1.3.1 Classification of glycosidases

#### *1.3.1.1 Traditional classification systems*

There are several methods that have been devised to classify glycosidases. The simplest of these systems is one devised by the International Union of Biochemistry and Molecular Biology (IUBMB) which classifies glycosidases according to the reaction they catalyze and their glycone specificity. While these rules are simple they are limited in value because they do not provide any information on the reaction mechanism or on the range of preferred substrates. Glycosidases can also be classified on the basis of their glycone specificity, the anomeric configuration of the glycosidic linkage cleaved and the stereochemical outcome of the reaction they catalyze (Scheme 1.2).

As mentioned previously, the specificity towards the glycone component is fairly stringent; glycosidases usually have optimal activity on one type of sugar over others and thus are classified on that basis: a galactosidase would be most active on galactosides but may also have activity (at a lower rate) on other glycosides such as glucosides, mannosides etc. Glycosidases are also subdivided based on the stereochemistry of the anomeric centre on the substrate they act on; they can either be α or β glycosidases and only cleave the corresponding glycoside (i.e. a β-glycosidase will only hydrolyze β-glycosides and not α-glycosides).




***Scheme 1.2****: Reactions catalyzed by retaining and inverting β-glucosidases and anomeric specificity.*

The most mechanistically useful classification is the “inverting” or “retaining” classification which is based on whether the enzyme cleaves the glycosidic bond with net inversion or retention of anomeric configuration (Scheme 1.2). If a β-glycosidase cleaves a β-glycoside and yields a hemi-acetal with the α-configuration as the first-formed product then it is an inverting glycosidase, but if it yields a hemi-acetal with the β-configuration then it is a retaining glycosidase. Therefore the overall name of a glycosidase is a combination of these three

classifications and an enzyme that catalyzes the hydrolysis of a β-galactoside more efficiently than other glycosides and with net retention of anomeric configuration (yielding β-galactose as a product) would be named a β-retaining galactosidase. This classification will be used throughout this thesis to describe the glycosidases worked on.

#### *1.3.1.2 Endo vs. exo glycosidases*

An additional classification used to categorize glycosidases that work on polysaccharides is the "exo" or "endo" classification which stems from the fact that the natural substrates for these glycosidases are polymers.[23] Exo-glycosidases cleave terminal sugar residues from either ends of the polymer. Exo-glycosidases typically release mono- or disaccharides from polymers and have a pocket shaped active site which is meant to accommodate the end of the polymer. In contrast, endo-glycosidases cleave in the middle of an oligosaccharide chain. Endo-glycosidases tend to have a groove-shaped active site with a number of subsites to accommodate the polymer substrate and bind with each monomeric subunit. While the classification of endo versus exo does not give any mechanistic information, it offers some correlation with the tertiary structure of some glycosidases. Figure 1.3 illustrates the difference in cleavage site between endo and exo-glycosidases.

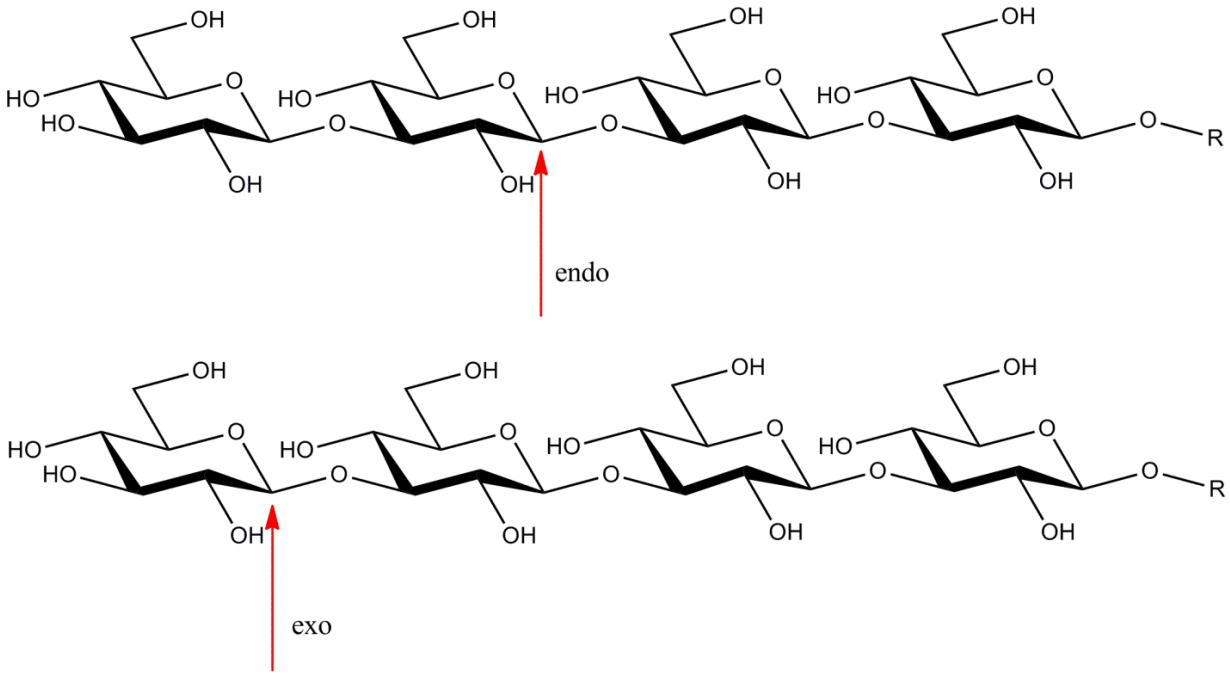



***Figure 1.3:*** *An illustration of the endo and exo cleavage sites of an oligosaccharide by glycosidases*

*1.3.1.3 Glycosidase families*

The previous classifications of glycosidases, while very useful, do not reflect the structural features of the enzymes, do not take into account evolutionary events such as divergent or convergent evolution and are not applicable to enzymes that have a broad specificity. To address these issues, glycosidases have been categorized into numerous families on the basis of their amino acid sequence similarity. This extremely useful classification system was developed by Henrissat and can be found at www.cazy.org.[24–27] There are currently 130 unique glycoside hydrolase families listed, though some have been deleted after being shown to not possess glycoside hydrolase activity.[28] Members within a family are predicted to have the same three-dimensional fold and catalyze reactions with the same stereochemical outcome, though there

are exceptions such as families 23 and 97 which contain both inverting and retaining enzymes.[29] As the tertiary structures and primary sequence similarities are often high within members of a family, predictions of key catalytic residues can be made.[30] As a result, extrapolation of catalytic residues for non-characterized members of the family is quite reliable generally. In addition to being grouped into families based on sequence alignment, glycosidases belonging to different families are also grouped into different clans based on similarities in tertiary fold, despite having low sequence similarity to each other.[26,31,32] It is presumed that these families are evolutionarily related. Therefore, the CAZY database enables one to predict the structure and function of a newly discovered glycosidase using only bioinformatics.

## 1.4 Mechanism of glycosidases

Glycosidases are extremely efficient at catalyzing the hydrolysis of the glycosidic linkage; the half life of this linkage under physiological aqueous conditions is 5 million years while glycosidases can catalyze this reaction under similar conditions with rates of up to 1000 $s^{-1}$.[20,21] The mechanism of glycosidases is formally a nucleophilic substitution at the anomeric centre carbon. If the nucleophile in the enzymatic reaction is a water molecule then the enzyme has hydrolase activity; this is the most common activity amongst enzymes which cleave the glycosidic bond. Another sugar or other alcohol can also act as the 'incoming nucleophile' or acceptor; if the enzyme is able to transfer the glycosyl donor onto an acceptor sugar then it has transglycosylation activity. The mechanisms by which either transglycosylation or hydrolysis

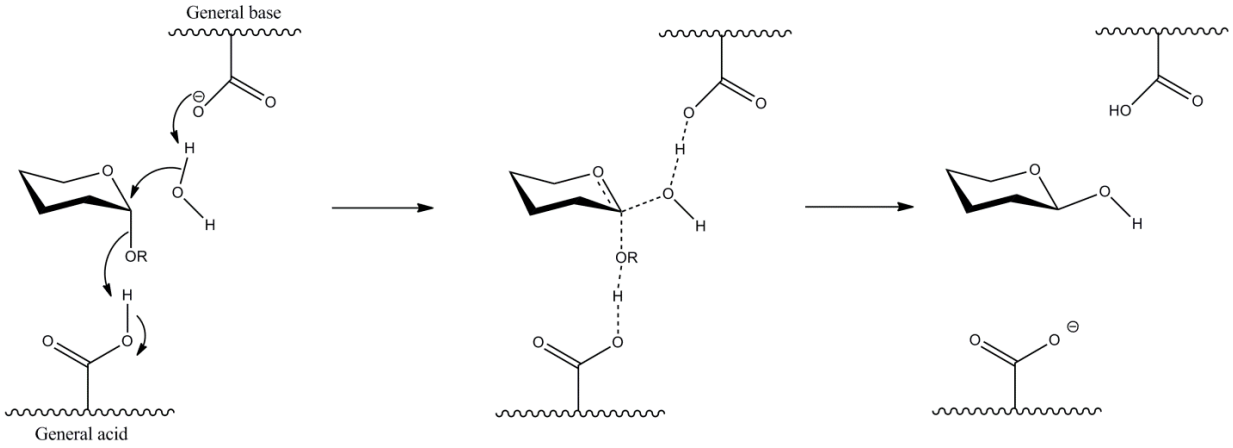
happen are the same and many glycosidases that catalyze hydrolysis also catalyze transglycosylation.

The original versions of the inverting and retaining mechanisms were first proposed by Koshland in 1953 and were based on the mechanism of analogous uncatalyzed reactions.[22,30] He proposed that inverting glycosidases worked by a single step nucleophilic displacement of the aglycone by water and that retaining glycosidases work via a double displacement mechanism that involves a covalent glycosyl-enzyme intermediate. The general principles of these mechanisms still hold though the specific details vary.[33–37] The specific mechanism(s) for both cases will be described in detail in the following sections.

### 1.4.1 Mechanism of inverting glycosidases

Koshland first proposed the mechanism of inverting glycosidases to be a single displacement of the aglycone by a water molecule. The reaction goes through an oxocarbenium ion-like transition state which is similar to that formed during acid-catalyzed hydrolysis (non enzymatic) of glycosides.[30] Evidence for the oxocarbenium ion-like transition state comes from moderate-to-large α-secondary deuterium kinetic isotope values. These values ranging from $k_H/k_D$ = 1.09-1.17 indicate a large positive charge development at the transition state. [38,39] Additionally, modeling of the transition state for an inverting glucosidase's hydrolysis of glucosyl fluoride shows that the anomeric centre has a trigonal planar geometry, consistent with a single-step nucleophilic displacement reaction, and that the endocyclic oxygen's bond order is 1.92,

consistent with that of an oxocarbenium ion-like transition state.[40] The oxo-carbenium ion transition state has its counterpart in retaining glycosidases and will be discussed in later sections. Inverting glycosidases have two active site catalytic carboxylic acid residues. One residue acts as a catalytic base which deprotonates the incoming water molecule. The other carboxylic acid acts as a catalytic acid which protonates the glycosidic oxygen assisting the aglycone in leaving. A scheme of this mechanism is provided in Scheme 1.3.




***Scheme 1.3*** *: Catalytic mechanism of inverting glycosidases*

Removal of the acid catalyst should slow the hydrolysis of substrates with poor leaving groups much more than substrates with good (activated) leaving groups (pKa

<7) which need little protonic assistance. However, removal of the base residue should substantially slow hydrolysis of all substrates. All crystal structures of inverting glycosidases show two active site carboxylic acid residues separated by a distance of 7-10.5 Å allowing for the substrate and nucleophile (water) to bind simultaneously.[41–44] Transglycosylation activity has not been detected in inverting glycosidases likely because the product of the hydrolysis

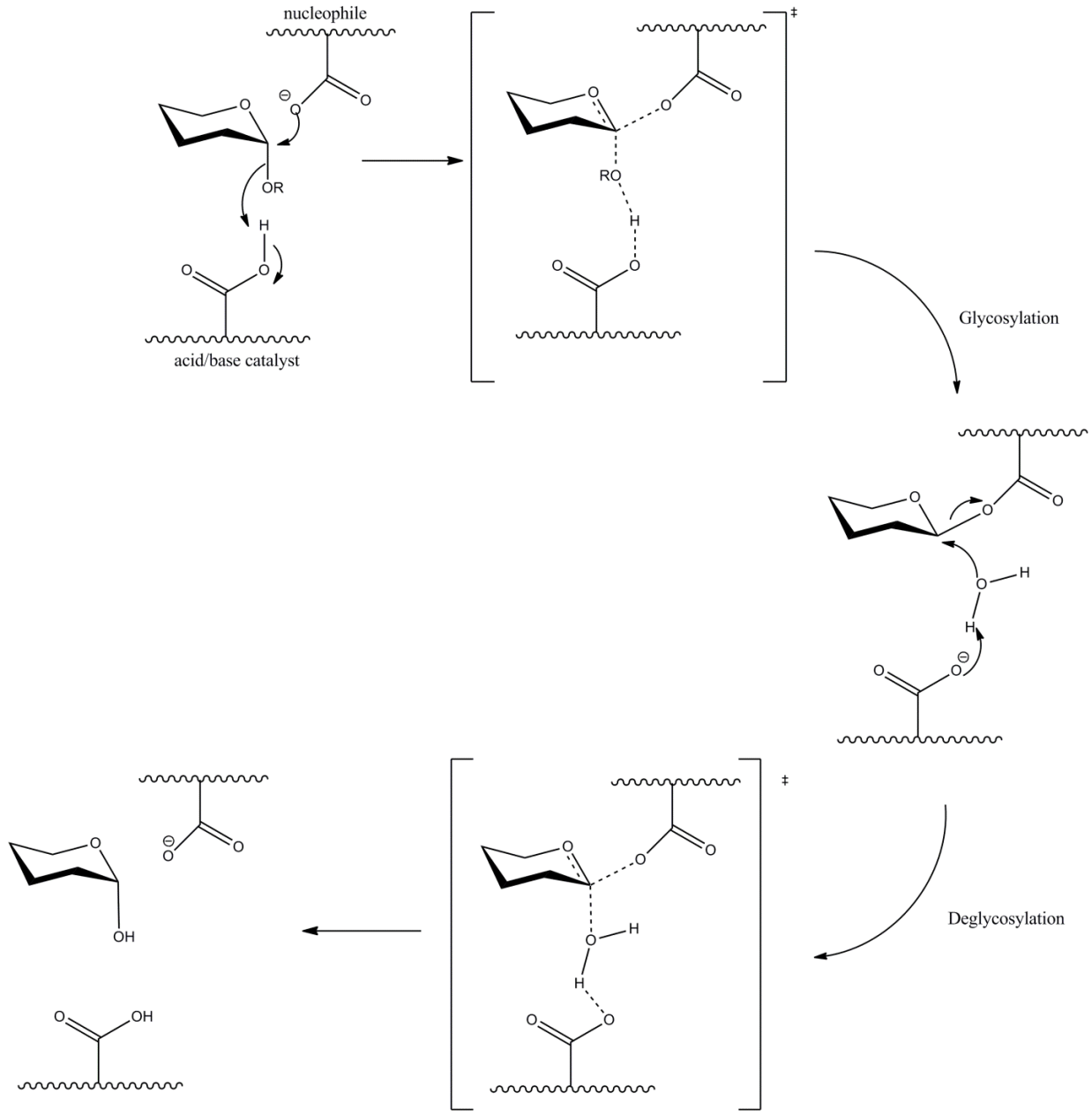
reaction will have the “wrong” or inverted configuration at the anomeric centre and therefore cannot be a substrate for the reverse reaction in order to fulfill the requirement of microscopic reversibility.[22]

Inverting glycosidases are also able to use not only the appropriate anomer of a glycosyl fluoride as a substrate but also the inverted or “wrong” anomer. Thus, for example, inverting β-amylase from sweet potato, whose natural substrate is the terminal disaccharide from α-(1,4) linked amylose, can use both α-maltosyl fluoride and β-maltosyl fluoride as substrates, converting them into β-maltose.[45] The amylase has been shown to perform an initial glycosyl transfer reaction between two β-maltosyl fluoride molecules to yield β-maltotetraosyl fluoride (where the maltose residues are linked α-(1,4)— consistent with the enzyme’s inverting activity). The α-linkages between the two maltosyl residues are then cleaved by the α-amylase according to the normal mechanism to yield β-maltose.[45]

### 1.4.2 Mechanism of retaining glycosidases

Koshland originally proposed a two-step double-displacement mechanism for retaining glycosidases.[30] The first step of this mechanism is the glycosylation step where a catalytic nucleophile, typically, but not always, a carboxylic acid residue, attacks the anomeric carbon of the glycone from the backside.[46] As a covalent bond is being formed between the catalytic nucleophile and the glycone, the aglycone’s departure is assisted by the general acid residue which protonates the aglycone. A covalent glycosyl-enzyme intermediate with the opposite

(inverted) anomeric configuration is thus generated upon departure of the aglycone. The second step of this mechanism is the deglycosylation step in which a water molecule (or other nucleophile) attacks the covalent glycosyl-enzyme intermediate resulting in the hydrolysis of the intermediate. The residue which acted as the general acid in the glycosylation step acts as a general base in the deglycosylation step to deprotonate the incoming water (or nucleophile) assisting in the attack. The attack of the incoming water on the glycosyl-enzyme intermediate inverts the configuration again, resulting in a net retention of stereochemistry. As previously mentioned other sugar moieties can act as the incoming nucleophile (acceptor), if present at suitably high concentrations resulting in the formation of another glycosidic linkage; this is referred to as transglycosylation. A summary of this mechanism can be seen in Scheme 1.4. Unlike inverting glycosidases, the distance between catalytic residues (typically carboxylates) is shorter, typically 5 Å apart to allow direct attack of the nucleophile on the anomeric centre.[41] Numerous studies have been performed to verify this mechanism and will be discussed in the following sections.




***Scheme 1.4:*** *General double displacement mechanism of retaining glycosidases*

### *1.4.2.1 The catalytic nucleophile*

Crystal structures of retaining glycosidases with bound ligands show a carboxylic acid residue positioned to act as a nucleophile (exceptions to the carboxylic nucleophile nucleophile exist and will be discussed briefly later in this section).[46] Sequence comparisons of retaining β-glycosidases within their families show complete conservation of these carboxylate residues supporting their importance in the function of these enzymes.[28] The role of the catalytic nucleophile has been elucidated through kinetic analyses and trapping experiments (discussed in section 1.4.2.3).

As the nucleophile residue is important for the formation of the glycosyl-enzyme intermediate, mutation of the residue to a catalytically incompetent residue (such as glycine, alanine etc) should render the enzyme essentially inactive. Such a result was observed with mutation of the catalytic residue in the β-retaining glucosidase from *Agrobacterium sp.* (Abg), resulting in a $10^7$ fold decrease in $k_{cat}$.[47] This loss of activity was "rescued" in the presence of a small nucleophilic anion, the azide ion, resulting in formation of a glycosyl-azide product with net-inversion of configuration.[41,48] This suggests that the azide ion replaces the catalytic carboxylate as the nucleophile. Chemical rescue using azide (or formate or acetate) ions is a common method for verifying the identity of the nucleophile residue after mutagenesis to a catalytically incompetent residue.[29,35,49–53]

Notable exceptions to the nature of the catalytic nucleophile exist, one of which is amongst hexosaminidases in families 18 and 20, and hyaluronidases in family 56 which utilize the oxygen atom on the N-acetamido group (present on the glycone) as an intramolecular nucleophile and hence no 'nucleophile' is present in the enzyme.[54–57] Other examples include the trans-sialidase from *Trypanosoma cruzi* as well as GH 33 sialidases and trans-sialidases from *Trypanosoma rangeli* and *Micromonospora viridifaciens* where a tyrosine residue plays the role of nucleophile.[58–60] In addition to the differences in catalytic nucleophile for these enzymes there are two families of retaining glycosidases, GH 4 and GH 109, which utilize a unique $NAD^+$-dependent redox elimination/addition sequence which is substantially different from the mechanism of retaining/inverting glycosidases discussed so far.[61–63] These mechanisms will not be further discussed in this thesis.

### *1.4.2.2 The oxocarbenium-ion like transition state and the covalent intermediate*

Both the transition states for the glycosylation and deglycosylation steps resemble oxocarbenium ions with the positive charge distributed between the ring oxygen and anomeric carbon as illustrated in Scheme 1.4 and partial double bond character between the two. This partial double bond requires C-5, O-5, C-1 and C-2 to all be on the same plane. In order for this to occur the pyranose ring must adopt either a half-chair or a boat conformation in the transition state. The structure of the oxocarbenium-ion-like transition state has been studied extensively through kinetic isotope effects and transition-state analogue inhibition studies.[33,46,64–66] A comparison of many of the rates of hydrolysis of the normal substrate versus

one deuterated at the anomeric carbon ($k_H/k_D$) for both glycosylation and deglycosylation steps showed that the ratio ($k_H/k_D$) was greater than unity in each case.[22,66,67] Such a positive kinetic isotope effects are indicative of a change of hybridization from $sp^3$ to $sp^2$ between ground state and transition state. This is consistent with the transition states for both glycosylation and deglycosylation having significant oxocarbenium-ion character. Since enzymes catalyze reactions by binding to the transition state preferentially over the ground state, transition-state analogues should act as good inhibitors of enzymes. This implies that compounds which resemble oxocarbenium ions will be tight-binding inhibitors of glycosidases. The characteristic features of the transition state include a positive charge at C1 or O5 with $sp^2$ hybridization and a coplanar arrangement of C5, O5, C1 and C2 (half-chair configuration).[68,69] A number of compounds which imitate either the shape or charge distribution stated above have turned out to be good inhibitors for retaining β-glycosidases (Figure 1.4).

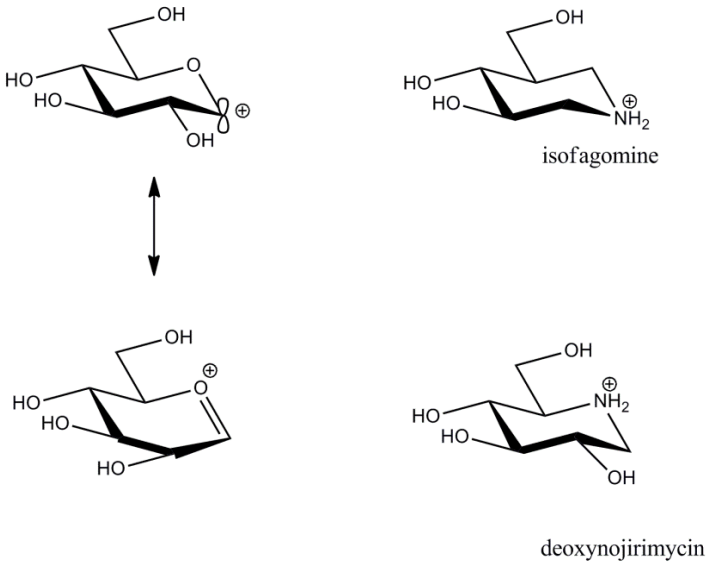



***Figure 1.4*** *Structures of deoxynorjirimycin and isofagomine, inhibitors which imitate the charge distribution of the transition state as illustrated*

The intermediate formed as a consequence of the glycosylation step for retaining glycosidases is covalent. This has been shown through kinetic isotope effects as a $k_H/k_D$ of greater than unity has been measured for the deglycosylation step in all studies, which proves that the transition state for deglycosylation has more $sp^2$ character than the preceding ground state.[33] In addition to kinetic isotope effects, the nature of the intermediate has been shown through trapping studies (described in the next section) and via x-ray crystallography.

### *1.4.2.3 Trapping the covalent intermediate*

Trapping the covalent intermediate provides direct evidence for the intermediary of such a species and can allow identification of the nucleophile residue through subsequent peptide mapping. Trapping of the covalent intermediate has been accomplished using several methods, all of which involve decreasing the rate of deglycosylation relative to the rate of glycosylation such that the intermediate accumulates, thereby "tagging" the nucleophile residue with a sugar. The reagents typically used for the labeling of retaining glycosidases are activated 2-deoxy-2-fluoro glycosides such as 2,4-dinitrophenyl β-2-deoxy-2-fluoro-galactoside (for galactosidases).[70,71] The mechanistic principle employed is that the electron-withdrawing fluorine on the carbon adjacent to the site of positive charge development, inductively destabilizes the transition states for both the glycosylation and deglycosylation steps, lowering the rate of each step. However the presence of a highly activated leaving group such as 2,4-dinitrophenol or fluoride ensures that the glycosylation step remains fast. If the difference in rates between the glycosylation and deglycosylation step is large enough, an accumulation of

the covalent glycosyl-enzyme intermediate is observed and if the deglycosylation step is slow enough the enzyme is inactivated.[72,73] A second method for intermediate trapping involves the use of activated substrates with a mutant at the acid/base position and is discussed below (1.4.2.4).[74]

Formation of the inactivated enzyme can be detected by electrospray ionization mass spectrometry (ESI-MS) through an increase in mass of the protein corresponding to the glycosyl moiety appended.[75–77] To identify the specific catalytic residue derivatized, the inactivated enzyme is then subjected to proteolytic digestion and the peptide fragments are separated and analyzed by mass spectrometry. The labeled residue can be identified by comparison with digests of untreated protein. The trapped intermediate can be detected by $^{19}F$ NMR and the trapped covalent intermediate has also been visualized using X-ray crystallography in a number of cases.[78–80]

### *1.4.2.4 Acid/base catalysis*

The acid/base catalyst is also extremely important in the hydrolysis of glycosidic bonds: indeed mutation of this residue to a catalytically incompetent residue drastically lowers rates.[52,81,82] The acid/base residue is typically also a carboxylate residue in retaining glycosidases—the exception being the Family 1 myrosinase from *Sinapsis alba* where the catalytic acid/base is replaced by a glutamine residue and the role of the residue is taken over by exogenous ascorbic acid.[83] As previously stated, the acid/base catalyst is usually 5.5 Å away from the nucleophile

residue and is initially protonated for the glycosylation step (acting as the general acid) and deprotonated for the deglycosylation step (acting as the general base). While the pKa of a typical carboxylic acid is 4.5, the pKa of the acid catalyst in the active site is generally higher, typically ranging between values of 6-8. This is a result of the electrostatic field that is generated by the nucleophile carboxylate residue which is 5.5 A away from the acid/base. This pKa shift maintains a favorable protonation state for the acid/base residue for the glycosylation step. The $pK_a$ of the acid/base residue drops upon formation of the glycosyl-enzyme intermediate, since the nucleophile carboxylate is covalently bonded to the glycosyl moiety and is no longer charged, thereby removing the electrostatic effects that were initially present. This $pK_a$ cycling has been studied in the xylanase from *Bacillus circulans* by $^{13}C$ NMR titration of the enzyme that had been $^{13}C$ labeled at the carboxyl group side chains of both the nucleophile and acid/base residue.[84] The $pK_a$ of the acid/base residue was shown to cycle between 6.7, initially, and 4.2 in the glycosyl-enzyme intermediate allowing it to act as general base. Neighboring residues also modulate the $pK_a$ of the acid/base residue through hydrogen bonding and electrostatic interactions. In the xylanase mentioned above, the acid/base catalyst is hydrogen bonded to an asparagine (Asn35).

Unlike the nucleophile residue, the acid/base residue's positioning is important but not as critical. Mutation of the acid/base catalyst (Glu172) in Bcx to Asp only decreased the second-order rate constant by 400 fold whereas mutation of the catalytic nucleophile from Glu to Asp decreased it 1600-5000 fold.[85–87] X-ray crystal structures of several glycosidases reveal that the acid/base carboxylic acid group can reside in one of two positions, either anti or syn, and function as a proton donor (Figure 1.5).[88–91] The anti or syn designation, defined by Vasella *et*

*al*, refers to the direction from which protonation takes place—either from the anti or syn side of the ring oxygen. [92] This designation was further supported by inhibition studies on a xylanase from *Cellulomonas fimi* (Cex) (an anti protonator) and Bcx (a syn protonator) with an imidazole sugar, which has lone pairs which are anti to the ring oxygen.[92,93] These inhibitors were predicted to interact better with an anti-protonator and indeed did since $K_I$ values with Cex were in the nanomolar range whereas they were in the millimolar range for Bcx.

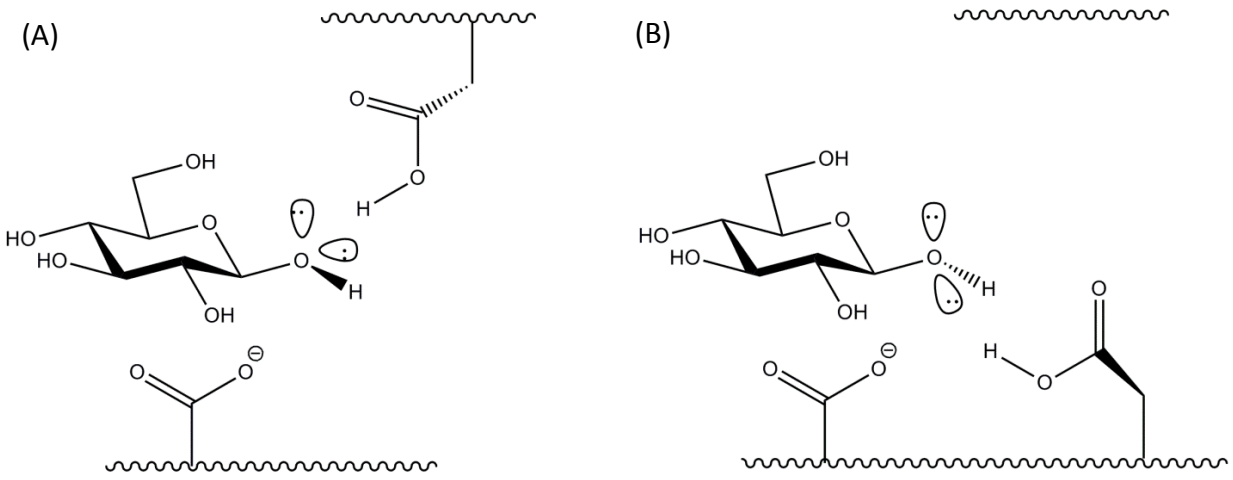



***Figure 1.5*** *Positioning of the acid/base catalyst with respect to the (A) syn versus (B) anti site of protonation*

In the absence of a crystal structure, sequence comparisons are used to identify highly conserved carboxylic acids which are subsequently mutated to non-ionizable amino acids and

the resulting mutants analyzed kinetically. A detailed kinetic analysis involves use of a range of substrates ranging from those with poor leaving groups that need acid catalysis, to those with 'good' leaving groups which do not require protonic assistance for departure. Comparing the activity of both wild type and acid/base mutant enzymes with these substrates is informative, as follows. As the catalytic acid/base residue provides protonic assistance to the departing aglycone, substrates with poor leaving groups are expected to show a much lower activity with mutants modified at that position than ones with good leaving groups; having an activity of approximately $10^5$ fold slower than the wild type. This is a result of the slowing of the glycosylation step of the hydrolysis reaction and is reflected in the second-order rate constant ($k_{cat}/K_m$). For substrates with activated leaving groups, formation of the glycosyl-enzyme intermediate proceeds at rates similar to that of wild type enzyme as the leaving group does not require protonic-assistance. As a result, the glycosyl-enzyme intermediate accumulates when activated substrates are used with an acid/base mutant because the deglycosylation step is also slowed due to removal of the base catalyst. Kinetically, this accumulation can be seen as an initial burst of release of the activated leaving group if sufficient enzyme is employed, and also reflected by an extremely low $K_M$ value. Such intermediate accumulation is also useful when identifying the catalytic nucleophile through mass spectrometry, NMR and/or x-ray crystallography. Additionally the observation of a much lower activity for substrates with poor leaving groups compared to good leaving groups in an acid/base mutant is characteristic of that residue being an acid/base residue.

A second method of identifying the catalytic acid/base residue involves nucleophilic small molecule rescue of activity with acid/base mutants. When using substrates containing good

leaving groups the deglycosylation step becomes rate-limiting in the acid/base mutant as the replacement residue is not able to deprotonate the incoming water. In such cases the incorporation of a nucleophilic ion such as azide (which does not require deprotonation) in the reaction mixture should accelerate the deglycosylation step.[94] The azide binds in the vacant site and reacts with the glycosyl-enzyme intermediate more rapidly than water, forming a glycosyl azide product. In the case of the acid/base mutant of Abg, a 300-fold increase in $k_{cat}$ was seen upon rescue with azide.[94] Unlike rescue of the nucleophile mutant by azide, which produces a glycosyl-azide product with inverted stereochemistry, rescue of the acid/base mutant produces a product with net retention of stereochemistry. The stereochemical analysis of the product from the azide rescue of the mutants can be used to distinguish between the nucleophile and acid/base residue as illustrated in Figure 1.6.

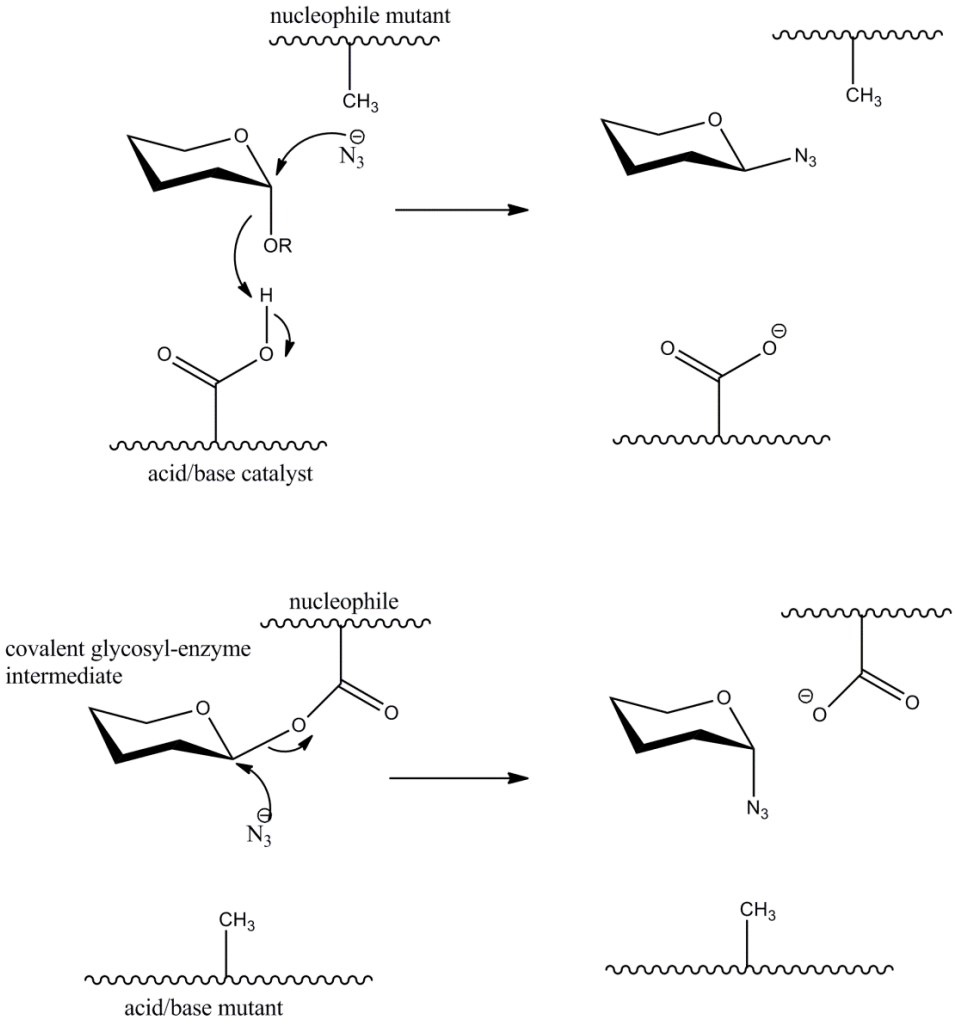



***Figure 1.6*** *: Distinguishing the acid/base mutant from the nucleophile mutant by the product of azide rescue*

*1.4.2.5 Non covalent interactions*

Enzymes catalyze reactions efficiently by binding to the transition state better than the ground state, thereby decreasing the activation energy as suggested by Linus Pauling.[95] This is achieved by non-covalent interactions between the enzyme and substrate which stabilize the geometry and charge of the transition state over the ground state resulting in a rate acceleration. Some of the major interactions, as determined by x-ray crystal structures, are hydrogen bonds between the hydroxyl groups of the substrate sugar and the enzyme.[88] Several methods have been used to determine the significance of these interactions, the first being systematic site directed mutagenesis of the residues that interact with the substrate and the second being the use of modified sugar substrates in which each hydroxyl is replaced by a hydrogen or fluorine neither of which are hydrogen bond donors. In each case kinetic parameters are measured to determine the effect of each substitution. The interactions at the 2-position are particularly important to catalysis as shown by the decrease in activity by a factor of $10^6$ fold for hydrolysis of a 2-deoxy-glycoside substrate by Abg (over the same substrate with a 2-hydroxyl group). Similar rate decreases are observed for other glycosidases when 2-deoxy-glycoside substrates are used. Interactions with the 2-hydroxyl can contribute 18-22 kJ/mol towards binding at the transition state whereas the other hydroxyl groups contribute only 3-7 kJ/mol each.[72] The 2-hydroxyl group sits close to the incoming nucleophilic amino acid and it has been shown to form a short hydrogen bond to the non-bonding carboxylic acid oxygen of the nucleophile, accounting for its importance in stabilizing the transition state.

## 1.5 Applications of glycosidases: towards the development of "universal" or "antigen-null" red blood cells

Current methods of modifying blood types involve the conversion of A- or B-types to type O through the use of glycosidases. The first example of this was the use of an α-galactosidase from coffee beans to convert blood type B cells to blood type O.[96] The major limitations of this method included the large (and costly) quantities of enzyme required and the inability to convert blood type A cells to O. Liu *et al.* made a significant advance in the conversion of A/B-types to O through their production of recombinant glycosidases, from *B. fragilis* and *E. meningosepticum*, which cleave both A and B monosaccharides (GalNAc and Gal) efficiently while leaving the red blood cell intact and fully functional.[97] Both these enzymes have pH optima close to physiological pH and are therefore suitable for cleavage of RBC's without destruction of the cells.

An endo-β-galactosidase from *C. perfringens*, EABase, was shown by Anderson et al. to be capable of hydrolyzing the intact A and B oligosaccharides from intact cell surfaces.[98] Since EABase can cleave both A and B trisaccharides, it is a more attractive target for the conversion of RBCs into universal blood cells since a single enzyme can be used for both blood types. This substrate specificity of EABase, therefore, makes it an attractive target for use in the hydrolysis of the entire trisaccharide from RBCs, and for the potential ability to transfer complex and biologically relevant sugars onto cell surfaces. An understanding of the mechanism of blood group antigen cleaving enzymes will aid in the generation of more efficient enzymes for the conversion of A/B/AB blood types to O or ABO-null cells. The engineering of highly efficient

enzymes that are capable of transferring the A and B antigens onto red blood cells or modifying pre-existing antigens on the cells would help overcome the shortage of donated blood by allowing interconversion between blood types. Such approaches would also be useful in the synthesis of artificial blood or in the generation of related antigens on other cell types.

Another method for the conversion of RBCs into antigen-null cells, which does not use glycosidases, is the camouflaging of RBCs through the use of cell surface-reactive hydrophilic macromolecules such as polyethylene glycol (PEG) in the presence of a non-reactive additive such as hyperbranched polyglycerol (HPG).[99] While these macromolecules efficiently camouflage all the minor antigens and the RhD antigens on the RBC surface, they do not mask the A or B antigens present. As the ABO antigens are the primary immunodeterminants involved for a large proportion of transfusion rejections, it is important to be able to cleave or mask these antigens efficiently. Therefore the development of a glycosidase which can efficiently cleave A/B antigens would hold the most significant promise in the quest for universal blood. However, to develop a truly antigen-null cell, one would have to combine both the polymer camouflaging and the enzymatic cleavage of ABO antigens. Progress on this front will be discussed in Chapter 3. Figure 1.7 summarizes the different methods which are in use and which can be used to generate antigen-null or ABO-null cells.

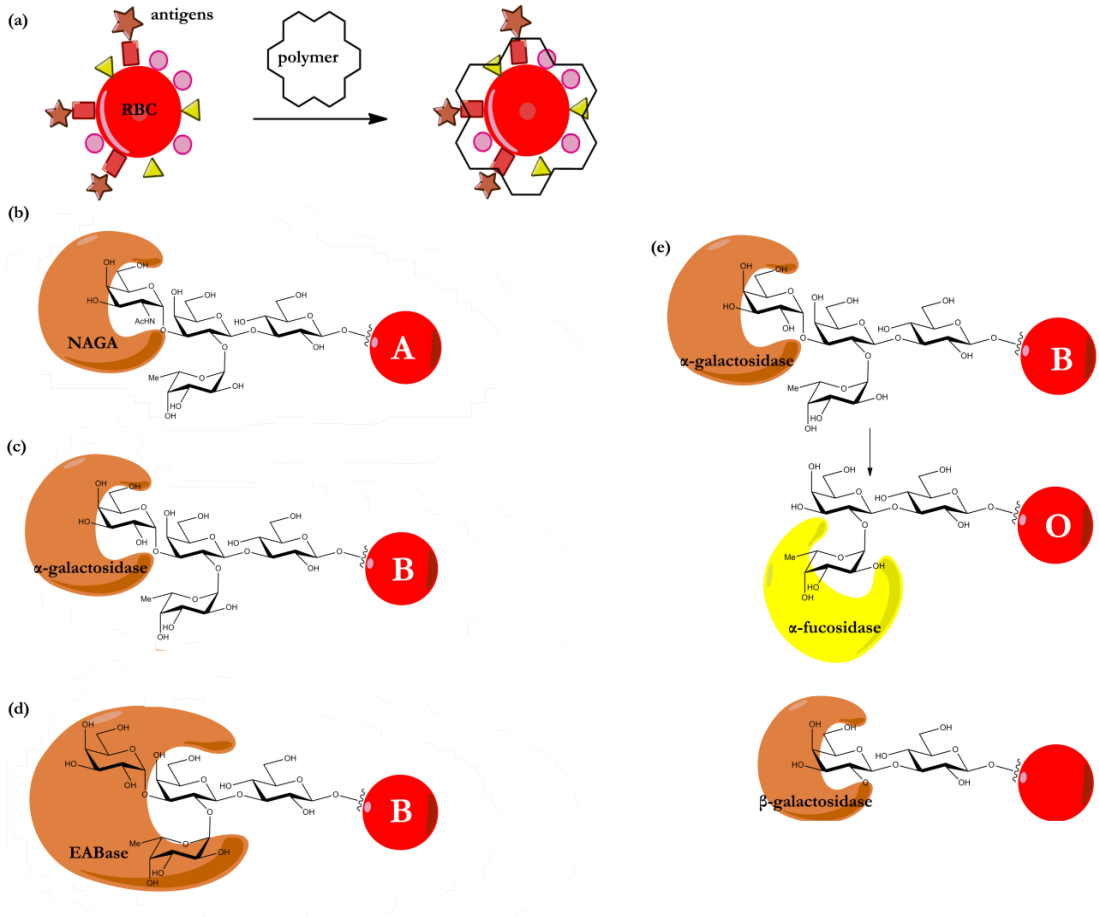



***Figure 1.7*** *A summary of the methods of generating universal blood cells: a) masking of antigens on a red blood cell by the use of a cell surface reactive polymer, b) removal of A-antigenicity through α-N-Acetylgalactosidamindases (NAGA) , c) removal of B-antigenicity through α-galactosidases, d) removal of both A and B antigenicity through blood group cleaving endo-β-galactosidases, and e) removal of antigens sequentially using multiple glycosidases*

## 1.6 Aims of this thesis

The ABO blood groups are of vital importance in blood transfusion and the development of an efficient enzyme conversion technology would substantially increase the availability of blood supplies for all ABO blood types. The current methods for conversion to O-type or "universal" blood involve exo-glycosidases which cleave the terminal N-acetylgalactosamine (for A type) or the terminal galactose (for B type) moiety. Additionally, these enzymes do not address the rare Bombay phenotype in which individuals have anti-H (antigens to the O-type) antibodies and would need blood lacking in the A/B/H antigens. While advances have been made in finding enzymes which can cleave those two moieties under physiological conditions with reasonable efficiency relatively few studies have been done on enzymes (endo-glycosidases) which could cleave off the entire trisaccharide antigen. Anderson *et. al*. isolated, cloned and purified an endo-β-galactosidase that was shown to be capable of liberating both the A and B trisaccharides from glycoconjugates.[98] At the time this enzyme showed no sequence similarity to the other 97 CAZY glycoside hydrolase families identified and was thus placed in a new family, GH98, for which the catalytic residues and mechanism were unknown. As this enzyme has a unique specificity for blood group antigens there was considerable interest in understanding its mechanism and identifying the catalytic residues. The overall objective of this study is to contribute to the understanding of enzymes that are useful for the conversion of blood into "universal" blood and to apply these enzymes in the generation of antigen-null blood cells.

The studies in this thesis were divided into three parts. The first is the understanding of the catalytic residues of the Family 42 exo-β-galactosidase from *Bacillus subtilis*. This enzyme was chosen as a model system for the methodology of mechanistic elucidation for the EABase enzyme and was a family for which the catalytic residues had not been identified at the time. The catalytic residues and mechanism will be identified through sequence analyses, mutagenesis, inactivation kinetics and mass spectrometry. In the second, and main part, of the thesis a detailed investigation on the mechanism and catalytic residues of the blood-group endo-β-galactosidase, EABase will be carried out. This involves the development of a suitable assay for characterization of EABase, sequence and modeled structural comparisons, site-directed mutagenesis, NMR studies and detailed kinetic characterization. Additionally, studies on the use of EABase to generate 'antigen-null' cells will be discussed. Finally, mechanistic studies on Family 29 α-fucosidases will be discussed along with a comparison of structural information (modeled and/or x-ray crystal structure) and sequence information to identify the catalytic residues of this family which are currently controversial and do not appear "conserved" throughout the family based on sequence alignment alone. An understanding of the catalytic residues of GH29 fucosidases will aid in the development of fucosylsynthases which can be used to synthesize substrates for use with the further evolution and testing of EABase and other blood-antigen cleaving glycosidases. Additionally GH29 fucosidases can be used to cleave the fucose residue from the H-antigen as an alternative to generating ABO antigen-null cells from endo-glycosidases such as EABase.

# Chapter 2: Model System For Mechanistic Elucidation: Family GH 42 β-Galactosidase From *Bacillus subtilis*

## 2.1 Family GH 42 galactosidases: introduction and sequence alignment

Glycosidases are well known to play crucial roles in biology, and are increasingly being used in biotechnology, for the cleavage of glycosidic bonds under mild conditions. β-galactosidases are amongst the more biotechnologically interesting glycosidases as they are widely used in the dairy industry for the generation of lactose-free products.[100,101] They are found in bacteria, archaea and eukaryotes and β-galactosidases from fungi (*Aspergillus oryzae* and *A. niger*) and yeast (*Kluveromyces fragilis* and *K. lactis*) are the most commonly used to produce lactose-free (or low-lactose) milk, syrup and oligosaccharide products.[102–104] Additionally, thermostable β-galactosidases from *Thermus sp. A4*, *B. subtilis* (this study) and *Alicyclobacillus acidocaldarius* (from Family 42) have been studied for this purpose as their tolerance to the higher temperature of pasteurization offers a greater advantage in generating lactose-free dairy products over thermolabile galactosidases.[102–104]

Exo-β-galactosidases are found in four principal carbohydrate-active enzyme (CAZY) families to date, GH1, GH2, GH35 and GH42, all of which belong to Clan GH-A. Enzymes in this clan, thus in all four of these families, are believed to use a two-step, double-displacement mechanism in which a covalent galactosyl-enzyme intermediate is formed and hydrolyzed via oxocarbenium ion-like transition states.[67,105] As previously mentioned, the active site of such enzymes contains two key carboxylic acid residues, the catalytic nucleophile, and the acid/base residue. The identification of these key carboxylic acids is one of the first steps in the mechanistic characterization of a retaining glycosidase family.[46] At the time of this study, the catalytic residues had been experimentally defined for the β-galactosidases from Families 1, 2 and

35[106,107], and only inferred for the Family 42 enzymes on the basis of the proposed membership of this Family in Clan GH-A[108], and from the three-dimensional structure of a complex of one of these enzymes (the β-galactosidase from *Thermus thermophilus*) with the product galactose.[102] Since β-linked galactosyl moieties are found extensively in biology, including blood group ABO antigens, and since β-galactosidases have direct use in biotechnology, the understanding of β-galactosidase mechanisms will allow for their efficient use in many applications. This study also serves as a basic model for the mechanistic characterization of a glycosidase; in this case, the family 42 β-galactosidase (YesZ) from Bacillus subtilis.

Family 42 glycosidases have mostly been identified in prokaryotic organisms with a few examples from archaea and fungi. GH42 also includes α-L-arabinosidase and β-D-fucosidase activities.[108,109] Several GH42 hydrolases have been identified from organisms present in extreme habitats where lactose would not be naturally prevalent, and their activities have been shown to be largely on galactooligosaccharides and galactans instead of lactose, suggesting that these enzymes are likely involved in *in vivo* plant cell wall degradation.[104,108,109] The β-galactosidase used in this study is from *Bacillus subtilis*, a gram-positive bacterium, capable of tolerating extreme environmental conditions. This bacterium is commonly found in soil where it plays a role in breaking down plant biomass.[110] Previous work on Family 42 galactosidases includes the determination of the stereochemistry (shown to be retaining) of the *T. thermophilus* A4 β-galactosidase by NMR and (as previously stated) a crystal structure of this galactosidase in complex with the product galactose.[102,104] The first step in identifying the potential catalytic residues for YesZ is a multiple sequence (amino acid) alignment of members of Family 42 and identification of conserved catalytically competent (aspartate or glutamate in

this case) residues. Comparison of the YesZ amino acid sequence from *B. subtilis* with those of other β-galactosidases from family 42 shows complete conservation of both the inferred catalytic nucleophile and acid/base residues as shown in Figure 2.1. The catalytic glutamate residues were inferred based on the crystal structure of *Thermus sp.* A4 in complex with galactose (Figure 2.2) and are conserved.

(A)

| | | | |
|---|---|---|---|
| *Thermus sp.* A4 β-Galactosidase | 306 | GRFWVM**E**QQPGPVNWAPHNPS | 326 |
| *B. Subtilis* β-Galactosidase | 289 | RPFWIL**E**TSPSYAASLESSAY | 309 |
| *C. perfringens* β-Galactosidase | 299 | KPFMMM**E**SSPSSTNWQPVAKL | 319 |
| *B. circulans* β-Galactosidase | 309 | QPFLLM**E**STPSSTNWQEVSKL | 329 |
| *G. stearothermophilus* β-Galactosidase | 297 | QPFILM**E**QVTSHVNWRDINVP | 317 |
| *T. maritima* β-Galactosidase | 110 | GRFWVM**E**QQAGPVNWAPYNLW | 130 |
| *T. neapolitana* β-Galactosidase | 308 | GRFWVM**E**QQAGPVNWAPYNLW | 328 |
| *Thermus sp.* T2 β-Galactosidase | 306 | GRFWVM**E**QQPGPVNWAPHNPS | 326 |
| *H. lucentense* β-Galactosidase | 305 | KPFWVM**E**QQPGDINWPPQSPQ | 325 |

(B)

| | | | |
|---|---|---|---|
| *Thermus sp.* A4 β-Galactosidase | 128 | GGLEAVAGFQTDN**E**YGC | 144 |
| *B. subtilis* β-Galactosidase | 132 | GRLPGLIGWQLDN**E**FKC | 148 |
| *G. stearothermophilus* β-Galactosidase | 135 | KNHPALKMWHVNN**E**YAC | 151 |
| *C. maltaromaticum* β-Galactosidase | 135 | ATHPGIVMWHVNN**E**YTC | 151 |
| *C. perfringens* β-Galactosidase | 131 | KDHPALILWHISN**E**FEG | 147 |
| *Y. pseudotuberculosis* β-Galactosidase | 144 | AHHPAVIGWHISN**E**YGG | 160 |
| *L. interrogans* β-Galactosidase | 128 | GNHPVVIGWQIDN**E**IGH | 144 |
| *T. neapolitana* β-Galactosidase | 128 | GKHPAVVGWQTDN**E**YGC | 144 |
| *H. lucentense* β-Galactosidase | 129 | ADNPHVAGWQTDN**E**FGC | 143 |

***Figure 2.1*** *Partial multiple sequence alignment of representative family 42 enzymes. (A) The region around the putative nucleophile residue: the nucleophile is in bold and underlined. (B) The region around the putative acid/base catalyst: the candidate residue is shown in bold and underlined.*

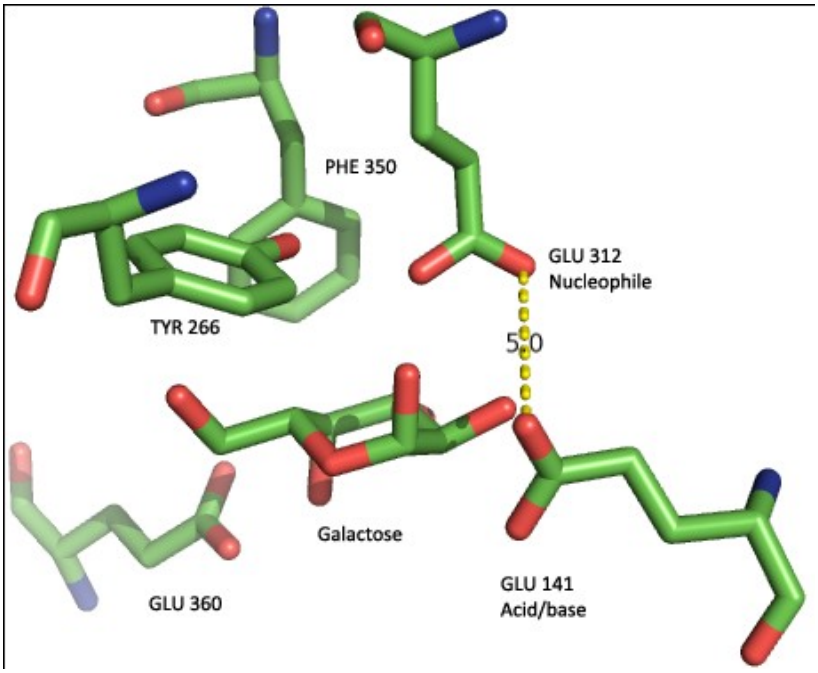



***Figure 2.2*** *Crystal structure of the active site of Thermus sp. A4 with bound galactose product. The putative acid/base residue and nucleophile are labeled and the distance between the two residues is shown in Å.*

The putative nucleophile and acid/base were tentatively assigned by Hidaka *et al.* based on the observation that they superimpose (structurally) well with the catalytic residues of *Escherichia coli* β-galactosidase (GH2; also in clan GHA) and since the environments around the nucleophile residue are similar between *Thermus sp.* A4 and *E. coli* β-galactosidases.[102] The crystal structure of *Thermus sp. A4* (Figure 2.2) shows E312, which is conserved, to act as the nucleophile residue and E141, also conserved, to be the acid/base residue. Based on these observations and the sequence alignments, E295 is the corresponding nucleophile residue in YesZ and E145 is the corresponding acid/base residue.

## 2.2 Kinetic characterization of wild type and mutants E295A (acid/base) and E145A (nucleophile) of YesZ

### 2.2.1 Kinetic parameters for the hydrolysis of para-nitrophenyl β-D-galactoside (pNP-Gal) by wild type YesZ

Previously studied GH42 β-galactosidases had not been extensively kinetically characterized and as a result the catalytic residues have only been inferred based on the crystal structure of *Thermus sp*. A4 β-galactosidase.[102,104,108,109,111,112] In order to determine valid assay conditions for YesZ (and acid/base and nucleophile mutants), studies on the pH dependence and thermal stability of the enzyme were performed. *B. subtilis*, as previously stated, is found in non-extreme environments, and shows optimal activity at 37$^{o}$C (tested at 25$^{o}$C, 65$^{o}$C and 37$^{o}$C; other data not shown). A chromogenic substrate, pNP-Gal, was used for the continuous spectrophotometric measurement of the hydrolysis reaction of YesZ at 37$^{o}$ in varying buffers. All measurements were taken in triplicate and the experimental procedures for kinetic analyses are outlined in detail in Chapter 5. Values of $k_{cat}/K_M$ for the hydrolysis of pNP-Gal by wild type YesZ were determined as a function of pH within the stability range measured (the stability range for the enzyme was measured by incubating the enzyme at varying pHs for 15 minutes then removing an aliquot and measuring activity against pNP-Gal at pH 7.0; only pH values for which >90% activity was retained were used in the pH profile) for each enzyme: 4.0-9.0 for wild type. These values are plotted in Figure 2.3. The pH profile of the wild type enzyme follows a typical bell-shaped curve with optimal activity at pH 7.0. The $pK_a$ values for the general base and general acid ionisable groups in the free enzyme, obtained from the plot of $k_{cat}/K_M$ versus

pH, are $pK_{a1}$= 5.7 ± 0.1 and $pK_{a2}$= 7.9 ± 0.1. According to the double displacement mechanism the catalytic nucleophile is required to be deprotonated in the glycosylation step. The lower $pK_a$, $pK_{a1}$, which is determined by the acidic limb of the pH profile, therefore, likely reflects the protonation state of this residue. Similarly, $pK_{a2}$, which is determined from the basic limb of the pH profile, should reflect the protonation state of the acid/base residue.

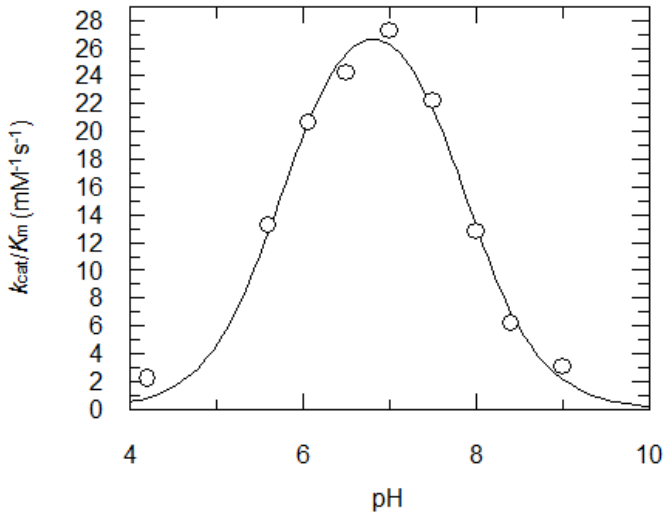



***Figure 2.3*** *pH Dependence of $k_{cat}/K_M$ for the hydrolysis of pNP-Gal by wild type YesZ*

Further studies on the stability of YesZ over time were carried out as some of the inactivation studies required longer incubation times at 37°C. A time course assay was performed to determine enzyme stability over a course of 3 hours. YesZ and its putative nucleophile (E145A) and acid/base (E295A) mutants were incubated at 37°C in 50 mM phosphate buffer at pH 7.0 and aliquots were taken every 15 minutes for the first hour and then every hour thereafter and assayed against pNP-Gal. All enzymes appear to be stable (>95% activity) under these conditions for the time period assayed.

Kinetic parameters for the hydrolysis of pNP-Gal by wild type YesZ, determined under the optimal conditions for YesZ (pH 7.0 and 37$^{o}$C) are listed in Table 2.1 below (in section 2.3.2).

### 2.2.2 Kinetic parameters for the hydrolysis of pNP-Gal by the E145A acid/base mutant and E295A (nucleophile) mutant

As previously mentioned, the stereochemistry of GH42 β-galactosidases was first determined to be retaining through $^{1}$H NMR spectroscopy of the hydrolysis of orthonitrophenyl galactoside (ONP-Gal) by *Thermus sp*. A4 β-galactosidase.[104] Additionally, the crystal structure obtained for that enzyme shows two carboxylate residues poised to act as a nucleophile and acid/base residue. However, no kinetic characterization of catalytically inactive mutants was performed; therefore the catalytic residues for this family had yet to be verified. The kinetic parameters for the acid/base (E145A) and nucleophile (E295A) mutants are summarized in Table 2.1

***Table 2.1*** *Kinetic parameters for the hydrolysis of pNPGal by wild type YesZ and its nucleophile (E295A) and acid/base (E145A) mutants (at pH 7.0 and 37$^{o}$C)*

| Enzyme | $k_{cat}$ ($s^{-1}$) | $K_M$ (mM) | $k_{cat}/K_M$ ($s^{-1}$ $mM^{-1}$) |
|---|---|---|---|
| **WT** | 81±4 | 3.0±0.2 | 27±2 |
| **E145A** | 0.13±0.01 | 0.047±0.03 | 2.8±0.1 |
| **E295A** | 0.008±0.01 | 3.2±0.5 | (2.5±0.2)* $10^{-3}$ |

Both the nucleophile and acid/base mutants were significantly less active than the wild type with an approximate 10000-fold decrease and 10-fold decrease in $k_{cat}/K_M$ respectively. The low

$K_M$ observed for the E145A mutant is most likely not a consequence of tighter reversible binding of the substrate, but rather a consequence of the accumulation of a covalent glycosyl-enzyme intermediate on the mutant. Such an accumulation occurs because the hydrolysis of the covalent glycosyl-enzyme intermediate is slowed substantially by removal of the base catalyst while formation of the intermediate remains fast for substrates with a good leaving group (such as the para-nitrophenyl moiety) that do not need acid catalytic assistance for departure. Since the expression for Km can be written as below; when $k_3 << k_2$ (deglycosylation is slower than glycosylation) an intermediate accumulates, and the $K_M$ value will necessarily be lowered.

$$K_M = \frac{k_{-1} \text{x } k_2}{k_1} \times \frac{k_3}{k_{2+k_3}}$$

Consequently the observation of a low Km represents supportive evidence for, though not yet full proof of, the role of E145 as the acid/base catalyst. The kinetic characterization of another GH42 enzyme, which came out shortly after this work was published, from *Alicyclobacillus acidocaldarius*, show similar results for the mutants of the nucleophile and acid/base residues (conserved by amino acid sequence), in which the mutants are significantly less active than wild type and in which the proposed acid/base residue also shows an expected drop in $K_M$.[103] Due to time constraints at the time of this study, a thorough kinetic analysis of the E145A and E295A mutants was not undertaken. However, a 30-fold increase in activity was observed upon addition of 100 mM sodium azide for the E295A (nucleophile) mutant. Surprisingly, no rescue was observed by Di Lauro *et al.* for the nucleophile mutant of *A. acidocaldarius* using azide but a 15-fold increase in activity was observed with 2 M formate. The acid/base mutant of *A.*

*acidocaldarius* was rescued (by 2-fold) with 200 mM azide (while no rescue was observed with formate).[103] $^1$H NMR analysis of a glycosyl azide product obtained from rescue indicated a β-configuration, as expected from the mutation of the acid/base residue of a retaining glycosidase. This provided further evidence for the assignment of the acid/base residue.

## 2.3 Inactivation of YesZ (and the acid/base mutant E145A) with 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-galactopyranoside (DNP2FGal)

The kinetic characterization of the wild type YesZ and mutants offers support for their proposed roles as acid/base and nucleophile; however, more direct evidence is needed for their unequivocal assignment. One method of identifying the catalytic nucleophile is through trapping of the covalent glycosyl-enzyme intermediate by decreasing the rate of deglycosylation relative to glycosylation thereby allowing for the accumulation of the intermediate.[70] The enzyme is inactivated (and the nucleophile is labelled) covalently by the 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-galactopyranoside reagent. As explained in section 1.4.2.3, the electron-withdrawing fluorine on C2 destabilizes the transition state for both the glycosylation and deglycosylation steps while incorporation of the 2,4-dinitrophenyl group speeds up the glycosylation step.[71,113] If the difference in rates between the steps is large enough then an accumulation of intermediate is observed and the enzyme is inactivated.

### 2.3.1 Time dependent inactivation of wild type YesZ and E154A

Kinetic parameters for inactivation were determined by pre-incubating 100 µl of the enzyme at 37 °C with DNP2FGal at a range of concentrations from 0.005 mM to 0.3 mM in a total volume of 140 µl. Aliquots were withdrawn at regular time intervals and added to 180 µL of pNPGal that had been pre-incubated in 50 mM phosphate buffer, pH 7.0 at 37 °C. The initial rates at each time point were plotted as a function of time to obtain time-dependent exponential decay curves from which $k_{obs}$ could be obtained for each inactivator concentration using the first-order rate equation: rate = $A_0e^{kt}$ + offset ($k = k_{i,obs}$), in GraFit 4.0. Values of the inactivation parameters $K_i$ and $k_i$ were obtained from the fits of a plot of $k_{obs}$ versus inactivator concentration to the equation: $k_{obs} = k_i \times [I]/(K_I + [I])$.[114]

Time-dependent inactivation of both the wild type YesZ and of the putative acid/base mutant (E145A) of YesZ was seen upon incubation of the enzymes with DNP2FGal as shown in Fig 2.4. Inactivation of the wild type enzyme was incomplete at lower inactivator concentrations as seen in Fig 2.4. This behaviour has been seen and characterized previously and arises when turnover of the intermediate via hydrolysis or transglycosylation occurs at rates close to those for its formation.[115] Complete inactivation was seen, however, for the acid/base mutant E145A, as expected. This change in behaviour arises because removal of the acid catalyst does not substantially affect glycosylation rate constants when a very good leaving group such as 2,4-dinitrophenol, that does not require protonic assistance for departure, is used.[53] However the absence of base catalysis in the mutant severely slows the hydrolysis of the glycosyl-enzyme intermediate, thus complete inactivation is observed. This is essentially the same phenomenon

as that which gives rise to low substrate $K_m$ values as described earlier. Replots of the pseudo-first-order rate constants for inactivation at each inactivator concentration for both enzymes yielded the following values for the inactivation parameters. Wild type YesZ: $k_i$ = (5.0 ± 0.3) · $10^{-3}$ $s^{-1}$, $K_I$ = 0.8 ±0.3 mM, $k_i/K_i$ = (6.3 ± 0.4) · $10^{-3}$ $s^{-1}$ $mM^{-1}$ and E145A: $k_i$ = (11.2 ± 0.7) · $10^{-3}$ $s^{-1}$, $K_I$ = 0.21 ± 0.02 mM, $k_i/K_I$ =(53 ± 3) · $10^{-3}$ $s^{-1}$ $M^{-1}$.

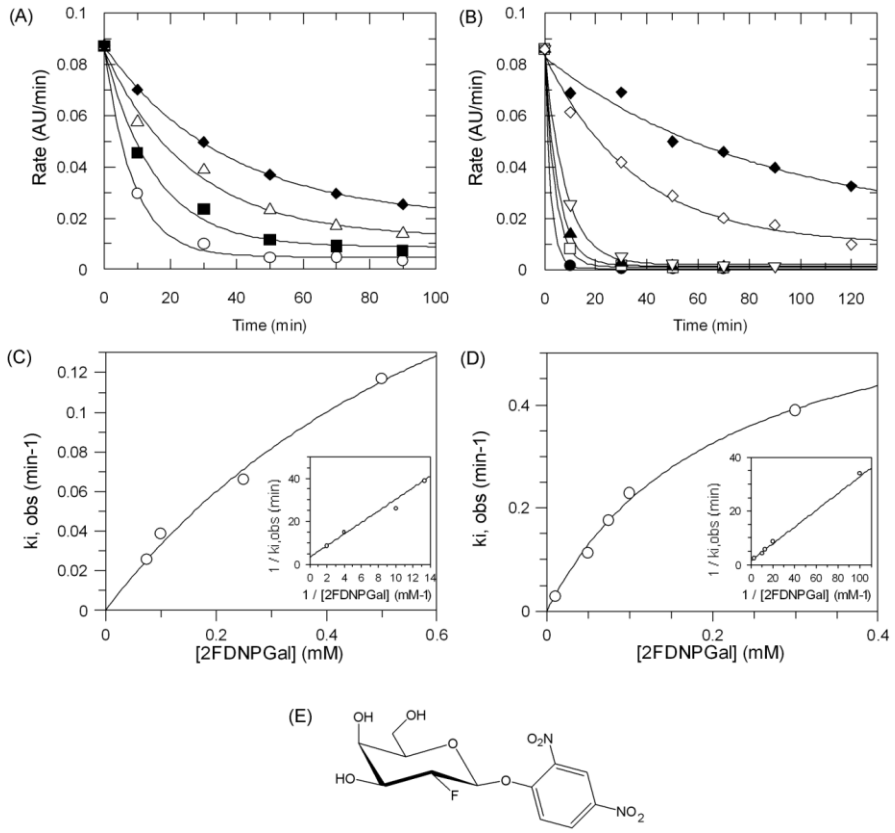



***Figure 2.4*** *Time dependent inactivation observed as a single exponential decay to a non-zero value when 2FDNPGal was incubated with (A) WT and (B) E145A mutant. Assays were performed at 37°C in 50mM phosphate buffer at pH 7.0. Inactivator (2FDNPGal) concentrations are as follows: (A) -◆- 0.075mM, -△-0.10mM, -■- 0.25mM, -○- 0.50mM; (B) -◆- 0.005mM, -◇- 0.01mM, -▽- 0.05mM, -▲- 0.075mM, -□- 0.10mM, -●- 0.30mM. A plot of $k_{i,obs}$, (obtained from fitting the curves in (A) and (B) to a single exponential decay with offset using Grafit 4.0) versus inactivator concentration gave Michaelis-Menten plots for (C) WT and (D) E145A from which $k_i$ and $K_I$ were obtained. The Lineweaver Burk plots are shown within. The structure of 2FDNPGal is illustrated in (e).*

A common experiment to confirm that inactivation is site-directed is through the demonstration of protection against inactivation by a competitive inhibitor. Isopropylthio-β-galactoside (IPTG), a competitive inhibitor of *E. coli* β-galactosidase, was tested as a potential competitive inhibitor of YesZ. An approximate $K_I$ value of 74 µM was calculated for IPTG from the Lineweaver-Burk plot of $1/V_o$ versus 1/[pNP-Gal] in the presence and absence of 100 µM IPTG (plot and formula found in Appendix B). Incubation of YesZ with 2F-DNPGal (0.25 mM), in the presence of 0.1 mM IPTGal, resulted in the reduction of the apparent inactivation constant ($k_{obs,I}$) from 0.10 $min^{-1}$ (without IPTGal) to 0.041 $min^{-1}$ ($k_{obs,protected}$) (Figure 2.5). The ratio of $k_{obs,protected}$ to $k_{obs,I}$ is 0.40 and the expected ratio is calculated from the following formula:

$$\frac{k_{obs,protected}}{k_{\ obs,inactivated}} = \frac{K_I + [I]}{K_I\left(1 + \frac{[CI]}{K_{CI}}\right) + \ 1}$$

where $K_I$ and [I] are of the inactivator and [CI] and $K_{CI}$ are of the competitive inhibitor. The expected calculated ratio from this formula is 0.45 which is in agreement with the observed ratio suggesting that inactivation by 2FDNPGal is site-directed. This further supports the hypothesis that the inactivation observed for YesZ is a result of the accumulation of a covalent glycosyl enzyme intermediate and will be further supported by mass spectrometric analysis of the inactivated enzyme.

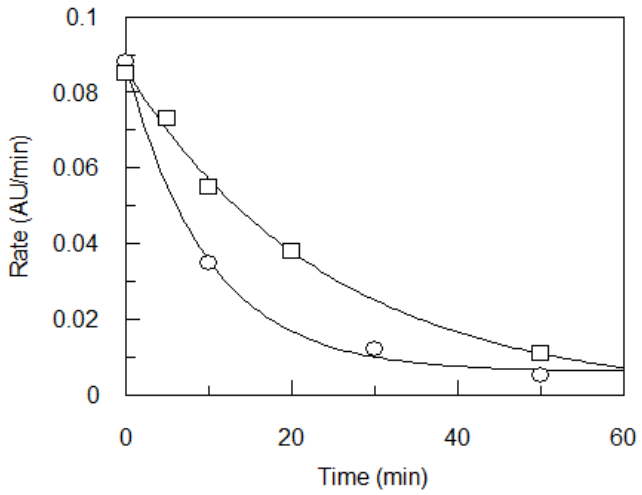



***Figure 2.5*** *Protection against inactivation of wild type YesZ with 0.2 mM 2FDNPGal given by 0 mM (○) and 0.1 mM (□) IPTGal.*

To provide further evidence that the inactivation is a result of the accumulation of a stable glycosyl-enzyme intermediate, the catalytic competence of the inactivated enzyme was studied. Wild type YesZ was incubated with 0.3 mM inactivator and excess inactivator was subsequently removed by dialysis then incubated in buffer at 37$^{o}$C and the return of activity associated with the regeneration of the free enzyme was monitored. The reactivation of YesZ in buffer is shown in Figure 2.6. The reactivation constant obtained from the fit of the data to the first order rate equation rate = A(1-e$^{kt}$) + C ($k$ = $k_{re}$) is $k_{re}$ = 0.06 h$^{-1}$ which corresponds to a half life $t_{1/2}$ = 11.6 h . This data supports the belief that the covalent glycosyl-enzyme intermediate formed was catalytically competent.

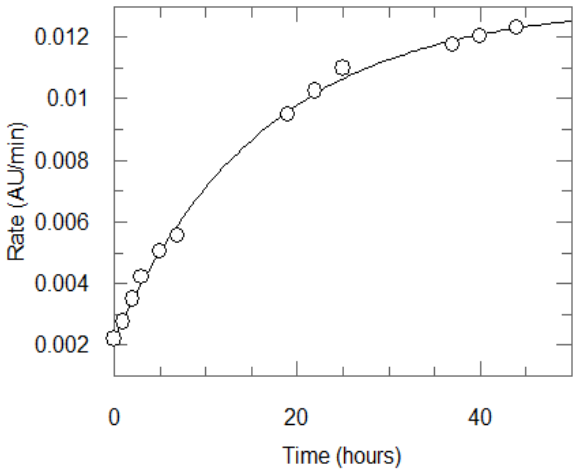



***Figure 2.6*** *Reactivation of 2-deoxy-2-fluoro-galactosyl galactosidase at 37°C in buffer (pH 7.0).*

## 2.4 Identification of the catalytic nucleophile of YesZ by labeling and mass spectrometric detection

### 2.4.1 Labeling of E154A by DNP2FGal and proteolysis

The inactivation of GH42 YesZ is likely due to the accumulation of a covalent glycosyl-enzyme intermediate, thereby allowing determination of the catalytic nucleophile through a comparison of the LC/ mass spectra of peptide digests of the inactivated protein with those of peptide digests of non-inactivated protein.[75] Since the inactivation of the acid/base mutant (E145A) by DNP2FGal proceeded essentially to completion from kinetic analyses, this enzyme form was selected for the peptide mapping studies as the covalent glycosyl-enzyme intermediate was relatively stable. A sample of the E145A (acid–base) mutant of YesZ (100 µl of 0.2 mg/mL in 50 mM sodium phosphate, pH 7.0) was incubated with 5 µL of DNP2FGal (3 mM in

water) for 2 h at 37 °C. Comparison of the mass spectra of the intact inactivated protein and control (unlabelled) protein will show whether the enzyme has been labelled. Subsequently both the intact protein and control sample (as above with 5 µL $H_2O$ instead of inhibitor) were subjected to proteolysis by diluting the samples with 10 µl of a pepsin solution (0.1 mg/ml in 50 mM phosphate buffer, pH 2.0) and 60 µL of 500 mM phosphate buffer, pH 2.0 and incubated for 3 h at room temperature to allow for the cleavage of the peptide bonds. Another enzyme commonly used to digest proteins is trypsin; this enzyme has a pH optimal of ~8 and an optimum temperature of 37$^{o}$C; use of this protease is not ideal since hydrolysis of the glycosyl ester linkage in the labelled peptide occurs under these conditions.

### 2.4.2 Electrospray ionization mass spectrometry

#### *2.4.2.1 LC/MS analysis of intact protein (E154A) (labelled and unlabelled)*

LC/MS analysis of the intact protein and peptide digests (for both labelled and unlabelled proteins) allows us to identify the residue covalently modified by mechanism-based inhibitors (DNP2FGal in this case). The stoichiometry of inactivation was first checked by measuring the intact protein mass by electrospray ionization mass spectrometry on a Q-TOF instrument. A mass, $[M + H]^+$, of 77 470 (±2) Da was measured for the unlabelled mutant, reasonably consistent with the predicted mass of 77 440 Da, while two peaks at 77 470 (±2) Da and 77 634 (±2) Da were observed for the labelled enzyme (Figure 2.7). Therefore, the mass difference of 164 between the two confirms the incorporation of a single 2-deoxy-2-fluorogalactosyl moiety

of mass 165. The presence of a peak corresponding to the mass of unlabelled enzyme in the inactivated sample indicates that the accumulation was probably incomplete.

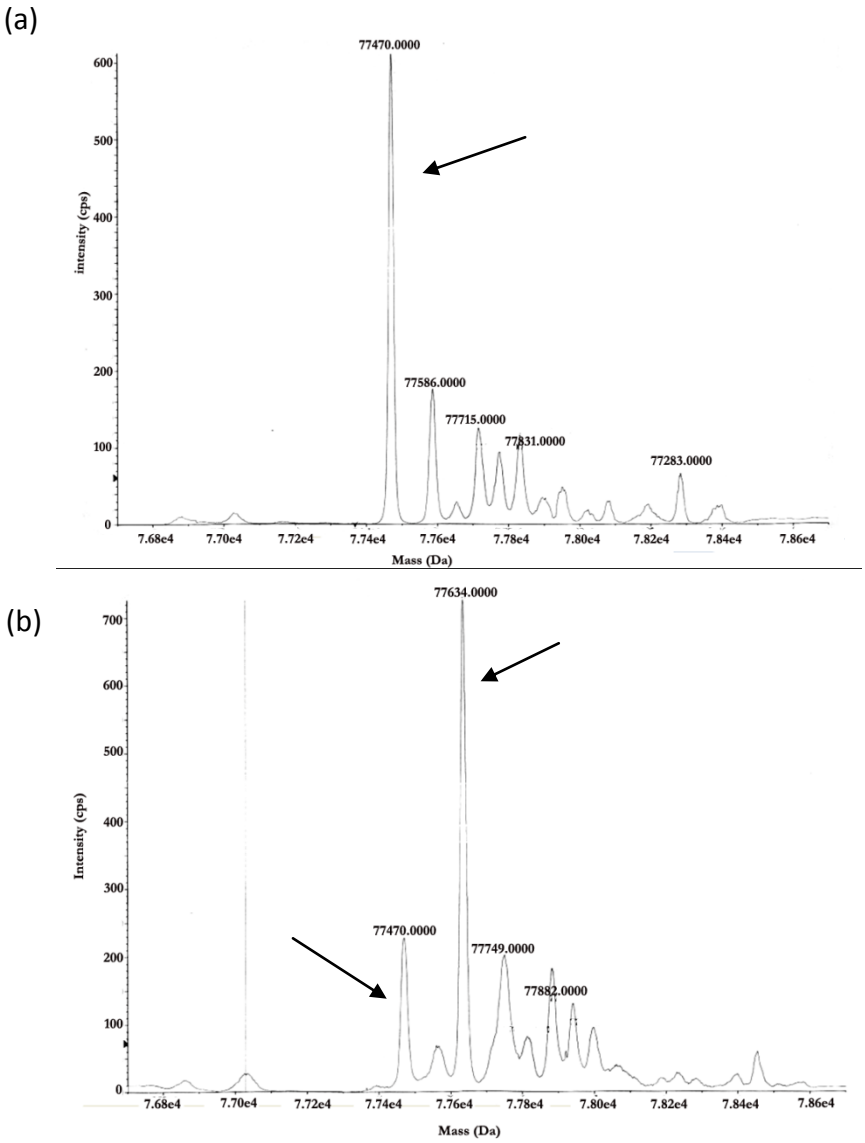



***Figure 2.7*** *(a) Detection of the 2-deoxy-2-fluorogalactosyl-label in intact protein by ESI-MS : Mass spectra of unlabelled enzyme. The arrow points to the unlabelled protein and (b) detection of the 2-deoxy-2-fluorogalactosyl-label in intact protein by ESI-MS : Mass spectra of labelled enzyme. Arrows point to the mass of the unlabelled protein at 77470 kDa and labelled protein at 77634 kDa. The label mass difference of 164 amu is detected between the two peaks.*

*2.4.2.2 LC-MS of peptic digests (labelled and unlabelled)*

As the LC-MS analysis of inactivated intact protein conclusively showed that it had been labelled, the peptide fragment containing the nucleophile residue would be expected to show a higher mass resulting from the covalently linked glycosyl moiety. The labelled peptide may be characterized by a different retention time on the RP-HPLC column when compared to the control unlabelled sample; this is used to screen the LC/MS data for the disappearance of a peptide ion present in the control. Additionally, as expected, the labelling of the peptide would cause a mass shift corresponding to the addition of the 2-fluoro-galactosyl moiety. Finally, the covalent glycosyl-enzyme ester bond can be broken by collision-induced fragmentation resulting in the loss of a neutral fluoro-glycosyl species. This allows the differentiation of glycosylated peptides from non-glycosylated peptides.

Samples of the inactivated E145A variant and non-inactivated control were subjected to pepsin digestion at pH 2.0 and the resultant peptic digests were, individually, separated on a reverse phase C18 column on an HPLC attached to an ESI-MS detector. When these mixtures were analyzed in separate experiments by scanning the total ion chromatogram in the normal LC/MS mode a large number of peaks corresponding to one or more peptide fragments were found in both cases. The peptide containing the 2-deoxy-2-fluorogalactosyl moiety was located by careful comparison of the two profiles since a mass difference of 165 Da would be expected. Comparison of the elution profiles of the digests of the control and inactivated E145A variants revealed essentially identical profiles except for a single peptide (*m*/*z* 1189.5) that was present in the labelled sample, but absent from the unlabelled enzyme (Fig 2.8). However, both profiles

contain a peak at *m*/*z* 1025.4 that corresponds, within error, to the equivalent unlabelled peptide (1189.5 minus the label of 165). The peptide of m/z 1189.5, which is likely the labelled peptide of interest, was then sequenced by collision-induced fragmentation. The major fragment peak observed is that of a singly charged ion of *m*/*z* 1025, which corresponds to the loss of the 2-fluorogalactosyl moiety as a neutral species, as explained previously, thereby confirming that this is indeed the peptide of interest.

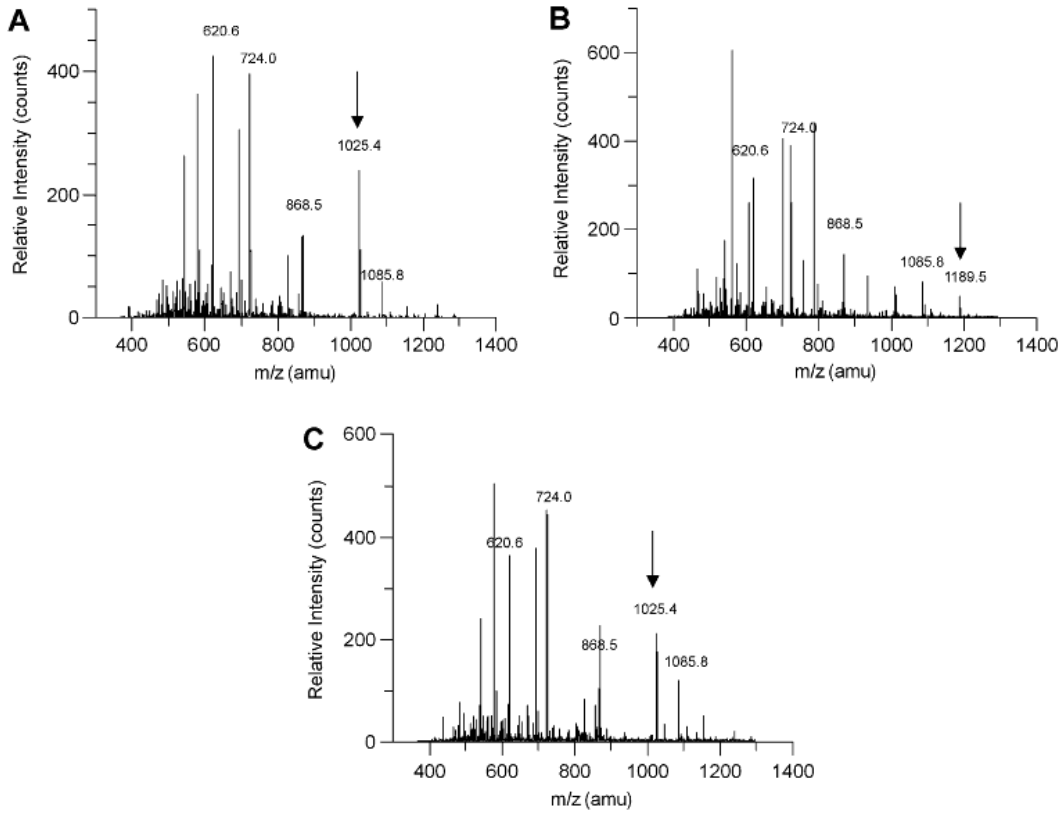
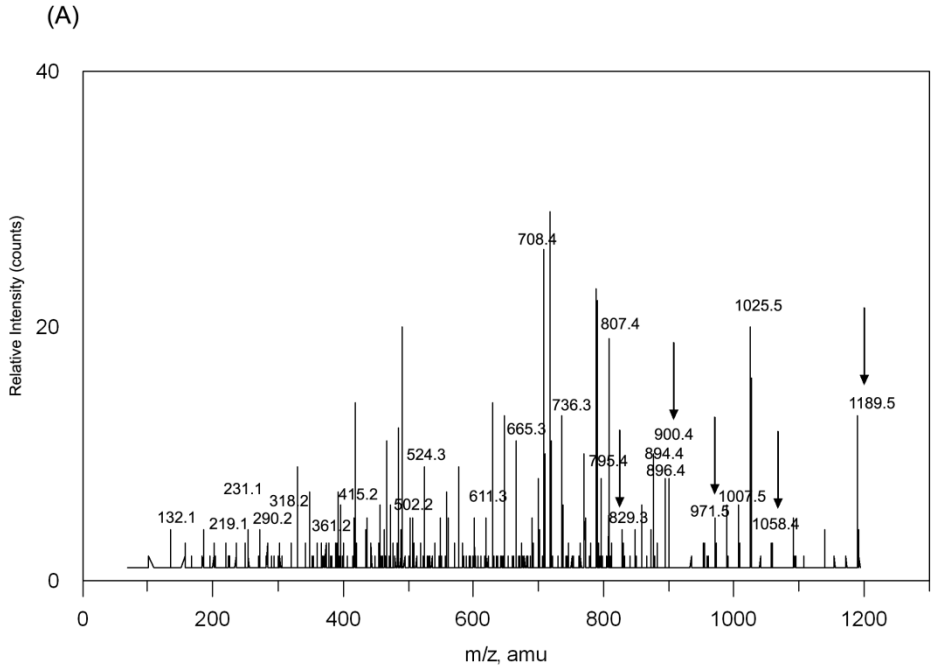



***Figure 2.8*** *Detection of the 2-deoxy-2-fluorogalactosyl-labelled peptide by ESI-MS mass spectra of (A) unlabelled enzyme digests eluted at 18 min and (B) and (C) labelled enzyme digests eluted at 16 and 15 min, respectively. Arrows point to the parent fragment of 1025 present in control and labelled digests and the labelled fragment at 1189 amu, which corresponds to the label mass difference of 164 amu.*

Further analysis of the fragmentation pattern of the parent peptide fragment of mass 1025 Da reveals the sequence ETSPSYAASL from the b-ion and y-ion peaks as shown in Fig 2.9. The spectrum also reveals four labelled peptide fragments of *m*/*z* 829, 900, 971, and 1058, which are consistent with peptides ETSPSY, ETSPSYA, ETSPSYAA, and ETSPSYAAS, each bearing the 2FGal moiety. Comparison of this sequence with that of YesZ derived from the genome sequence shows that this peptide corresponds to the region of sequence starting with glutamic acid 295. Since this sequence contains only one carboxylic acid and since carboxylic acids act as the enzymatic nucleophile in all known retaining glycosidases with the exception of sialidases, which use a tyrosine and Family 4 glycosidases, which use an entirely different mechanism these results suggest that, as predicted, the nucleophile in this enzyme is Glutamate 295.[59,63] Assignment of Glu295 as the catalytic nucleophile is consistent with the complete conservation of this residue in the sequences of Family 42 galactosidases, as was shown in Fig 2.1.




(B)

y-ions

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| unlabelled | 1025 | 896 | 795 | 708 | 611 | 524 | 361 | 290 | 219 | 132 |
| labelled | 1189 | | | | | | | | | |

E–T–S–P–S–Y–A–A–S–L

b-ions

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| unlabelled | | 231 | 318 | 415 | 502 | 665 | 736 | 807 | 894 | 1007 |
| labelled | | | | | | 829 | 900 | 971 | 1058 | |

***Fig. 2.9.*** *ESI-MS/MS analysis of the 2-deoxy-2-fluorogalactosyl-labelled peptide (A) MS/MS daughter-ion spectrum of the labelled peptide. Observed b-ion and y-ion fragments are shown in the spectra. (B) Fragmentation pattern of the peptide and assignment of the singly charged y- and b-ions from the fragment ion spectra. Masses shown in the pattern are those observed in the spectrum. Glutamate is shown as the predicted labelled residue.*

The function of E295 as the nucleophile has been unequivocally confirmed through the observation of a labelled peptide within which E295 is the only catalytically competent residue. The acid/base residue of this enzyme is tentatively labelled as E145, supported by the kinetic characterization of the alanine mutant in which a 10-fold decrease in $k_{cat}/K_M$ is observed along with a 64-fold decrease in $K_M$ suggesting the accumulation of a covalent glycosyl-enzyme intermediate—which was further confirmed through the labelling studies which used the E145A mutant to obtain the covalent trapped intermediate. Due to the time-sensitive nature of this work at the time, a detailed characterization of the acid/base residue was not performed. This would have included azide rescue and a comparison of the kinetics of the hydrolysis of substrates with different leaving group $pK_a$'s by YesZ and its E145A mutant. Azide rescue was seen by Di Lauro *et al.* on the acid/base residue of the GH42 β-galactosidase from *A. acidocaldarius.*[103] This acid/base mutant also yielded a glycosyl azide product with retention of stereochemistry. This stereochemistry is expected for azide rescue of the acid/base residue since the azide ion replaces the general base in the deglycosylation step. Inversion of stereochemistry is expected if the azide ion replaces the nucleophile residue in the glycosylation step. The observation of a galactosyl-azide product with retention of stereochemistry by Di Lauro *et al.* provided the first solid evidence for the acid/base residue for GH42 enzymes.[103]

## 2.5 Discussion and conclusions

The 2-fluorosugar inactivator approach combined with mass spectrometry is a powerful tool for identifying the catalytic nucleophile of retaining glycosidases. Reaction of 2FDNPGal with the E145A mutant of YesZ formed a covalent glycosyl-enzyme intermediate that was identified through MS analysis of peptic digests. This allowed for the unequivocal identification of Glu295 in the β-galactosidase YesZ from *Bacillus subtilis* as the catalytic nucleophile within the sequence ETSPSYAASL. Therefore, by analogy, this identifies the homologous residue in all members of Family GH42; these results are consistent with the conservation of this residue amongst GH42 enzymes and was also observed in the kinetic analysis of *A. acidocaldarus* by Di Lauro *et al*. Glu245 was tentatively assigned as the acid/base catalyst, as suggested by the conservation of this residue in the sequences of Family 42 galactosidases (Fig. 2.1), and by the crystallographic results (Fig. 2.2). This assignment is strengthened through kinetic analysis of the alanine mutant at that position, and the accumulation of a glycosyl-enzyme intermediate (which is characteristic of an acid/base mutant). The acid/base residue was unequivocally determined by Di Lauro *et al.* as E157 (in *A. acidocaldarus*) through $^{1}$H NMR analysis (shown to be a β) of the azide rescue product formed from the hydrolysis of 2-nitrophenyl β-galactoside by the E157G mutant. This completed the experimental assignment of catalytic nucleophiles in known retaining β-galactosidase families at the time and illustrates the fundamental methods by which catalytic residues in retaining glycosidases, which follow the double-displacement mechanism can be elucidated. Such methods will be initially used to determine the catalytic residues of other retaining glycosidases allowing for the manipulation of those enzymes to generate synthetic enzymes or more efficient glycosidases. Further this study reduced the

number of current Clan GH-A members for which unequivocal experimental assignment had not yet been achieved (at the time) to five. Future work could be performed on this family to generate glycosynthases for the formation of β-galacto-oligosaccharides.

# Chapter 3: Mechanistic Analysis of The Blood Group Antigen-Cleaving Endo-β-Galactosidase From *Clostridium perfringens*

## 3.1 Family GH 98 endo-β-galactosidases

### 3.1.1 Introduction

Blood transfusion is a vital treatment for a large range of medical cases, both acute and chronic, such as injuries (both external and internal), surgery, sickle cell anemia, thalassemia and neo-natal complications.[116–120] As a result, there is a high demand for donated blood or blood substitutes. Red blood cells contain a variety of antigens on their cell surface which may be protein, carbohydrate, glycoprotein or glycolipid-based.[121] The major blood group system is the ABO system which is based on carbohydrate antigens, and the secondary major blood group system is the RhD system based on protein antigens. The ABO blood groups vary by the presence of two sugar antigens on the surface of red blood cells: a terminal alpha-1,3-linked N-acetylgalactosamine (A-antigen) or an alpha-1,3-linked galactose (B-antigen), both of which are absent in the O-blood type (Figure 3.1).[7,8] All individuals have antibodies to the antigen that they lack, those in group A have Anti-B, those in group B have anti-A and those in group O have both antibodies; individuals in group AB do not have antibodies to either antigen and are thus universal acceptors. Transfusion of an incorrect blood type results in the lysis of the incompatible blood cells by agglutination of the red blood cells with antibodies, which can result in death. Shortages in donated blood supplies can be fatal in emergency situations, therefore development of a set of enzymes that can modify blood cells to either remove A or B antigens or display either or both antigens would overcome this severe problem. Within the ABO blood system, group O blood donors are in the highest demand because their red cells

contain no A or B antigens allowing them to be transfused into patients with any of the A, B, AB, or O blood types (Figure 3.2).

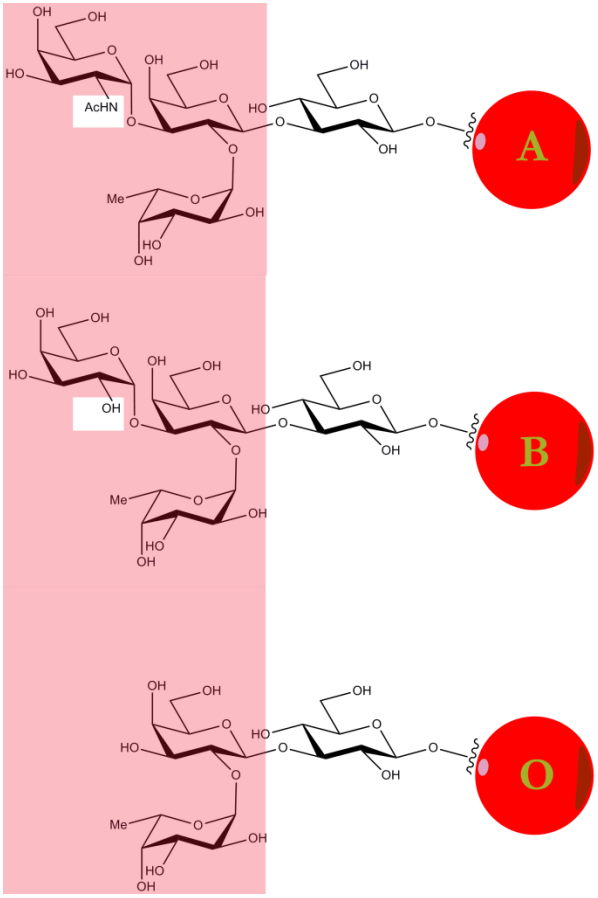



***Figure 3.1:*** *Structure of the A, B and O blood group antigens; the red area indicates the antigen structure: a trisaccharide for types A and B and a disaccharide for type O. The white rectangle highlights the difference between A and B antigens*

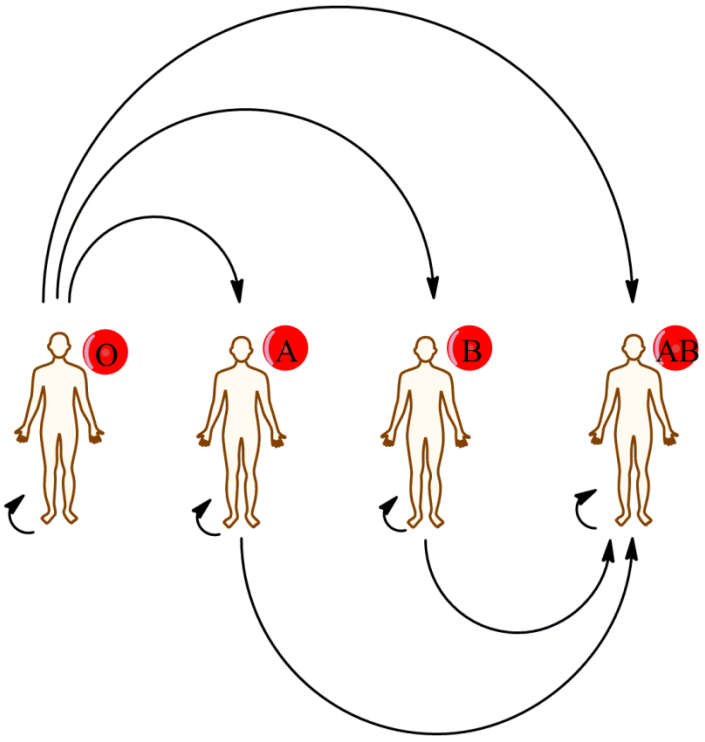



***Figure 3.2:*** *Donor and recipient blood type compatibility based on the ABO antigen system; the direction of the arrow indicates direction of blood donation: O-types are "universal" ABO donors and AB-types are "universal ABO acceptors; each type can donate to individuals of the same type.*

Currently proposed enzymatic methods to convert A or B blood to O involve the use of N-acetylgalactosaminidases, for conversion of A to O type, or α-galactosidases for the conversion of B blood to O (Figure 3.3).[97] In the first attempts at this approach, the α-galactosidase from coffee beans was used to convert type B blood to type O.[96] This approach required large amounts of enzyme (1-2 g of enzyme per unit of group B RBC's at low pH) and had to be performed at lower than optimal pH values due to the pH optimum of the enzyme; making this

approach highly inefficient. Conversion of type A blood has been significantly more difficult as this antigen exists in two major subgroups, $A_1$ and $A_2$, which differ in the density of antigens on the cell surface (with $A_1$ being 5 times more dense), the repeating A-antigen structure in $A_2$ cells and in the identities of the internal linkages—Gal-β(1-4)-GlcNAc for $A_2$ subgroups and Gal-β(1-3)-GalNAc for $A_1$ subgroups (Figure 3.4). Conversion of $A_2$ subtype to O was performed using chicken liver α-N-acetylgalactosaminidase under similar conditions to those used with the coffee bean galactosidase, also rendering this method inefficient. More recently Liu *et al.* identified two enzymes, a GH109 α-N-acetylgalactosaminidase from *Elizabethkingia meningosepticum* and a GH110 α-galactosidase from *Bacteriodes fragilis* that are capable of more efficient cleavage of $A_2$ and $A_1$ types (GH109) and B types (GH110) under physiological pHs (6.8).[97] While methods currently exist for the cleavage of individual A or B antigens at a reasonable, but still not apparently cost effective efficiency, an enzyme capable of the liberation of both A and B antigens, and potentially of the addition of these antigens back onto RBCs, would be unique and has the potential to be significantly more efficient.

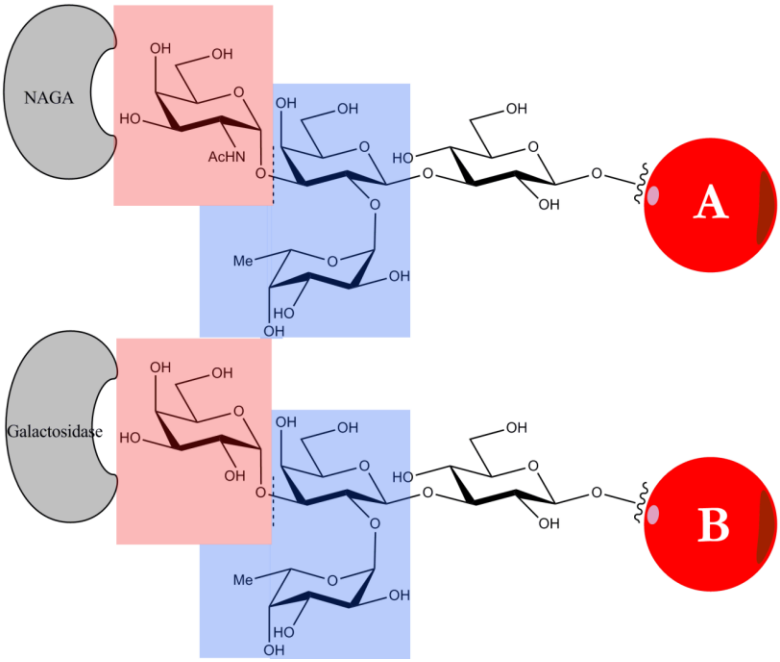



***Figure 3.3:*** *Current methods for conversion of A and B RBC's to "universal" O type RBCs; NAGA is the N-acetylgalactosaminidase from GH110 while galactosidase refers to the α-galactosidase from GH109. The red rectangles indicate the sugar being hydrolyzed and the blue rectangle indicates the O-antigen that is left after hydrolysis resulting in "universal" O-type blood*

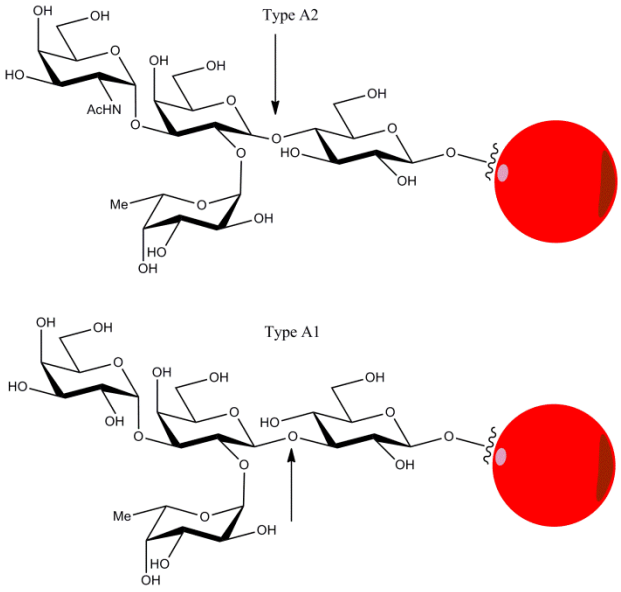



***Figure 3.4:*** *Structure of Type I and Type II A-antigens*

EABase, an endo-β-galactosidase capable of liberating the terminal trisaccharides from glycoconjugates constituting both the A and B blood-group antigens, was initially isolated by Anderson *et al* as a contaminant within commercial preparations of sialidases from *C. perfringens.*[98] *Clostridium perfringens* is a gram-positive pathogenic bacterium commonly found in soil, insects and the intestines of mammals and humans.[122–126] *C. perfringens* is known to cause food poisoning and gas gangrene; the latter of which is caused by the release of hydrolases and enterotoxins by the bacteria to destroy host tissues.[122,125,127,128] Due to the lack of significant amino acid sequence similarity to the other 97 CAZY glycoside hydrolase families identified at that time the authors assigned EABase to a new glycoside hydrolase family, GH98.[98] The mechanism of enzymes within this family, and the identities of the catalytic residues had not been determined experimentally, such information was needed to contribute to the understanding of this class of endo-β-galactosidases. Additionally, the unique substrate specificity of EABase makes it an attractive target for use in the hydrolysis or transfer of complex and biologically relevant sugars onto cell surfaces, either with the native enzyme or through the use of mutant forms (glycosynthases) engineered for this purpose. This would bring us closer to the efficient conversion of blood to “universal blood”, potentially overcoming some of the rejection issues that transplant patients undergo and possibly allowing for the synthesis of both natural and unique antigen structures on RBCs and other cells.[14,17,129,130] An understanding of the mechanism will not only guide the possible development of glycosynthases but also probe the basis for the directed evolution of this enzyme into a more

efficient A/B-antigen cleaver. Additionally, it would reduce the need for specific enzymes (α-N-acetylgalactosaminidase/ α-galactosidase) for the individual cleavage of A and B blood groups.

This project aims to unambiguously determine the mechanism of EABase from *C. perfringens*, and by extension other enzymes within glycoside hydrolase family 98, and use these enzymes in the modification of blood cells. Specifically, we will attempt to: identify the stereochemical outcome of the enzymatic hydrolysis using NMR spectroscopy, develop an efficient assay for kinetic analyses of the enzyme, identify the catalytic residues through mutagenesis and kinetic characterization, and use EABase to modify blood cell surfaces. These results will help to clarify the mechanism by which this family of hydrolases operates. The identification of catalytic residues and substrate specificity will allow for the development of synthetic or evolved enzymes for modifying red blood cells.

### 3.1.2 Previous work done on GH98

Anderson *et al.* first identified the enzyme as a contamination in a commercial preparation of clostridial sialidases.[98] Recombinant EABase was cloned and expressed from *C. perfringens* and shown to have activity against A and B erythrocytes as well as A and B trisaccharide-containing glycoconjugates. To demonstrate these abilities FACS was used to sort cells that had agglutinated with anti-A or anti-B antibodies. Agglutination was shown to be effectively eliminated in both A and B-antigen-containing cells treated with EABase. Products of the hydrolysis reaction and their linkages were determined using NMR spectroscopy, though their

methods did not allow the stereochemical outcome of the reaction to be determined due to rapid mutarotation. Additionally, their studies, using glycoconjugates with different core linkages, gave insight into the substrate specificity of EABase. Specifically, EABase was shown to preferentially cleave the endo-β-galactosyl linkage of the type 2 core chain (Gal-β(1,4)-GlcNAc; present in the $A_2$ subgroup ) over the type 1 core chain (Gal-β(1,3)-GlcNAc; present in the $A_1$ subgroup). Additionally, EABase was shown to not be able to hydrolyse substrates lacking an L-fucose linked α(1, 2) to the penultimate-galactosyl residue. While these studies have given considerable insight into the substrate requirement of EABase, the mechanism had not been investigated at all.

The only work present at the time that gave any suggestion for the mechanism of EABase was a paper by Rigden which had predicted EABase to act via a retaining mechanism on the basis of amino acid sequence information and predicted fold recognition.[131] His predicted secondary structure model showed similarities in fold to families 1, 2, 5, 10, and 17, all of which are retaining GH families. Through this, and with no experimental data, the catalytic nucleophile and acid/base residues in the case of a retaining mechanism were suggested to be E354, D467 and D429.

### 3.1.3 Sequence alignment and potential catalytic residues

Alignment of the amino acid sequences of enzymes that are functionally related or in the same family is used to determine the identity of highly conserved residues such as catalytically relevant residues. In most glycosidase families the nucleophile and acid/base (retaining mechanism) or acid and base residues (inverting mechanism) are fully conserved. Sequence alignment of the full gene sequence of EABase with those of other Family 98 hydrolases using BLAST (basic local alignment search tool) reveals several conserved aspartate and glutamate residues within the predicted catalytic domain region (Figure 3.5). Aspartate and glutamate have been shown to act as the catalytic residues in the majority of glycosidase mechanisms. From Figure 3.5, it can be seen that three of the residues, E354, D453 and E506 are each surrounded by three to four additional conserved residues. As this region is conserved, these residues might be the most likely to serve as catalytic residues. E467 is not fully conserved, which suggests that it likely does not play a role in catalysis; whereas D429 is conserved, but the region flanking the residue is not, possibly indicating that this residue might not play a catalytic role. These five residues, E354, D429, D453, E467, and E506, were chosen for mutation to alanine based on a combination of Rigden's predictions and additionally, based on the residues in the most conserved regions.

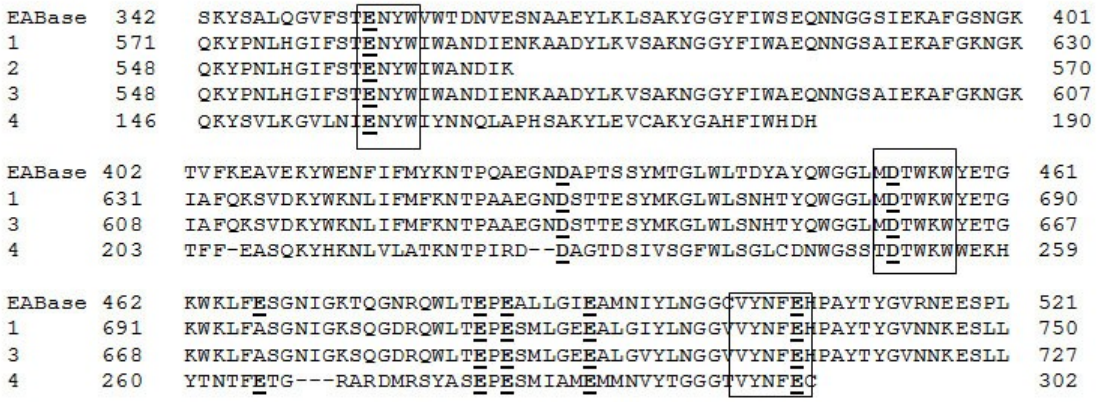



***Figure 3.5:*** *Sequence homology of GH 98 enzymes. Putative catalytic aspartates and glutamates are highlighted in bold and underlined. Boxes surround regions where 4 or more amino acids are conserved adjacent to putative catalytic residues. Organisms are as follows: 1: [Streptococcus pneumoniae CGSP14]; 2: [Streptococcus pneumoniae G54]; 3: [Streptococcus pneumoniae SP3-BS71]; 4: [Streptococcus pneumoniae TIGR4]*

## 3.2 Design and synthesis of a chromogenic substrate for stereochemical and kinetic analyses

### 3.2.1 Enzymatic synthesis of β-DNP-Gal-α-(1, 3)-Gal

In order to develop an assay that could be directly monitored and rapidly obtain clean NMR spectra of the stereochemical outcome of the reaction, two potential chromogenic substrates for the kinetic analysis of EABase were synthesized.

The first was β-DNP-Gal-α-(1, 3)-Gal, synthesized enzymatically by the incubation of donor UDPGal and acceptor DNP-Gal (synthesized according to previously published protocols) [132]

with bovine α-(1, 3)-galactosyltransferase.[133,134] The enzymatic reaction produced only the desired disaccharide product which was purified away from starting materials by C18 reverse-phase Sep-Pak column chromatography. While Anderson *et al.* had shown that the 2-fucosyl residue was important for hydrolysis by EABase, it was hoped that the presence of an activated leaving group such as the 2, 4-dinitrophenolate moiety, would overcome that need.

When purified β-DNP-Gal-α-(1, 3)-Gal was incubated with EABase at concentrations up to 150 μg/mL no reaction was observed by TLC, UV-vis spectrophotometry or by mass spectrometry over a temperature range of 25-37$^{o}$C, pH range of 5-7.5 and time range of 0-48 hours. This result further validated the requirement for a 2-fucosyl residue for activity. As a result a substrate containing the 2'-fucose was necessary. At the time of the proposal, no known fucosyltransferases were stable enough in-vitro to synthesize sufficient amounts of the desired product enzymatically and the chemical synthesis of this trisaccharide is challenging—so an alternative was sought.

### 3.2.2 Use of engineered *E. coli* cells to produce A-trisaccharide and A-pentasaccharide (Randriantsoa *et al.*)

Drouillard *et al.* illustrated a method by which complex oligosaccharides could be synthesized inside *E. coli* cells that had been engineered to express the enzymes necessary to form the desired linkages.[135] As illustrated in Scheme 3.1(a) (Section 3.2.3), M. Randriantsoa engineered *E. coli* cells to express both an α-(1,3)-N-acetylgalactosaminyltransferase and an α-(1,2)-

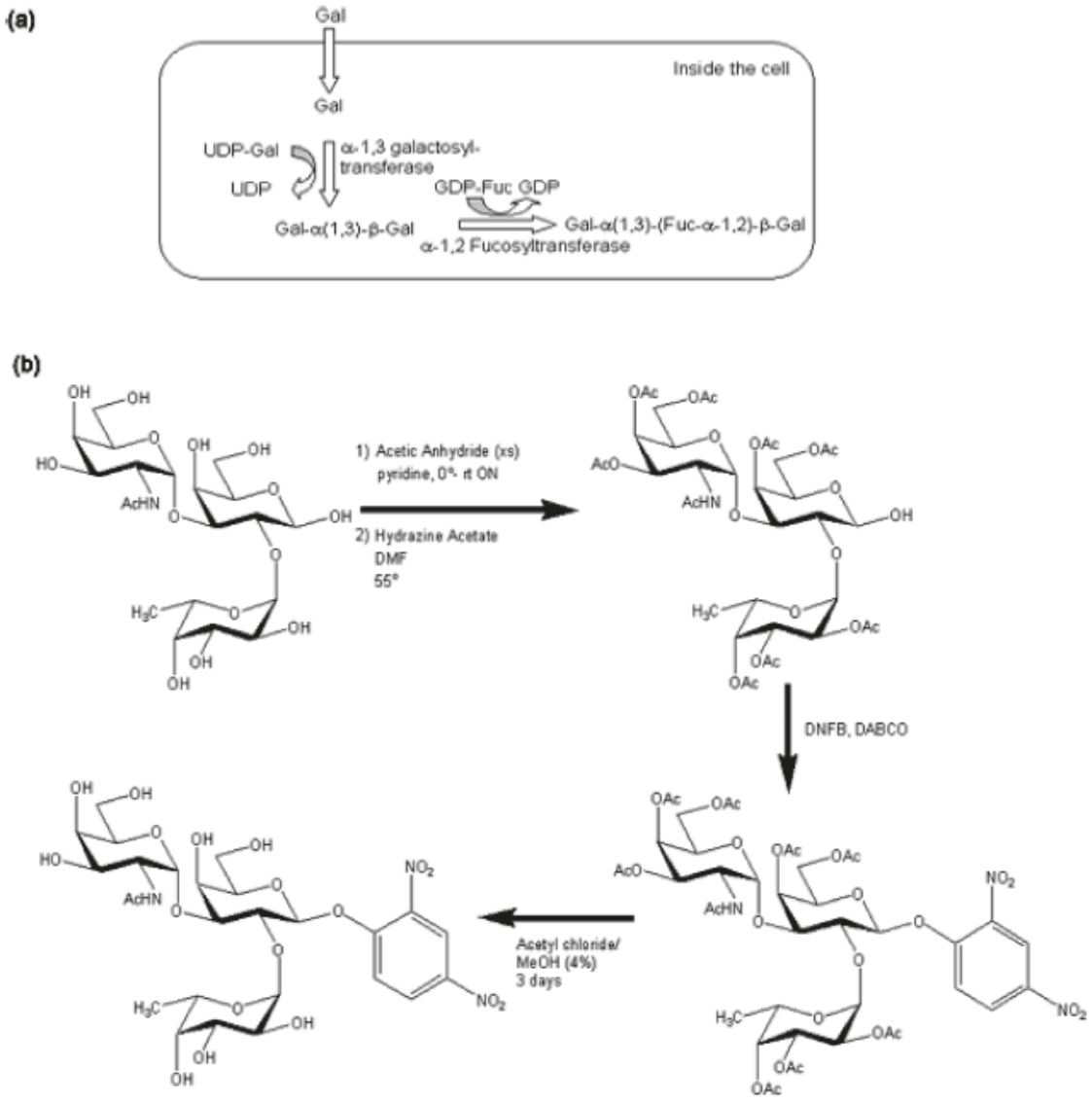
fucosyltransferase, allowing the synthesis of a suitable substrate for EABase. As mentioned previously, in-vitro oligosaccharide synthesis using glycosyltransferases is limited in this case by the stability of known α-(1,2)-fucosyltransferases, which makes it difficult to synthesize reasonable amounts of A/B-trisaccharides for use in the characterization of EABase. However, when expressed within the engineered *E. coli* cells, the fucosyltransferase is stable and active and the cells are capable of gram-scale synthesis of complex oligosaccharides. This therefore overcomes the limitations of *in vitro* enzymatic or chemical synthetic methods.

Initial attempts to synthesize the DNP-trisaccharide focused on the feeding of DNP-Gal itself to these metabolically engineered *E. coli* cells in the hopes that the other sugars would be transferred onto it. Unfortunately this did not prove successful as hydrolytic release of the 2,4-dinitrophenol moiety likely caused cell death.[136–138] As the only available alternative approach, a sample of A-trisaccharide was synthesized in the *in vivo* cell system and subsequently chemically converted to the 2,4-dinitrophenyl glycoside as shown in Scheme 3.1(b) (Section 3.2.3).[132,139]

### 3.2.3 Chemical synthesis of DNP-A-trisaccharide

The synthesis of DNP-A-trisaccharide used approaches that were previously described for the synthesis of other glycosides containing the DNP leaving group as shown in Scheme 3.1(b). In short A-trisaccharide was acetylated in acetic anhydride/pyridine then deacetylated at the anomeric position using hydrazine acetate in dry DMF. In the next step 2,4-dinitrofluorobenzene and 1,2-diazabicyclo[2.2.2]octane (DABCO) were added to the protected trisaccharide hemiacetal to give the protected DNP-β-A-trisaccharide product. This was then

deprotected using distilled acetyl chloride in methanol and purified over a reverse phase C18 Sep-Pak column. Detailed information and NMR data are provided in Chapter 5.




*Scheme 3.1: Schematic of synthetic route toward a chromogenic substrate for kinetic analysis of EABase: (a) cell-based synthesis of A-trisaccharide and (b) chemical synthesis of DNP-A-trisaccharide.*

## 3.3 Cloning and mutagenesis

### 3.3.1 Cloning and expression of wild type EABase

To facilitate the characterization of EABase from *C. perfringens* with the goals of obtaining large quantities of protein readily and generating variants of this enzyme easily, the gene encoding the enzyme was cloned into pET21-a and expressed in *E. coli* as follows. Primers containing the coding region of the gene starting at the N-terminal peptide and containing the NheI and XhoI restriction sites were used for PCR amplification of the *C. perfringens* genome (ATCC 10543), yielding a single product of ~2.3 kbp which was fully sequenced and shown to contain the full length open-reading frame. The EABase gene product was subcloned into pET21-a vector using the XhoI and NheI sites and transformed into *E. coli* BL21(DE3) cells. Plasmids were extracted and subsequently fully sequenced from colonies grown overnight. These were all shown to contain the full length gene encoding EABase, starting with the N-terminal peptide.

A single colony was induced in 2 L of LB broth as described in Experimental Procedures, and the enzyme containing a His6 tag was purified by nickel-affinity chromatography and Mono-Q ion-exchange. The gene in this plasmid consisted of a 2325 bp open reading frame encoding a protein consisting of 774 amino acids including the linkers and His6 tag; the calculated theoretical molecular weight of this protein is Mr = 88147.3 Da. The purified enzyme was resolved by SDS-PAGE and migrated as a band with an approximate MW of 88 kDa. The approximate yield of protein was 15 mg protein from 1 L of LB medium.

### 3.3.2 Site directed mutagenesis of conserved catalytic residues: E354, D429, D453, E467, and E506

Site directed mutagenesis was used to generate alanine mutants of putative catalytic residues. Mutagenesis using the QuikChange method yielded colonies for the E354A, E467A, and D429A mutants. DNA sequencing results showed that some of the colonies contained the single desired mutation (a few colonies contained undigested wild type plasmid). No colonies were obtained for the D453A and E506A mutations using the QuikChange method despite numerous attempts, presumably because of unwanted secondary structure formation in the polynucleotide. For these mutants, the 4-primer method was used. Sequencing of selected colonies showed that they contained the single desired mutations.

### 3.3.3 CD spectra of all mutants illustrating proper folding

The conformational integrity of the mutant proteins was verified by circular dichroism spectroscopy. The CD spectrum of a protein reflects its three dimensional structure; characteristic CD spectra are obtained for each of the three basic secondary structures (helices, sheets, and coils) and characteristic spectra are also obtained for unfolded peptides [140,141]. The CD signal for an unfolded peptide is around zero in the 210-220 nm range, corresponding to the absorbance of the peptide bond; this allows for the monitoring of unfolding/folding. As seen in Figure 3.6 CD spectra for all mutants are essentially identical to those of the wild type in all cases, confirming that all the mutants were folded correctly.

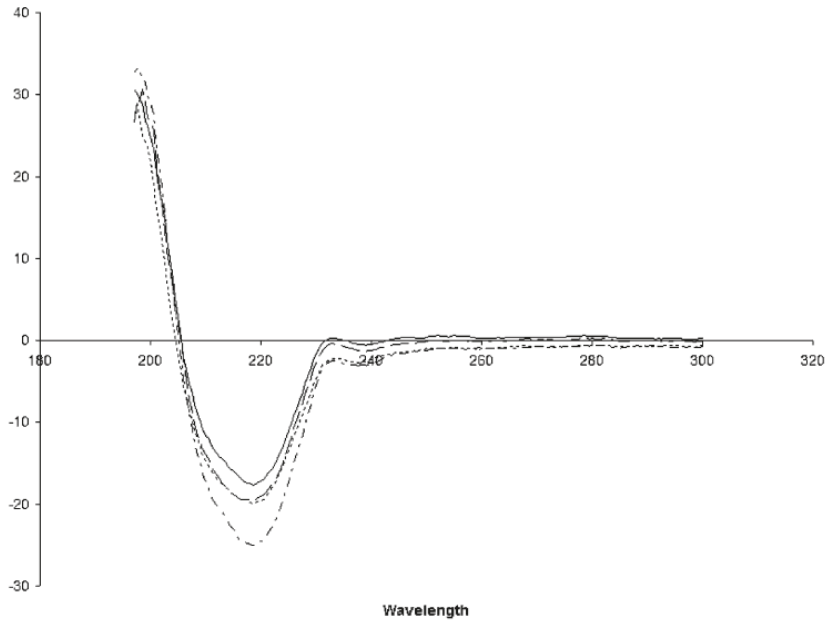



***Figure 3.6****: CD Spectra of wild type enzyme (- - -), E354A (— ), D453A(— —), andE506A(— - —). Similar concentrations of protein were used to obtain each spectrum (between 1 and 3 μg).*

## 3.4 Stereochemical outcome of EABase catalyzed hydrolysis of DNP-A-trisaccharide

### 3.4.1 Attempts at stereochemical determination using methanolysis

The classification of a glycosidase's catalytic mechanism as inverting or retaining is generally experimentally verified by the stereochemical outcome of the glycosidase-catalyzed reaction through direct $^{1}$H NMR analysis of the product(s) from the enzyme-catalyzed reaction. Two methods can be used; the first method is the real time monitoring of the chemical shift and determining the coupling constant of the anomeric proton with H2 of the initial product formed as the reaction progresses. Mutarotation of the product sugar is a major limiting factor of the second method but can often be overcome by the addition of sufficient enzyme that the

majority of starting material is converted to product in a short time in a solution that is at neutral pH and free of buffers that are known to increase the rate of mutarotation.[142] The second method is to carry out a methanolysis reaction where the product, a methyl glycoside, cannot undergo mutarotation. The stereochemistry of the glycoside so formed serves as a proxy for the first formal hydrolytic product--since it is highly unlikely that a different mechanism would be followed.[143] However, this approach will only work for retaining glycosides, since transglycosylation is not possible for inverters.

As no convenient low molecular weight substrate for real time NMR monitoring of EABase was initially available at the time of these studies, methanolysis was the method chosen. Blood group A+ porcine gastric mucin is a commercially available substrate for EABase. Since blood group A+ porcine gastric mucin (A+ PGM) is not highly soluble and is inhomogeneous, real time NMR measurements were not feasible. $A^{+}$PGM was briefly tested as a potential substrate for the real time NMR experiment; however spectra obtained did not show clear signals for the anomeric proton (or other protons) from the A-trisaccharide liberated, likely as a result of low concentrations of the product present. Other limitations included the timescale of the reaction (typically on the scale of hours) and the large quantities of $A^{+}$PGM required to liberate sufficient A-tri. Unfortunately, attempts to generate an O-methyl glycoside by reaction of A+ PGM and EABase in 5 M methanol were unsuccessful. TLC analysis and mass spectra showed the presence of the hydrolyzed A-trisaccharide after 1, 2, 6 and 12 hours of incubation but not of the O-methyl glycoside of the trisaccharide. Thus the stereochemical outcome of the reaction could not be determined in this way. The presence of A-trisaccharide in only those reactions containing enzyme and A+PGM and not in those that did not contain EABase, suggests that it is

unlikely that EABase was unstable in the presence of 5 M methanol as enzyme-catalyzed hydrolysis was occurring. One possible alternative explanation of these findings was that EABase is an inverting enzyme, since inverting glycosidases generally cannot catalyze methanolysis. However, such a null result was certainly not definitive evidence.

### 3.4.2 NMR studies using A+PGM and DNP-A-tri

Once we had the synthetic substrate DNP-β-A-trisaccharide on hand, $^{1}$H NMR spectroscopy was used to determine the anomeric stereochemistry of the initially formed products from EABase catalyzed hydrolysis. Figure 3.7 (a) shows a series of $^{1}$H-NMR spectra following the time course of the reaction from the point prior to enzyme addition (t = 0) to 68 min after the addition of enzyme. An expansion of the anomeric proton region of the $^{1}$H-NMR spectrum of the DNP-β-A-trisaccharide substrate is shown at t= 0 min. The doublet at δ 5.30 ppm (J=7.7 Hz) corresponds to the anomeric proton of the β-galactosyl moiety, and the large peak at δ 4.68 ppm arises from residual HOD. Spectra taken at time intervals reveal a steady decrease in the intensity of the peak from the substrate anomeric proton δ 5.30 ppm (J=7.7 Hz) as enzyme catalyzed hydrolysis occurs. This is accompanied by the gradual appearance of a peak at δ 5.18 ppm, J = 3.2 Hz due to the anomeric proton of the initially formed product. This J value is in the range of H1-H2 coupling constants typically found for an α-configured sugar. Over this time-course, no signal was observed in the region between δ 4.65-4.50 ppm where the anomeric proton from the β-configured product would be expected to show. However, $^{1}$H- NMR analysis of the same sample, using the same water suppression protocol, after overnight equilibration at ambient temperature, clearly revealed a resonance at δ 4.52 ppm (J=7.3 Hz) which is in the

range of H1-H2 coupling constants typically found for the β-configured product (Figure 3.7 (a)). These results confirm that the anomeric proton peak of the β-configured trisaccharide was not hidden under the residual HOD peak or as a result of the water suppression protocol. Mutarotation of the α-configured trisaccharide to the β-configured trisaccharide appears to be very slow at ambient temperatures and under the conditions used for this $^{1}$H-NMR experiment. As further confirmation that the α-anomer is the first formed product, a linear plot of the intensity of the anomeric peaks of the product and substrate as a function of time (Figure 3.8) shows that the rate of substrate depletion equals the rate of formation of the α-anomer of the trisaccharide product (slopes are identical: 0.015/min), showing that no other products were initially formed. These results show that the product released by EABase is the α-configured trisaccharide and thus that this enzyme uses an inverting mechanism. These results are contrary to the initial predictions by Rigden.[131]

Both the Sp4GH98 and Sp3GH98 endo-β-galactosidases were also shown by Higgins *et al.* to follow an inverting mechanism by $^{1}$H-NMR time course studies (similar to ones performed in this study for Sp4GH98 only using the Lewis$^{Y}$ tetrasaccharide) and by x-ray crystallographic studies in a publication that appeared essentially simultaneously with our own.[144] NMR studies showed that the α-anomer of the H-disaccharide product formed first, supporting the inverting mechanism for GH98 enzymes that had been found in our studies. The Sp4GH98 and Sp3GH98 enzyme active sites were shown to have identical arrangements of conserved active site residues to each other, and also showed that the equivalent (by sequence comparison) residues to D453 and E506 in those enzymes coordinated a water molecule that lies 3.1 A beneath the C1 residue of galactose (Figure 3.9). The positioning of the water in their crystal structures is

consistent with an inverting mechanism where both the glutamate and aspartate residues act as general bases. Based on our NMR studies and the NMR and X-ray crystallographic studies performed by Higgins *et al.* GH98 enzymes use an inverting mechanism to hydrolyze substrates.

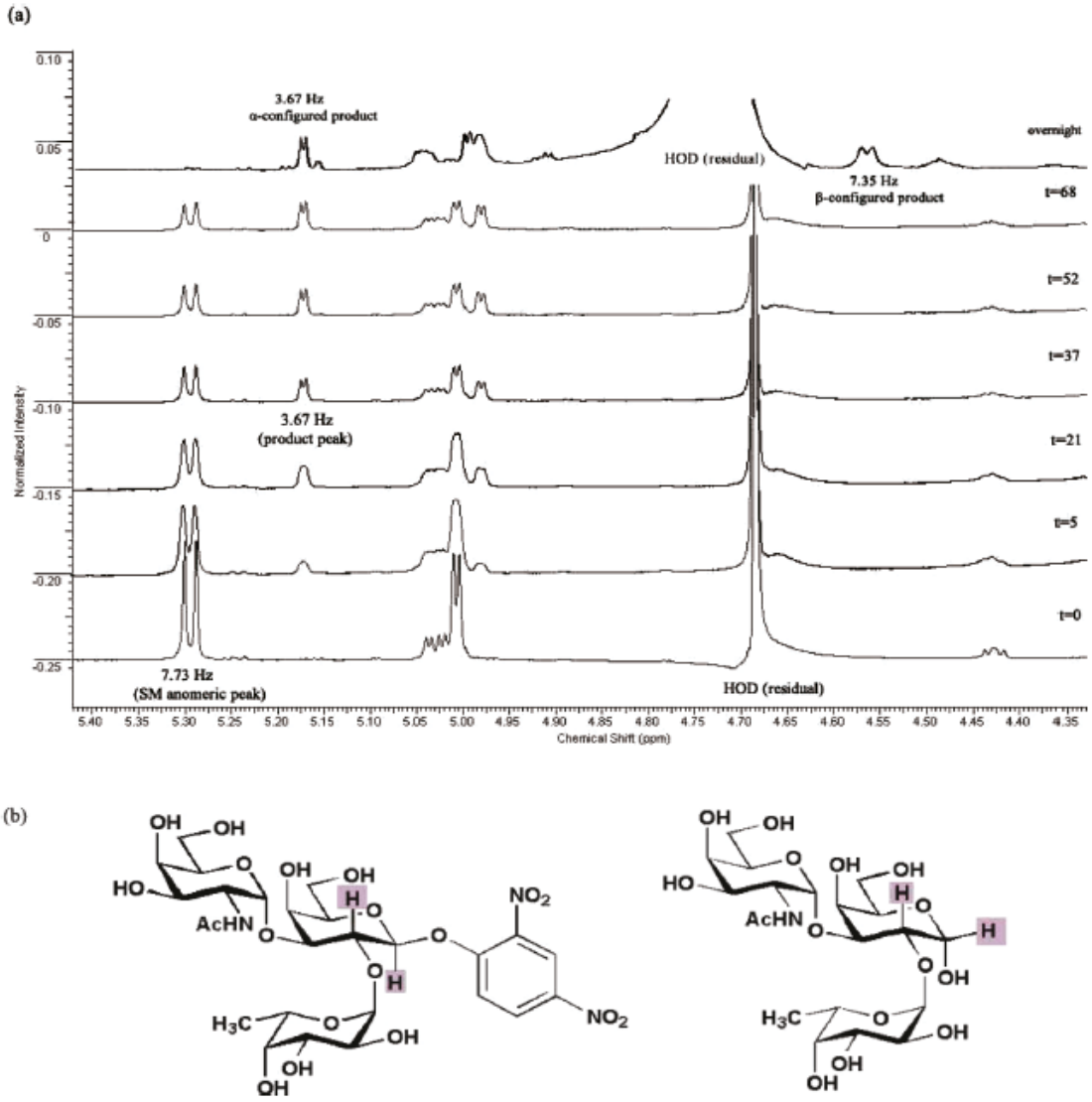



***Figure 3.7:*** *Determination of stereochemical outcome by $^1$H-NMR spectrometry. (a) $^1$H-NMR spectra of reaction mixtures containing DNP-A-trisaccharide to which EABase has been added, recorded after times indicated (in minutes); (b) structure of the starting material, DNP-A-trisaccharide, with the H1 and H2 protons shaded in purple (left) and the structure of the product formed in the alpha configuration with H1 and H2 protons shaded in purple (right).*

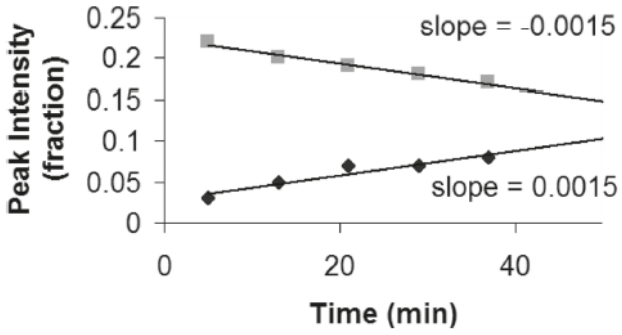



***Figure 3.8:*** *Comparison of the decrease in intensity of substrate peak with time (negative slope) with the increase in intensity of product peak with time (positive slope).*

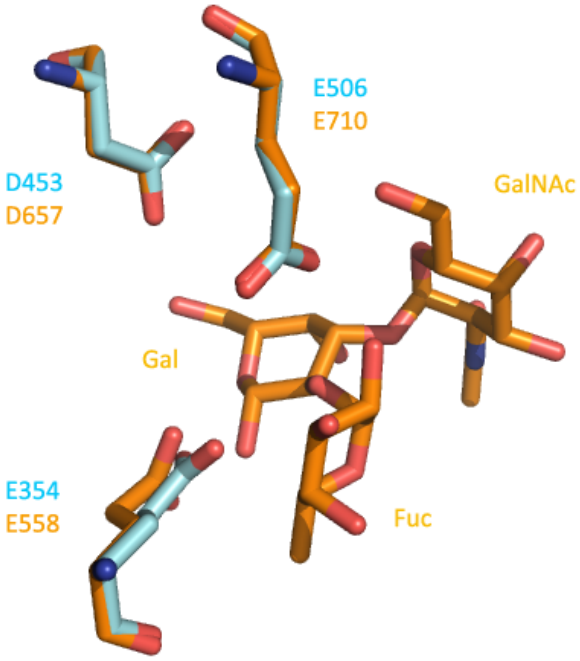



***Figure 3.9:*** *SSM Superimposition of the active site of Sp3GH98 in orange (pdb: 2WMI; coordinates obtained through a crystal structure by Higgins et. al) onto the active site of EABase in cyan (coordinates obtained through Phyre2 modelling—no crystal structure is available for EABase): The catalytic triad is shown with the equivalent residues aligning in the crystal structure of Sp3GH98.*

## 3.5 Kinetic characterization of wild type EABase and E354A, D429A, D453A, E467A, and E506A mutants

### 3.5.1 Kinetic parameters for WT EABase and mutants

#### *3.5.1.1 Using DNP-A-tri as a chromogenic substrate in a continuous assay*

Inverting glycosidases proceed through a concerted, single-displacement mechanism in which one carboxylic acid residue acts as a base and another as an acid. Since these residues are critical to the mechanism, we expect their removal to have a significant effect on catalysis: removal of the acid catalyst should slow the hydrolysis of substrates with poor leaving groups much more than substrates with good leaving groups (pKa <7) and removal of the base residue should substantially slow hydrolysis of all substrates. Michaelis-Menten kinetic parameters for the hydrolysis of DNP-β-A-trisaccharide by wild type EABase and each of the five mutants are summarized in Table 3.1. These activity measurements provide excellent insight into the roles of the residues, as follows.

The mutants E506A and D453A were completely inactive when assayed with the highly activated substrate, DNP-β-A-trisaccharide, even when assayed at high concentrations of substrate and enzyme. The lack of activity is likely due to the important roles of the mutated residues in deprotonation and/or positioning of the water molecule for attack at the anomeric center. Confirmation that the substrate was intact and that no inhibitors were present was provided by the addition of wild type EABase directly to the assays, leading to rapid turnover. The structural integrity of the mutants was verified by the CD spectra of these mutants which

showed that they were folded similarly to wild type. Additionally, upon the basis of precedent with the inverting β-glycosidase CenA from C. fimi, this would suggest that one or both of these residues functions as the base catalyst. It is quite common to find two residues acting in concert as catalytic acids/bases in glycosidases.[145–147] Alternatively, one of the carboxylic acids may play a particularly important role in coordinating substrate hydroxyls, possibly bridging across an adjacent pair as is seen in GH13 enzymes.[37,148,149] These findings are also consistent with the crystal structures obtained of Sp3GH98 and Sp4GH98 which show the equivalent residues coordinating a water molecule as described previously.

The $k_{cat}$ value for the E354A mutant was only 10 times lower than that for the wild type enzyme. Because this mutant retains considerable activity, this residue is unlikely to serve as a base but may serve as the acid catalyst. As the dinitrophenol moiety is a good leaving group, it is not surprising that the $k_{cat}$ value was not impacted more significantly since its departure does not require acid-catalytic assistance as much as do leaving groups with higher $pK_a$ alcohols. The $K_M$ of E354A is reduced by a factor of 2 compared to the wild type, which is of marginal significance but may be a result of more favourable interactions of the aromatic leaving group with a less polar residue at position 354.

The $k_{cat}$ and $K_M$ values for the D429A and E467A mutants are similar to those of the wild type EABase, indicating that they are not important in catalysis. A summary of these results is provided in Table 3.1.

***Table 3.1:*** *Kinetic parameters for wild type EABase and mutants with DNP-A-trisaccharide*

| | $k_{cat}$ ($min^{-1}$) | $K_M$ (μM) | $k_{cat}$ / $K_M$ ($\mu M^{-1}$ $min^{-1}$) |
|---|---|---|---|
| WT | 105±5 | 64±3 | 1.6±0.1 |
| E354A | 10±1 | 31±3 | 0.32±0.02 |
| D453A | NH | NH | NH |
| E506A | NH | NH | NH |
| E467A | 93±5 | 59±4 | 1.6±0.1 |
| D429A | 81±4 | 61±3 | 1.3±0.1 |

NH = No hydrolysis detected

No other detailed kinetic studies have been performed on mutants of GH98 enzymes. Kinetic data were obtained by Higgins *et al.* for wild type Sp4GH98 and Sp3GH98 enzymes using natural substrates and analyzed using a stopped assay and high performance anion exchange pulsed amperometric detection (HPAE-PAD).[144] Mutants modified at the equivalent residues to E354 in Sp3GH98 and Sp4GH98 were shown to have sufficiently low activity in their crystalline state in complex with natural substrates to allow for the structures of these enzymes with their intact substrates to be determined. While no kinetic parameters were determined for these mutants, the low activity observed with natural substrates supports the assignment of E354 as the catalytic acid. The rationale and further experiments to support this hypothesis are discussed in the next section.

### 3.5.2 Analysis of enzymatic cleavage of natural substrates

As mentioned previously, the observation of only very low activity of the proposed acid mutant of EABase using a natural substrate would provide support for its assignment. Initially the only natural substrate available was A+ porcine gastric mucin. The ability of the purified enzyme and

mutants to hydrolyze A+PGM was tested by TLC (Figure 3.10). The wild type EABase efficiently hydrolyses the substrate with release of A-trisaccharide. However, even when incubated with enzyme concentrations up to 1000 times higher than those used for the wild type EABase, no hydrolysis whatsoever of $A^+$ PGM by mutants of the putative base catalysts, D453A or E506A, is observed. This is completely consistent with observations with the activated substrate. By comparison, a small amount (≈20%) of substrate hydrolysis by the putative acid catalyst mutant, E354A, was observed after 16 h, but only when the mutant enzyme was used at concentrations 20 times higher than that in the wild type EABase experiment. This allows an approximate estimation of a 1000-fold rate reduction for nonactivated substrates as a consequence of the removal of E354. These results support the assignment of E354 as the acid catalyst since the removal of the acid residue leads to non-activated substrates being hydrolyzed far slower than activated substrates.

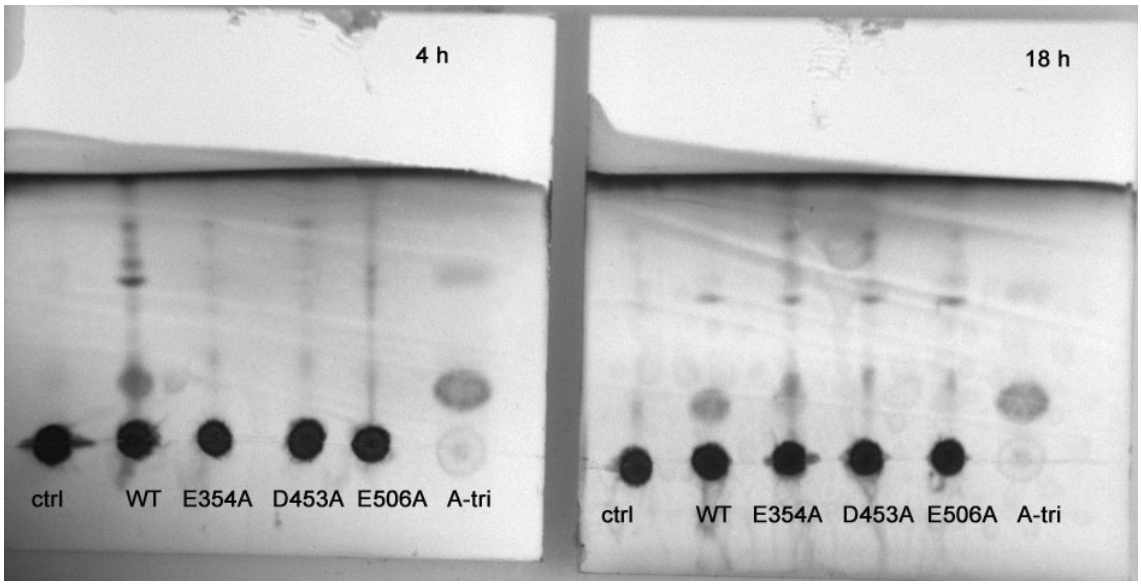



***Figure 3.10*** *TLC showing the reaction of natural substrate A⁺PGM with wild type and mutant enzymes. Lanes are as indicated control A⁺PGM in 20 mM sodium acetate pH 6.0, wild type, E354A, D453A, E506A. Two plates are shown, one after 4 hours and the other after 18 hours to illustrate the difference in rate of hydrolysis between the mutants.*

Though the approximation of a 1000-fold rate reduction for the hydrolysis of $A^+$PGM by the acid mutant was determined, further validation of E354A as the acid residue was performed by obtaining kinetic parameters for the hydrolysis of a substrate with a poor leaving group. The natural substrate A-pentasaccharide (structure shown in Figure 3.11) was used with the disaccharide being the poor leaving group. The hydrolysis of this natural substrate cannot be monitored directly therefore kinetic parameters for the hydrolysis of A-pentasaccharide were determined spectrophotometrically using a stopped reducing sugar assay with p-hydroxybenzoic acid hydrazide (HBAH). Acid hydrazides react with reducing sugars under alkaline conditions to yield yellow anions as one of the reaction products which can be monitored at 420 nm. Control reactions which did not contain EABase were also run. As can be seen in Table 3.2, kinetic parameters of $K_M$=384 ± 42 µM and a $k_{cat}$ =4.1 ± 0.2 min-1 were

obtained with the wild type EABase. Interestingly, this $K_M$ value is 6-fold higher than that obtained with DNP-A-trisaccharide implying that the dinitrophenyl moiety binds more tightly than the natural sugar at this site. Furthermore, the $k_{cat}$ is significantly lower than that for the aryl glycoside. The rates of hydrolysis of A-pentasaccharide by E354A were far lower, an approximate $k_{cat}$ value of (3.7 ± 0.1) x $10^{-3}$ $min^{-1}$ being obtained at saturating (3 mM) substrate concentrations. Thus, replacement of Glu354 with Ala results in an 1100-fold rate reduction for nonactivated substrates completely consistent with the earlier estimate. As was elaborated above, such kinetic behaviour with an activated and unactivated substrate is completely consistent with a role for E354 as the acid catalyst.

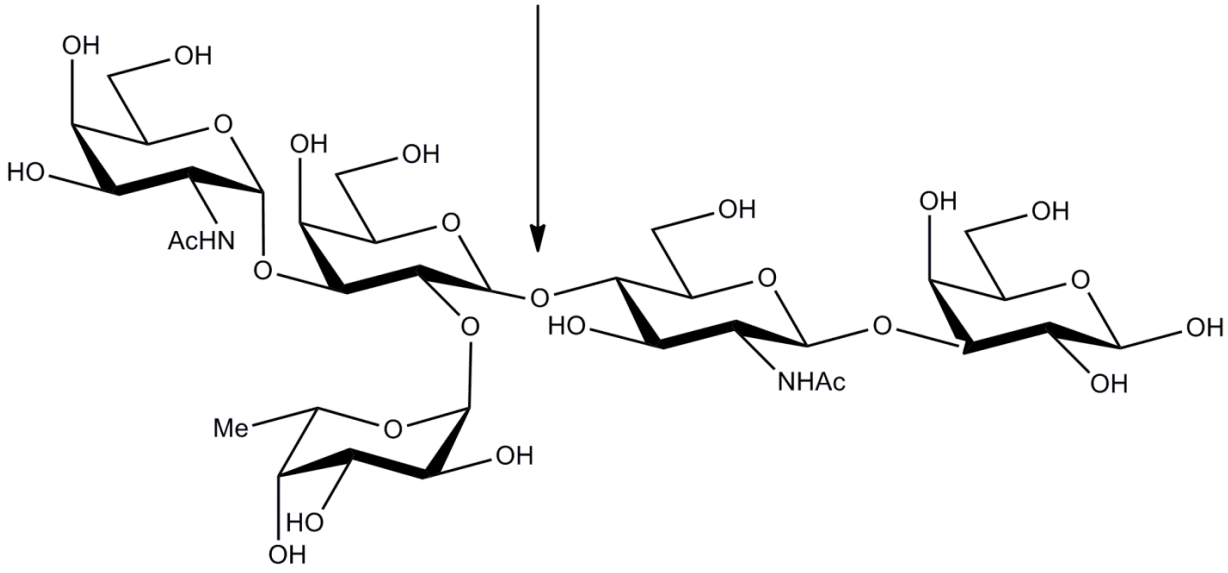



*Figure 3.11 Structure of the A-pentasaccharide. The arrow is the position of enzymatic hydrolysis*

***Table 3.2:*** *Kinetic parameters for wild type EABase and mutants with A-pentasaccharide*

| | $k_{cat}$ ($min^{-1}$) | $K_M$ (μM) | $k_{cat}$ / $K_M$ ($\mu M^{-1}$ $min^{-1}$) |
|---|---|---|---|
| WT | 4.1 ± 0.2 | 384 ± 42 | (1.1 ± 0.1) * $10^{-2}$ |
| E354A | (3.7 ± 0.1) * $10^{-3}$ | ND | ND |

ND = Not determinable

Indeed inspection of the crystal structures of Sp4GH98 and Sp3GH98 shows that the equivalent acid residue is 2.5 Å away from O1 of the reducing end galactose of the trisaccharide or disaccharide products and forms a strong hydrogen bond which would allow delivery of the proton to the glycosidic oxygen, consistent with its role as a general acid.

### 3.5.3 pH dependence of activity

Values of $k_{cat}$ /$K_M$ for the hydrolysis of DNP-A-trisaccharide by wild type EABase and the E354A mutant were determined by the substrate depletion method as a function of pH within the stability range for each enzyme: 5.0-9.0 for wild type and 5.5-8.5 for E354A. These values are plotted in Figure 3.12. The pH profile of the wild type enzyme is a typical bell-shaped curve with optimal activity at pH 6.0, reflecting the ionisation of the acid and the base residues. The $pK_a$ values for the general base and general acid ionisable groups in the free enzyme, obtained from a fit of each limb in the plot of $k_{cat}$ /$K_M$ versus pH using GraFit 4.0, are respectively $pK_{a1}$=5.0 and $pK_{a2}$=7.1. However, the value of $pK_{a1}$ should be taken only as an estimate since the data obtainable for this limb of the profile were limited due to the instability of the enzyme at lower

pH values. The pH profile of the E354A mutant was significantly different from that of the wild type enzyme, with the $k_{cat}/K_M$ values remaining constant over the whole pH range at which the mutant was stable. The profile could not be obtained at sufficiently low pH values to report on the acidic limb of the pH dependence. However, the absence of a basic limb in the profile of the mutant is completely consistent with the notion that the residue mutated, E354, is indeed the acid catalytic residue because the activity of the E354A mutant is essentially independent of pH in the range over which the activity of the wild type enzyme decreased severely with pH. These observations, along with previous kinetic data, strongly support the assignment of E354A as the acid residue.

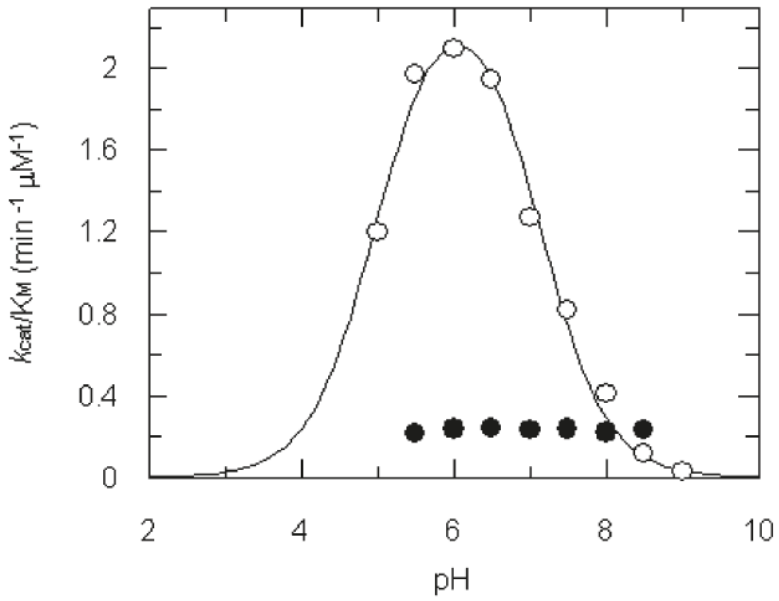



***Figure 3.12:*** *pH Dependence of $k_{cat}/K_M$ for the hydrolysis of DNP-A-trisaccharide by wild type EABase (○) and its E354A mutant (●).*

### 3.5.4 Attempts at rescue using azide, formate and acetate

Rescue of activity of mutants using small external nucleophiles such as azide and formate has been shown in many retaining and some inverting glycosidases.[49,150–153] No rescue was seen upon addition of formate or azide anions to any of the EABase mutants. However, this lack of rescue is consistent with studies on other inverting β-glycosidases, in which acid and/or base residues have been mutated.

## 3.6 Applications of EABase: removal of B-antigens from RBCs using EABase

Unfortunately the observation that EABase is an inverting glycosidase considerably complicates efforts to generate a glycosynthase for the modification of surfaces of blood cells since retaining glycosidases have proved the most malleable in this manner: few inverters have been converted.[154–159] However, EABase can indeed be efficiently used to cleave both A and B antigens off cell surfaces as demonstrated by Andersen *et al.* who were able to remove the antigenicity of both a commercial supply of type A red blood cells (92%) and type B red blood cells (100%). Use of this enzyme alleviates the need for specific enzymes for each blood type and a mix of enzymes for AB types. The use of this enzyme or other A or B antigen-removing enzymes does not mask protein antigens such as Rhd antigens on blood cells so other approaches are needed.

Strategies involving the physical masking of the blood cells using PEG or HPG covalently attached to the cell surface have been previously demonstrated and are currently being further researched by members of Dr. Jay Kizhakkedathu's lab.[160–163] However, the polymer strategy alone does not mask the antigenicity caused by A or B antigens on the RBCs since these project far from the surface. In order to generate completely antigen-null red blood cells, a combination of polymer grafting and enzymatic hydrolysis of A/B antigens is being used. Nick Rossi of Dr. Jay Kizhakkedathu's lab demonstrated that EABase was able to completely remove antigenicity from B type red blood cells as determined by a visual assay involving agglutination of blood samples over time when the appropriate A or B antibodies were added to the enzyme-modified blood samples. In combination with the use of covalently linked polymers, EABase was able to assist in making antigen-null cells from B-type RBCs. However, antigenicity was not fully removed from A type blood. This is likely due to the heterogeneous nature of the A antigens on the cells and the different linkages (-Gal-β1,3-GlcNAc/-Gal-β1,4-GlcNAc) present which are hydrolysed to different extents by EABase (described previously). Additionally, the combination of polymer and enzymatic hydrolysis is still fairly inefficient, precluding its use on a clinical scale.

A fascinating observation made by Rossi *et al.* was that inclusion of non-reactive polymers such as dextran, PEG, or icodextrin during polymer grafting reactions dramatically increased the amount of polymer that was grafted onto the red blood cells.[164] This observation led to the idea that addition of the non-reactive polymers may also improve enzymatic cleavage of A/B antigens from the red blood cells. Indeed, when these additives were used in combination with EABase it was shown by Dr. Rossi that significantly less enzyme was needed to obtain

essentially complete cleavage of the B-antigens from red blood cells. Agglutination time increased over 10-fold, even at the lowest enzyme concentration measured (10.5 U EABase). The mechanism by which these additives increase the efficiency of the hydrolysis of these antigens was initially hypothesized to be due to hydrophobic interactions between the additives, cell surface and enzyme that channeled the enzyme close to the cell surface. However, the effect which draws the enzyme close to the surface of the blood cell is a depletion interaction. The principles of this interaction are illustrated in Figure 3.13 and are as follows: when two large objects (the red blood cell and the enzyme in our case) are far apart the smaller object (additive polymer in our case) is free to occupy any of the space not occupied by the enzyme or red blood cell (Figure 3.13(a)), however when the red blood cell and enzyme are at a distance that is equal to or smaller than the diameter of the additive polymer, the additive polymer cannot occupy the space between the blood cell and enzyme (Figure 3.13(b))—this creates an entropically unfavourable situation in which a "void volume" is present between the two larger objects. Bringing the red blood cell and enzyme together (Figure 3.13(c)) eliminates that "void" space and is entropically favourable because the additive polymer is not excluded from a space not occupied by the enzyme or blood cell. [165]

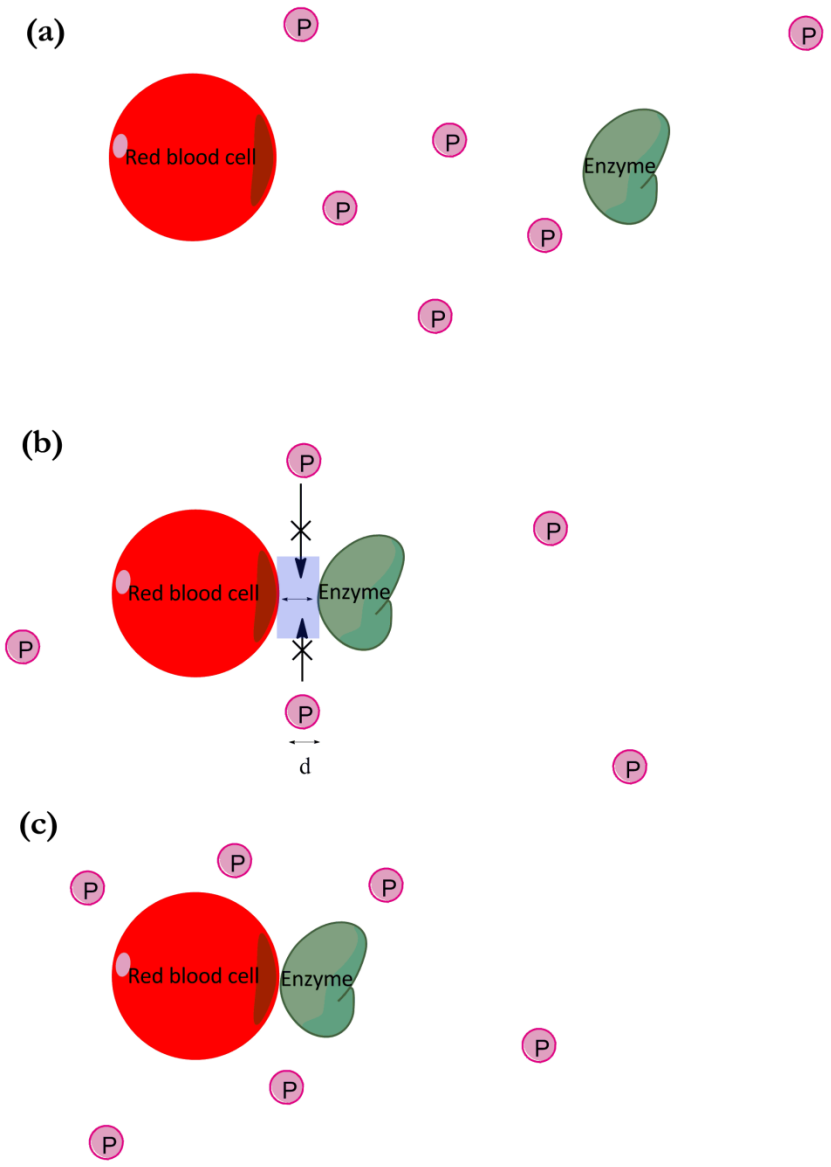



*Figure 3.13 The depletion interaction: (a) the enzyme and red blood cell are far enough apart that the additive polymer (pink circles labeled P) can occupy any volume not occupied by the red blood cell or enzyme (b) void volume (blue) created when the red blood cell and enzyme are at a distance of ≤ diameter of the additive polymer, and (c) removal of the void volume when the cell and enzyme come in contact*

To test that these rate enhancements in the presence of additives were due to the depletion interaction and not due to an interaction (or stabilization) of the enzyme itself, effects on the rate of hydrolysis of DNP-A-tri by EABase *in vitro* by PEG 8000, dextran (40 K and 80 K) and icodextrin were investigated at various concentrations of the additives. From Table 3.3 it is clear that these additives did not affect the activity of wild type EABase in the hydrolysis of DNP-A-trisaccharide. A-pentasaccharide was also used as a substrate to determine whether these additives had any effect on the cleavage of non-activated substrates, and again no effect was observed. These data indicate that the activity increase for the hydrolysis of B-antigens on the surface of RBCs observed in the presence of the additives is not a result of any special interaction between the additive and EABase that somehow optimised the enzyme. This is not surprising as it is unlikely that these additives would be able to modify the active site.

***Table 3.3*** *Kinetic parameters obtained for the hydrolysis of DNP-A-tri by EABase in the presence of several additives.*

| **Polymer [mg/mL]** | | **$k_{cat}$ ($min^{-1}$)** | **$K_M$ (μM)** |
|---|---|---|---|
| Wild Type | | 108 ± 7 | 62.1 ± 2.3 |
| Dextran(40) | 500 mg/mL | 108 ± 11 | 63.1 ± 1.0 |
| | 650 mg/mL | 105 ± 6 | 59.9 ± 1.2 |
| | 800 mg/mL | 110 ± 9 | 64.7 ± 1.4 |
| Dextran (80) | 500 mg/mL | 112 ± 9 | 66.5 ± 4.0 |
| | 650 mg/mL | 102 ± 4 | 59.6 ± 3.2 |
| | 800 mg/mL | 100 ± 5 | 64.3 ± 2.4 |
| PEG 8000 | 500 mg/mL | 101 ± 4 | 63.9 ± 2.0 |
| | 650 mg/mL | 113 ± 10 | 61.7 ± 1.8 |
| | 800 mg/mL | 105 ± 6 | 64.2 ± 2.4 |
| Icodextrin | 500 mg/mL | 108 ± 8 | 62.8 ± 1.9 |
| | 650 mg/mL | 103 ± 6 | 61.3 ± 2.2 |
| | 800 mg/mL | 106 ± 4 | 62.6 ± 3.0 |

The effects of additives on enzyme-catalyzed hydrolysis of surface antigens and artificial glycosides was also tested using α-(1,2)-fucosidases from GH29 to illustrate that the lack of any increase in catalytic activity was not EABase specific. The additives were also shown to increase the rate of defucosylation of the A and B antigens on the surface of red blood cells by the α-L-fucosidases The hydrolysis of para-nitrophenyl α-L-fucoside (pNPαfuc) by two fucosidases in this family, one from *T. maritima* and the other from *B. thetaiotaomicron* BT2970, was monitored in the presence of varying concentrations of the same additives and compared to a control without any additive. The results, summarized in Table 3.4 and 3.5, show that unsurprisingly no effect on activity is observed.

***Table 3.4*** *Kinetic parameters obtained for the hydrolysis of pNPαfuc by T. maritima α-fucosidase in the presence of several additives.*

| **Polymer [mg/mL]** | | **$k_{cat}$ ($min^{-1}$)** | **$K_M$ (μM)** |
|---|---|---|---|
| Wild Type | | 346 ± 16 | 31 ± 1 |
| Dextran(40) | 500 mg/mL | 358 ± 11 | 33 ± 1 |
| | 650 mg/mL | 345 ± 16 | 39 ± 2 |
| | 800 mg/mL | 350 ± 19 | 34 ± 2 |
| Dextran (80) | 500 mg/mL | 362 ± 16 | 35 ± 2 |
| | 650 mg/mL | 355 ± 12 | 41 ± 2 |
| | 800 mg/mL | 302 ± 18 | 39 ± 3 |
| PEG 8000 | 500 mg/mL | 354 ± 8 | 44 ± 2 |
| | 650 mg/mL | 359 ± 9 | 46 ± 2 |
| | 800 mg/mL | 361 ± 12 | 42 ± 2 |
| Icodextrin | 500 mg/mL | 343 ± 9 | 35 ± 1 |
| | 650 mg/mL | 360 ± 7 | 31 ± 2 |
| | 800 mg/mL | 321 ± 10 | 31 ± 2 |

***Table 3.5*** *Kinetic parameters obtained for the hydrolysis of pNPαfuc by B. thetaiotaomicron BT2970 α-fucosidase in the presence of several additives.*

| **Polymer [mg/mL]** | | **$k_{cat}$ ($min^{-1}$)** | **$K_M$ (mM)** |
|---|---|---|---|
| Wild Type | | 319 ± 39 | 26 ± 1 |
| Dextran(40) | 500 mg/mL | 302 ± 31 | 28 ± 1 |
| | 650 mg/mL | 311 ± 22 | 26 ± 1 |
| | 800 mg/mL | 305 ± 26 | 21 ± 1 |

| **Polymer [mg/mL]** | | **$k_{cat}$ ($min^{-1}$)** | **$K_M$ (mM)** |
|---|---|---|---|
| Dextran (80) | 500 mg/mL | 336 ± 41 | 31 ± 2 |
| | 650 mg/mL | | |
| | 800 mg/mL | 326 ± 35 | 29 ± 1 |
| | | 329 ± 35 | 27 ± 2 |
| PEG 8000 | 500 mg/mL | 298 ± 22 | 29 ± 2 |
| | 650 mg/mL | | |
| | 800 mg/mL | 296 ± 26 | 31 ± 3 |
| | | 297 ± 31 | 31 ± 2 |
| Icodextrin | 500 mg/mL | 321 ± 28 | 26 ± 1 |
| | 650 mg/mL | | |
| | 800 mg/mL | 318 ± 26 | 27 ± 1 |
| | | 320 ± 24 | 28 ± 2 |

The concentrations of additives used were in the same range as those used to show a significant increase in catalytic efficiency of EABase (as well as α-fucosidase and α-acetylgalactosaminidase) in the cleavage of B-antigen (the fucosyl or N-acetylgalactosaminyl residues) from RBCs. From these results it is clear that the additives do not affect enzyme activity directly under the conditions tested.

## 3.7 Conclusions and future directions

Based on the findings in this chapter, EABase is an inverting endo-β-galactosidase. Both activated and nonactivated substrates were used to kinetically characterize EABase and its mutants (E354A, D429A, D453A, E467A, and E506A) at pH 6.0, 37 $^{o}$C.

Hydrolysis of DNP-β-A-trisaccharide by EABase follows normal Michaelis-Menten kinetics with an apparent $K_M$ of 64 ± 3 μM and a $k_{cat}$ of 105 ± 5 $min^{-1}$. Mutation of two putative active site residues, D453 and E506, to alanine resulted in complete loss of activity, strongly suggesting that one or both of these residues functions as the base catalyst. The kinetic data also strongly suggest that E354 is the acid catalyst since the activity of the E354A mutant with nonactivated natural substrates is 1100-fold lower than that of the wild type enzyme, while its activity is only 10-fold lower when assayed with an activated aryl glycoside substrate (DNP-β-A-trisaccharide). Further support is obtained through comparison of pH profiles for the wild type and E354A mutants: mutation of the acid catalyst eliminates the basic limb from the bell-shaped pH dependence of $k_{cat}/K_M$ seen for the wild type enzyme. These findings are supported by x-ray crystallographic data and NMR studies performed on the only two other GH98 enzymes that have been characterized: Sp3GH98 and Sp4GH98.

Unfortunately since EABase is an inverting glycosidase, efforts to generate a glycosynthase for the modification of surfaces of blood cells are considerably more difficult. Unlike retaining glycosidases, inverting glycosidases do not possess true transglycosylation activity since there is no glycosyl-enzyme intermediate to partition; thus, only the inefficient thermodynamic

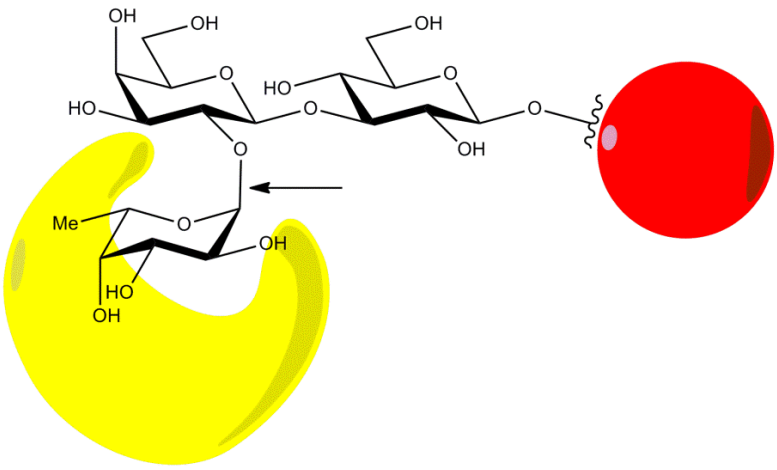
approach can be employed with wild type enzymes. Furthermore, only limited success has been attained in generating glycosynthases in inverting glycosidases, generally by the mutation of the general base residue, making this a somewhat less promising route.[154] However, through some combination of saturation mutagenesis of conserved residues, it may prove possible to engineer such an activity in the future.

Currently, EABase is being used in combination with the covalent grafting of HPG onto red blood cells to completely mask all antigenicity present in B-type RBCs in Dr. J. Kizhakkedathu's lab; the activity of EABase *in-vivo* is significantly improved in the presence of additives such as dextran though these additives have no effect on the activity of EABase with small substrates. Additional studies showing that this effect is observed with different types of cell surface modifying enzymes (such as glycosynthases and glycosyltransferases) would illustrate a more broad applicability to the method and would allow for an exciting and inexpensive method of increasing the activity of these enzymes *in vivo.* Additionally, to tackle the problem of EABase's low hydrolytic activity on A-type RBCs the directed evolution of EABase may assist in evolving an enzyme that is more efficient in the cleavage of the Gal-β(1,3)-GlcNAc linkage allowing for efficient removal of A-antigens off both $A_1$ and $A_2$ red blood cells rendering this an efficient combined method for the production of antigen-null red blood cells.

# Chapter 4: GH29 α-L-Fucosidases: Insight Into The Catalytic Acid/Base And Mechanistic Differences Within The Family

## 4.1 Introduction: family GH29 α-L-fucosidases

L-Fucosylated glycoconjugates are widely found in mammalian tissues such as the liver, brain and spleen as well as on red blood cells as ABO blood group and Lewis antigens. Fucosylated glycans are, not surprisingly, heavily involved in many physiological processes including antigenicity and immune responses, signal transduction and the adhesion processes in pathogens.[166–171] GH29 α-L-fucosidases are exo-glycosidases and most commonly hydrolyze α-(1,2)-linkages to galactose but can also cleave α-(1,3)-, α-(1,4)-, and α-(1,6)-linkages to N-acetylglucosamine residues.[172–175] A scheme of the reaction catalyzed by α-L-fucosidases which cleave the α-(1,2)-linkage is shown below (Scheme 4.1). The fucosidases from microorganisms and almond emulsin cleave only the α-(1,2)-linkage while those from mammalian sources or marine gastropods have a broader specificity, cleaving some, or all of the linkages mentioned above. Understanding the catalytic mechanism and active site residues of this family of fucosidases is important from a knowledge standpoint and will help to enable the modification of cell surface antigens and blood group antigens. Additionally, it will enable the development of synthetic enzymes for the synthesis of A/B/H antigens or for cleavage of fucosyl residues on antigens or cell surfaces.




*Scheme 4.1: Cleavage of the α-(1,2)-linkage by an α-L-fucosidase (yellow). The bond cleaved is indicated by an arrow and the substrate shown is the blood group H antigen (red blood cell in red)*

GH29 fucosidases have been shown to be retaining enzymes, since the first formal product of hydrolysis retains the same anomeric stereochemistry as the substrate.[176–178] Retaining glycosidases, as previously described, use a double displacement mechanism with two catalytic carboxylic acids at the active site—a nucleophile and an acid/base residue. The catalytic nucleophile was initially determined to be an aspartate residue (Asp224 in *Thermatoga maritima* fucosidase (Tmfuc)) through mass spectrometric analysis of a proteolytic digest of a fucosidase inactivated by the mechanism-based covalent inactivator, 2-deoxy-2-fluoro-α-L-fucosyl fluoride.[178] Sulzenbacher *et al.* later published the crystal structure of this same fucosidase from *Thermotoga maritima* along with kinetic data, both of which supported the assignment of a glutamate (Glu 266) as the catalytic acid/base.[179] The absence of observable hydrolysis of a substrate bearing a relatively poor leaving group (para-nitrophenol (pNP)) by the E266A mutant, yet significant observable hydrolysis by the same mutant if a substrate with a

good leaving group (fluoride) was used strongly supports E266's assignment as the acid/base. Additionally, azide rescue of the mutant provided strong support for the assignment.

Assignment of the catalytic residues in this family appeared to be straightforward until studies on the human α-fucosidase (Hfuc) by Liu *et al.* showed that the acid/base residue was not conserved between Hfuc and Tmfuc, despite both being members of GH29.[176] A graphical representation of the active sites of Hfuc, Tmfuc, and other selected fucosidases is found in the later sections. Neither sequence alignments nor alignments of the modelled structure of human fucosidase to the crystal structure of *T. maritima* were able to show conservation of the acid/base residue. Support for Liu *et al.'s* assignment of the acid/base residue as Glu289 came through detailed kinetic characterizations, including trapping of the intermediate using the E289G mutant, and a low $K_M$ for the mutant with activated substrates, azide rescue of the mutant, and isolation of a fucosyl-azide product with net retention of stereochemistry, which is the stereochemistry expected if rescue is observed with the acid/base mutant.[176]

Yet another GH29 α-L-fucosidase, from an Archaea, *Sulfolobacter solfotaricus* (Ssfuc), uses a catalytic diad as the acid/base; of which one carboxyl aligns structurally (in the model of Ssfuc) to the acid/base identified in Tmfuc.[180–183] The nucleophile residue is completely conserved, through sequence alignment of GH29 α-L-fucosidases, however, there is no observed conservation of the acid/base residue in this family; this makes the prediction of the acid/base residue difficult for GH29 fucosidases and it is unclear how E289 facilitates acid/base catalysis in Hfuc given that the modelled distance between this residue and the nucleophile would be 10.8

Å. A summary of the enzymes which have crystal structures and which are characterized is provided in Table 4.1.

***Table 4.1*** *Comparison of the structural homologies in the nucleophile residue and acid/base residue (from both Tmfuc and Hfuc) between GH29 α-L-fucosidases that either have solved crystal structures and/or have been kinetically characterized. The listing of kinetically characterized fucosidases does not include the characterizations of the enzymes in this chapter.*

| Species | Nucleophile | Acid/Base #1 (aligns with Tmfuc) | Acid/Base #2 (aligns with Hfuc) | |
|---|---|---|---|---|
| | | | | *Have crystal structures* |
| *BT2192* | D188 | E234 | E275 | Bacterial |
| *BT2970* | D229 | E288 | E306 | Bacterial |
| *BT3798* | D199 | E240 | NONE | Bacterial |
| *Bifido. Long. 2336* | D172 | E217 | E281 | Bacterial |
| | | | | *Have crystal structures and are kinetically characterized* |
| *T. marit* | D224 | E266 | E281 | Bacterial |
| | | | | *Kinetically characterized, and no available crystal structure* |
| *Sulfolo* | D150 | E200 | NONE | Archaea |
| *Homo sapiens A1* | D225 | NONE | E289 | Eukaryote |

*Species listed are as follows: BT2192: Bacteroides thetaiotaomicron 2192* (AAO77299)*; BT2970: Bacteroides thetaiotaomicron 2970* (AAO78076)*; BT3798: Bacteroides thetaiotaomicron 3798* (AAO78903)*; Bifido. long. 2336: Bifidobacterium longum sp. 2336* (ACJ53394)*; T. marit: Thermotoga maritima* (AAD35394)*; Sulfolo: Sulfolobacter solfotaricus* (AAK43159); *Homo sapiens A1: human* (AAA52481).

The aims of this chapter are to be able to predict the identity of the acid/base residue for all members of GH29 using available methods, since sequence comparisons do not reveal a conserved catalytically competent residue, and to experimentally verify the predicted residue. The method of predicting the acid/base residue will combine modelling of the structures of those fucosidases for which there is no available crystal structure, and comparing these models

with crystal structures of other fucosidases, including Tmfuc. The predictions obtained from the structural analysis will be combined with information from sequence alignments of GH29 fucosidases along with phylogenetic information on enzyme sources. A specific aim of this chapter is to experimentally verify the identity of the predicted acid/base residue in several *Bacteroides thetaiotaomicron* (BT) GH29 α-L-fucosidases along with the identity of the nucleophile residue. The development of a system for predicting the catalytic acid/base will contribute to our understanding of the mechanism of GH29 α-L-fucosidases and perhaps shed light on their evolutionary divergences and substrate specificities as well as provide useful information for the conversion of these enzymes into fucosynthases or for their use in defucosylation of antigens and other fucosylated oligosaccharides.

## 4.2 GH29 structure modeling and alignment of known crystal structures and PHYRE generated models to the crystal structure of Tmfuc

The structure of a protein is more conserved than its amino acid sequence, therefore, comparisons of structures are more likely to lead to reliable predictions of the catalytic residues, especially in families for which characterized enzymes with crystal structures exist. The majority of the enzymes in GH29, however, do not have associated crystal structures, thus requiring homology modeling to generate models which can then be used for comparison. The PHYRE2 (Protein Homology/analogY Recognition Engine) server was used to generate all GH29 models [184]. In short, models are generated by first finding a sequence alignment to other proteins and generating a hidden Markov Model of the submitted amino acid sequence (HMM)

and then comparing that HMM to a database of HMMs for all known structures of enzymes and generating a model based on the comparisons.[184] This method is cited to predict protein structure better than most other available methods.[184] The PHYRE2 generated models (for α-L-fucosidases from BT4136 (BT4136fuc), BT1625 (BT1625fuc), human (Hfuc), rat (Rfuc), dog (Dfuc)*, Streptococcus sp*. TIGR4 (TIGR4fuc), *Streptococcus sp.* 142 (Strep142fuc), *Sulfolobacter solfotaricus* (Ssfuc), and numerous other fucosidases in this family which are not directly referred to in this chapter) and known crystal structures (for GH29 α-L-fucosidases from BT2192 (BT2192fuc), BT2970 (BT2970fuc), BT3798 (BT3798fuc), and *Bifidobacterium longum* 2336 (BiAfcB)) were all compared to the crystal structure of the GH29 α-L-fucosidase from *Thermotoga maritima* (and to each other, though as the results are redundant they are not presented here). The comparisons between structures were made using a SSM-superimpose (secondary structure matching superimpose) in WinCOOT. The results obtained from the structural comparisons can be separated into three groups: Group S1: structures in which the acid/base residue overlaps with that of Tmfuc; Group S2: structures that undergo a conformational change for the acid/base residue to overlap with that of Tmfuc; and Group S3: structures in which the acid/base residue does not align with that of Tmfuc and would need to undergo a different conformational change (these are limited to the non-plant eukaryotic fucosidases). The nucleophile residue is shown to align/overlap throughout GH29 members both in sequence alignments and structural alignments; this is not surprising considering the nucleophile is the most important catalytic residue for retaining glycosidases.

### 4.2.1 Group S1: Structures in which the acid/base residue overlaps with Tmfuc's acid/base residue

All of the GH29 α-L-fucosidases from *Bacteroides thetaiotaomicron* (BT2192, 2970 and 3798) which have a crystal structure are in Group S1. Figure 4.1 shows a structural alignment of the regions surrounding the nucleophile and acid/base between a typical structure from a Group S1 α-L-fucosidase and to the crystal structure of Tmfuc. Not only does the nucleophile (Asp) align, but also the tryptophan residue (Trp 222 in Tmfuc) involved in substrate binding. Partial structural alignment (not all members of this group have a residue in this position) is observed for the residue which structurally aligns with the acid/base identified for the human fucosidase (Hfuc). There is little structural alignment of the residues neighboring the acid/base residue. The lack of conservation of neighboring residues may be due to the fact that they do not face the active site and therefore are not involved in substrate binding or positioning or hydrogen bonding with other active-site residues or the substrate. From Figure 4.1, the distance between the acid/base residue (E288 in BT2970fuc) and the nucleophile (D229 in BT2970fuc) is 5.0 Å, a distance commonly seen between the acid/base and nucleophile residues in retaining glycosidases.[185] E306 in BT2970fuc (the residue which structurally aligns with Hfuc's acid/base) appears to be too far from the nucleophile (10.8 Å) to provide acid/base catalysis without a conformational change. However, it is clear from the crystal structure of BT2970 with a covalently bound fucosyl moiety, that no such change occurs. Furthermore, E306 is obstructed

by Trp 227 and as a result would not be able to interact with the fucosyl-enzyme intermediate. Though it appears clear that E288 should act as the acid/base residue in BT2970fuc, both E306 and E288 were chosen for kinetic analysis in order to experimentally confirm the identity of the acid/base residue.

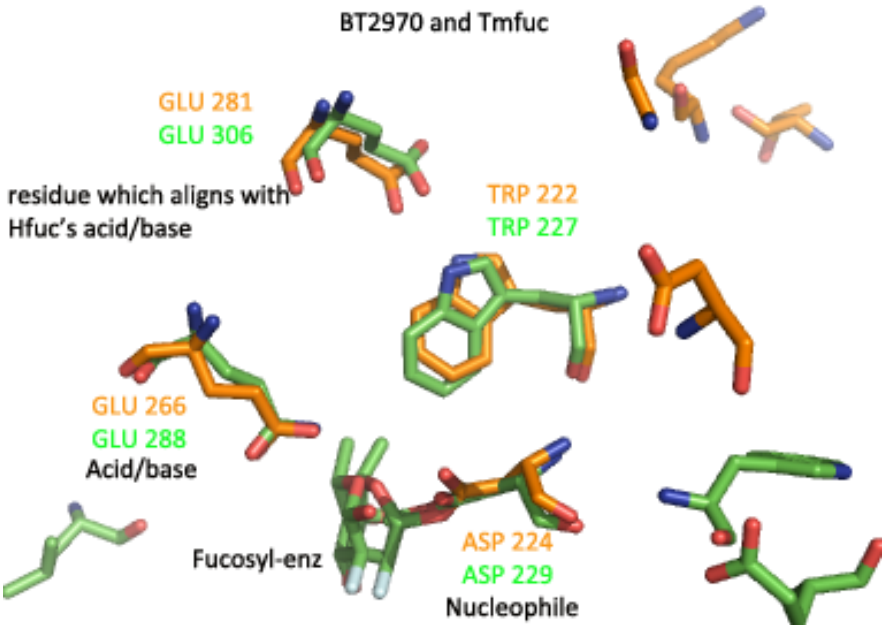



***Figure 4.1*** *PyMol illustration of a WinCOOT generated SSM-superimpose of the crystal structures of BT2970 and the 2-fucosyl enzyme of T. maritima α-L-fucosidase. Only selected active site residues are shown; the structure of Tmfuc is shown with an orange backbone and the structure of BT2970 is shown with a green backbone. The colors of the residues correspond to the colors of the structures.*

### 4.2.2 Group S2: structures in which the acid/base residue overlaps through an induced-fit mechanism with Tmfuc's acid/base residue

The acid/base residues of members in this group do not directly align with that of Tmfuc (or members of Group S1). Figure 4.2 shows the position of the proposed acid/base residue, when no substrate is bound, in the crystal structure of *Bifidobacterium longum* 2336 α-L-fucosidase (BiAfcB) and the position of the acid/base residue in a representative model of the α-L-fucosidase from BT4136. In both the crystal structure and model shown, the proposed acid/base residue is positioned 16-18 Å away from the nucleophile residue—this distance is also observed in models of other group S2 fucosidases. From Figure 4.2, E249 in BT4136 and E217 in BiAfcB (the proposed acid/base residues), is located on a loop that looks like it could fold in towards the active site and become appropriately positioned to provide catalysis.

At the time of this study, no experimental evidence for the proposed conformational change was available, and the only crystal structure available for this group was that of BiAfcB (with no bound ligand). Several predictions for the acid/base residue for members of this group were made from partial multiple sequence alignments and the major prediction for the acid/base residue for BT1625 and BT4136 is Glu 249 (numbering is the same for BT4136 and 1625). A more detailed explanation for the prediction of Glu 249 as the acid/base is found in Section 4.2.3.

Members of Group S2 show, unsurprisingly, complete structural overlap of the nucleophile, and also show overlap of the tryptophan residue involved in substrate binding. Additionally,

structural overlap is seen between Hfuc's acid/base residue and E305 (BT41236) or E281 (BiAfcB) ; though the ~10 Å distance between those residues and the corresponding nucleophile residue and their position in the active site (partially obstructed by the Trp residue), make it seem like a less favourable candidate for acid/base within this group.

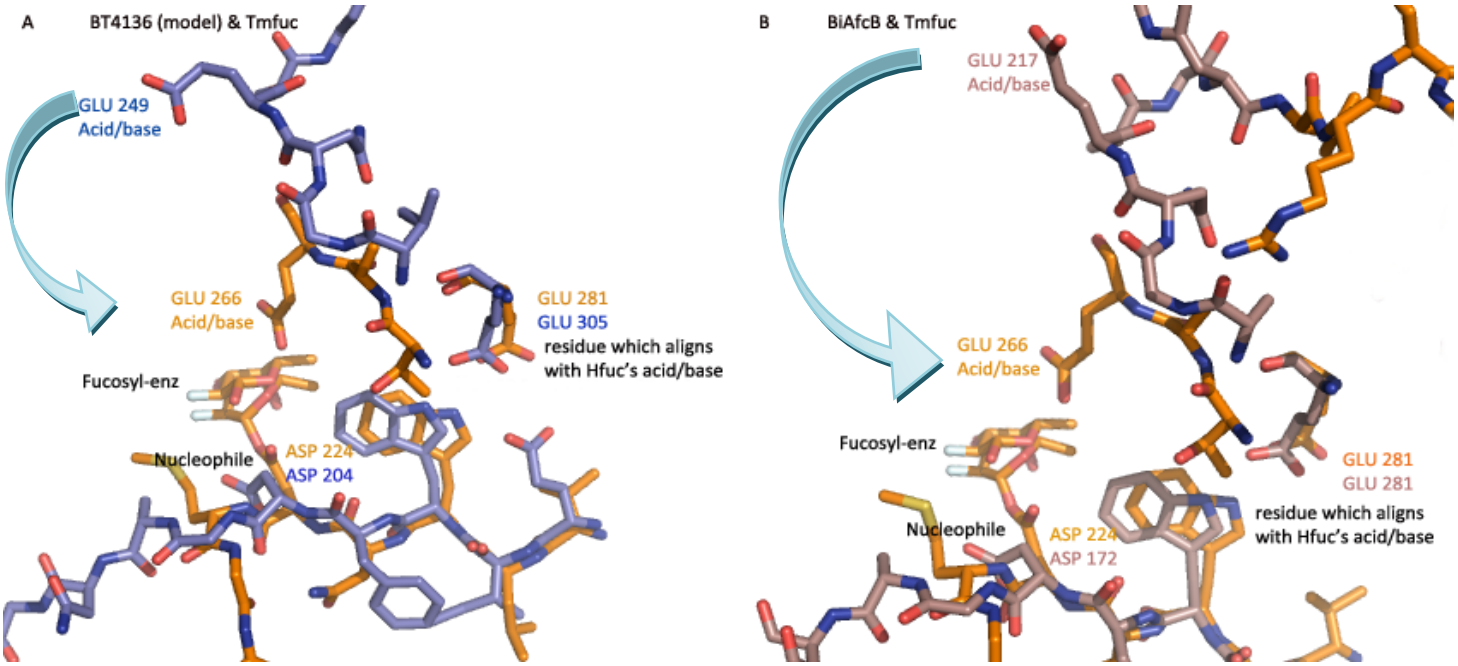



***Figure 4.2*** *Pymol generated illustrations of the WinCOOT SSM-superimpose between A) the crystal structure of the 2-fucosyl enzyme of Tmfuc (orange) and the model of BT2970fuc (blue) and B) the crystal structure of Tmfuc (orange) and the crystal structure of BiAfcB (purple). The conformational change is indicated with the blue arrow.*

Very recently, and after kinetic characterization of both BT4136 and BT1625's α-L-fucosidases were completed for this study, a crystal structure illustrating the induced-fit mechanism for BiAfcB in complex with deoxyfuconojirimycin (DFJ) was solved by Sakurama *et al*. From this

crystal structure (Fig 4.3), Glu 217 (also seen in Fig 4.2) is positioned 20.8 Å away from the nucleophile before ligand binding (in green), but becomes 5.8 Å away after binding (in purple), positioning it well to offer acid catalysis to the departing aglycone.[186] Direct evidence (through crystal structures) for the proposed conformational change supports the prediction of the acid/base residue for BT4136 (and BT1625) α-L-fucosidases as Glu 249. The glutamate (E305) corresponding to the acid/base in Hfuc was also selected for mutagenesis and kinetic characterization. A few other catalytically competent residues were selected for mutagenesis in BT1625fuc and BT4136fuc; D307, E286, E294 and E298. Rationales are provided with their kinetic characterization in section 4.7.

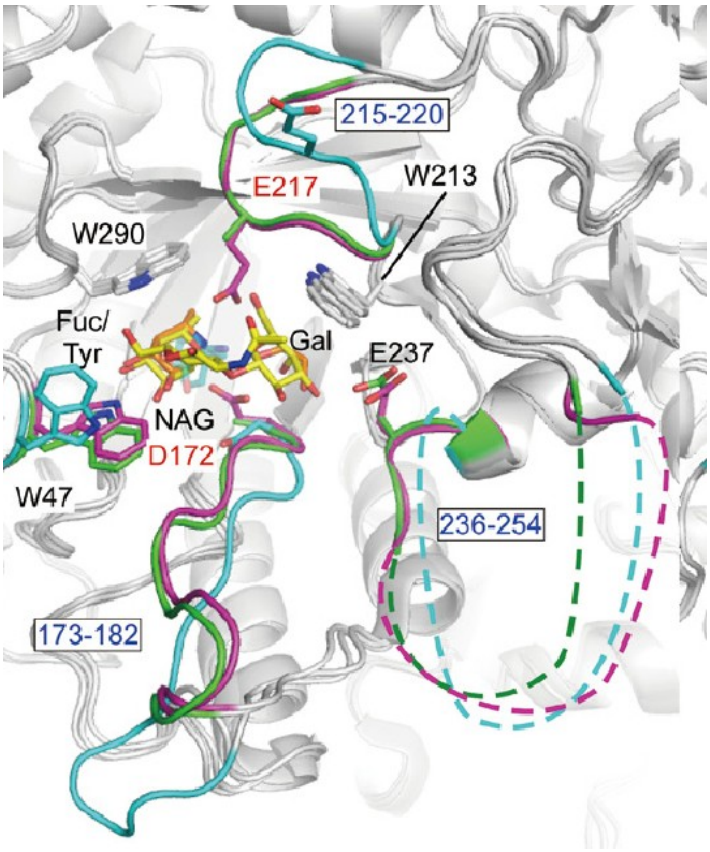



*Figure 4.3: Induced fit movement of BiAfcB upon ligand binding; (a) ligand-free enzyme is shown in blue and (b) the enzyme-DFJ complex is shown in purple. Residue E217 (label in red) moves from its position 20.8 Å away from the nucleophile to 5.8 Å in the enzyme-DFJ complex. This research was originally published in The Journal of Biological Chemistry. Sakurama, H., Fushinobu, S., Hikada, M., Yoshinda, E., Honda, Y., Ashida, H, Kumagai, H., Yamamoto, K., Katayama, T. 1,3-1,4-α-L-Fucosynthase that specifically introduces Lewis a/x antigens into type-1/2 chains. Journal of Biological Chemistry. 2012; [published online ahead of print March 26, 2012]. © The American Society for Biochemistry and Molecular Biology*

### 4.2.3 Group S3: GH29 α-L-fucosidases for which the acid/base residue does not align with that of Tmfuc

The members in Group S3 are limited to non-plant eukaryotic GH29 α-L-fucosidases and include the characterized human α-L-fucosidase. The human fucosidase in this group was the first fucosidase examined that contradicted the findings of Tarling *et al.* for the assignment of

acid/base residue in GH29. Modelling of Hfuc and comparison of the model to the crystal structure of Tmfuc (Figure 4.4) shows that there is no residue which would overlap with Tmfuc's acid/base residue. Additionally, unlike BiAfcB (and Group S2 α-L-fucosidases), no loop containing the acid/base residue is seen that could fold in and act through the same induced-fit mechanism that BiAfcB undergoes. The acid/base residue experimentally identified for Hfuc (E289) is 10.8 Å away from the nucleophile.

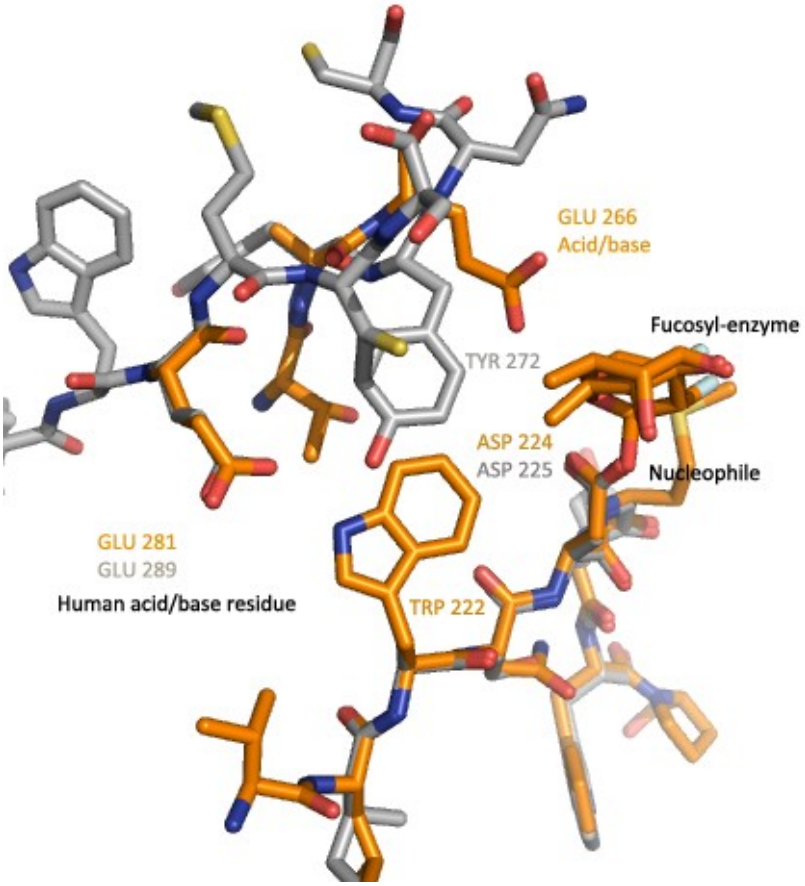



***Figure 4.4*** *Pymol generated illustrations of the WinCOOT SSM-superimpose between A) the crystal structure of the 2-fucosyl enzyme of Tmfuc (orange) and the model of Hfuc (grey). Important residues are listed and colour coded by their structure.*

Some of the major differences between Hfuc and Tmfuc include the presence of an alanine instead of a Trp residue near where the substrate binds (for Tmfuc) and the presence of a Tyr residue from another loop in its place. The tyrosine is not conserved structurally in other bacterial fucosidases or eukaryotic fucosidases; and it appears as if it could hydrogen bond with Glu 289 and assist in a conformational change (different from that observed in BiAfcB) upon substrate binding. Evidence that the active sites of Hfuc and Tmfuc are different also came from kinetic analysis of the alanine mutant of E70 (in Hfuc) which showed the mutant to possess similar activity and substrate binding as the wild type. The analogous residue in Tmfuc, E66, was shown to be important for substrate binding. As an induced-fit occurs upon substrate binding for BiAfcB, it would not be surprising if a different induced-fit occurs for Group S3 fucosidases.

These structural comparisons of GH29 α-L-fucosidases with the crystal structure of Tmfuc allow for a more visual method of predicting the acid/base residue—a prediction which will be verified through kinetic characterization in the following sections.

## 4.3 Phylogenetic tree analysis and amino acid sequence alignment with other GH29 fucosidases

The observation that three different structural classes of GH29 fucosidases exist was not straightforward. As there is no sequence-conserved acid/base in this family, identification of the acid/base residue in the loop of Group S2 structures came through comparison of the sequences of GH29 fucosidases based on their phylogenetic tree branches. A phylogenetic tree illustrating the evolutionary divergences in amino acid sequence for GH29 α-L-fucosidases was constructed by Sulzenbacher *et al.* (Figure 4.5).[179] From this tree (refer to Figure 4.5), two main branches of fucosidases exist from the first (earliest) divergence; the first consists of Class A and C and the second consists of Class B. Class A (highlighted in red) consists of several bacterial fucosidases, Class B consists of bacterial and plant fucosidases (highlighted in green), and Class C consists of non-plant eukaryotic fucosidases (highlighted in yellow).

Several interesting observations can be made when comparing the models (or crystal structures) of the fucosidases in each Class:

- all members of **Class A** belong in **Group S1**—their modeled acid/base residue structurally overlaps with Tmfuc
- all members of **Class C** belong in **Group S3**—they have a putative acid/base residue that is modeled to structurally align with the model of Hfuc
- **Class B** fucosidases are found in **Group S1 and S2**
- Class A fucosidases do not show general sequence alignment of the acid/base residue with each other

- Sequence alignment of all the α-L-fucosidases in Class B shows that the acid/base residue (corresponding to Tmfuc) is conserved (Figure 4.6b) with the exception of the two members at the bottom of the Class; these two are the most divergent members of Class B. The acid/base residues of the last two members are E208 and E221 which are just slightly left of the alignment (determined by their structural alignment with BiAfcB).

Since Class B fucosidases were found in both Group S1 and Group S2, the position of the residue which structurally overlapped with Tmfuc's acid/base could be determined for those fucosidases in Group S1. The position of those residues in the corresponding amino acid sequence allowed for the determination of the position of the acid/base residue in Group S2 fucosidases as all Class B fucosidases sequence sequence-aligned at that position. This observation allowed for the tentative assignment of the acid/base residue in Group S2 GH29 α-L-fucosidases and indicated that the loop would undergo a conformational change. The identification of the acid/base residue for Group S2 fucosidases through this method is supported by the (very recently published) crystal structure of BiAfcB in complex with deoxyfuconojirimycin (DFJ) obtained by Sakurama *et al.* showing the induced-fit and the position of the acid/base in the loop overlapping with Tmfuc's acid/base. The nucleophile residue is fully conserved, as expected, by sequence in Classes A, B, and C i.e. throughout all members in GH29 (Figure 4.6a). The residue corresponding to the human α-L-fucosidase's acid/base residue is not fully conserved in Class A or B (although it shows high conservation in Class B).

The outcome of all these comparisons (and experimental evidence of the induced-fit mechanism for BiAfcB) is that prediction of the acid/base is possible through structural comparison (which was made possible through phylogenetic analysis and sequence comparisons for Group B and Group C).

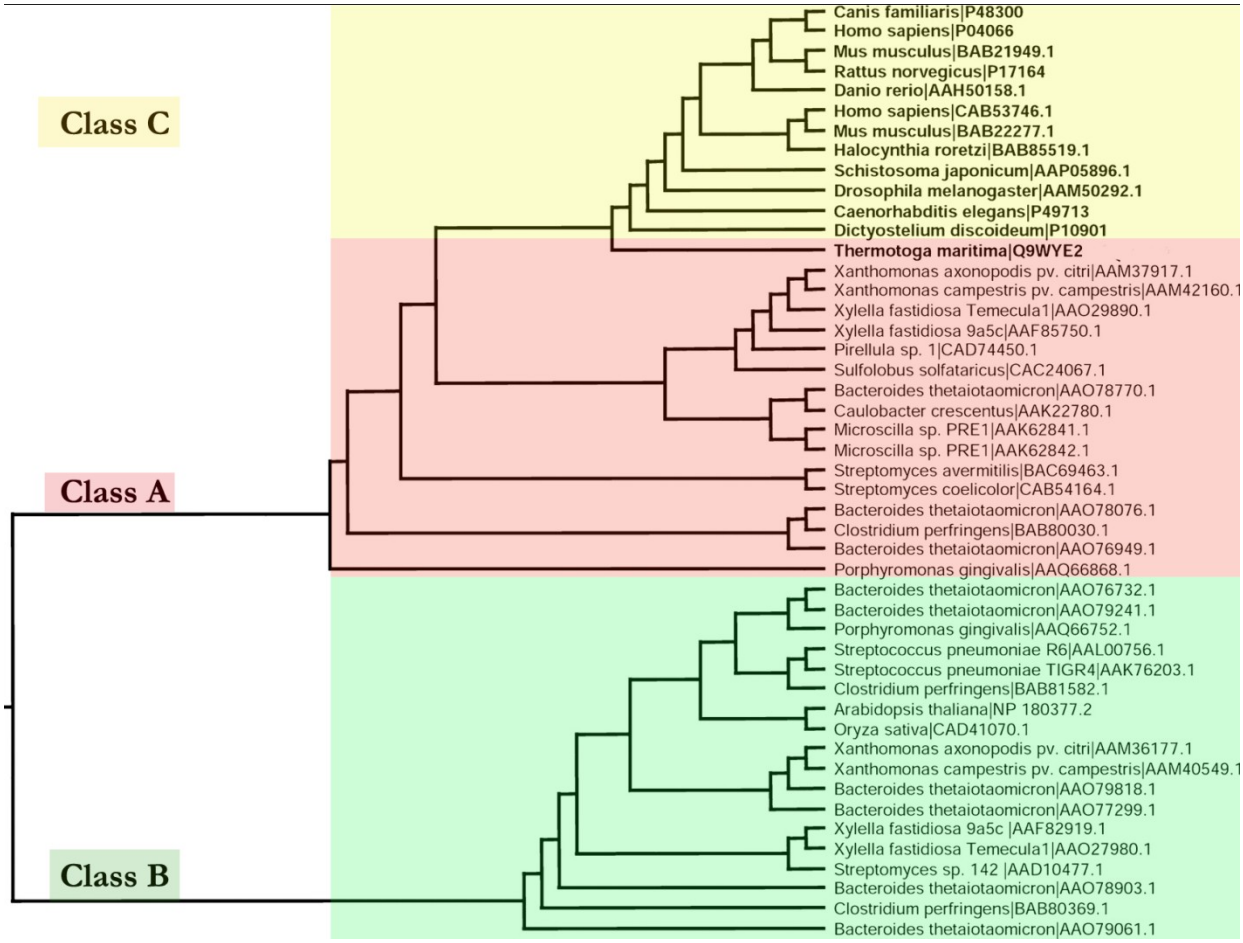



***Figure 4.5*** *Phylogenetic tree of GH29 α-L-fucosidases. The figure is adapted from a figure originally published in The Journal of Biological Chemistry. Sulzenbacher G, Bignon C, Nishimura T, Tarling CA, Withers SG, Henrissat B, Bourne Y. Crystal structure of Thermotoga maritima alpha-L-fucosidase. Insights into the catalytic mechanism and the molecular basis for fucosidosis.. Journal of Biological Chemistry. 2004; 279:13119-28 © The American Society for Biochemistry and Molecular Biology*

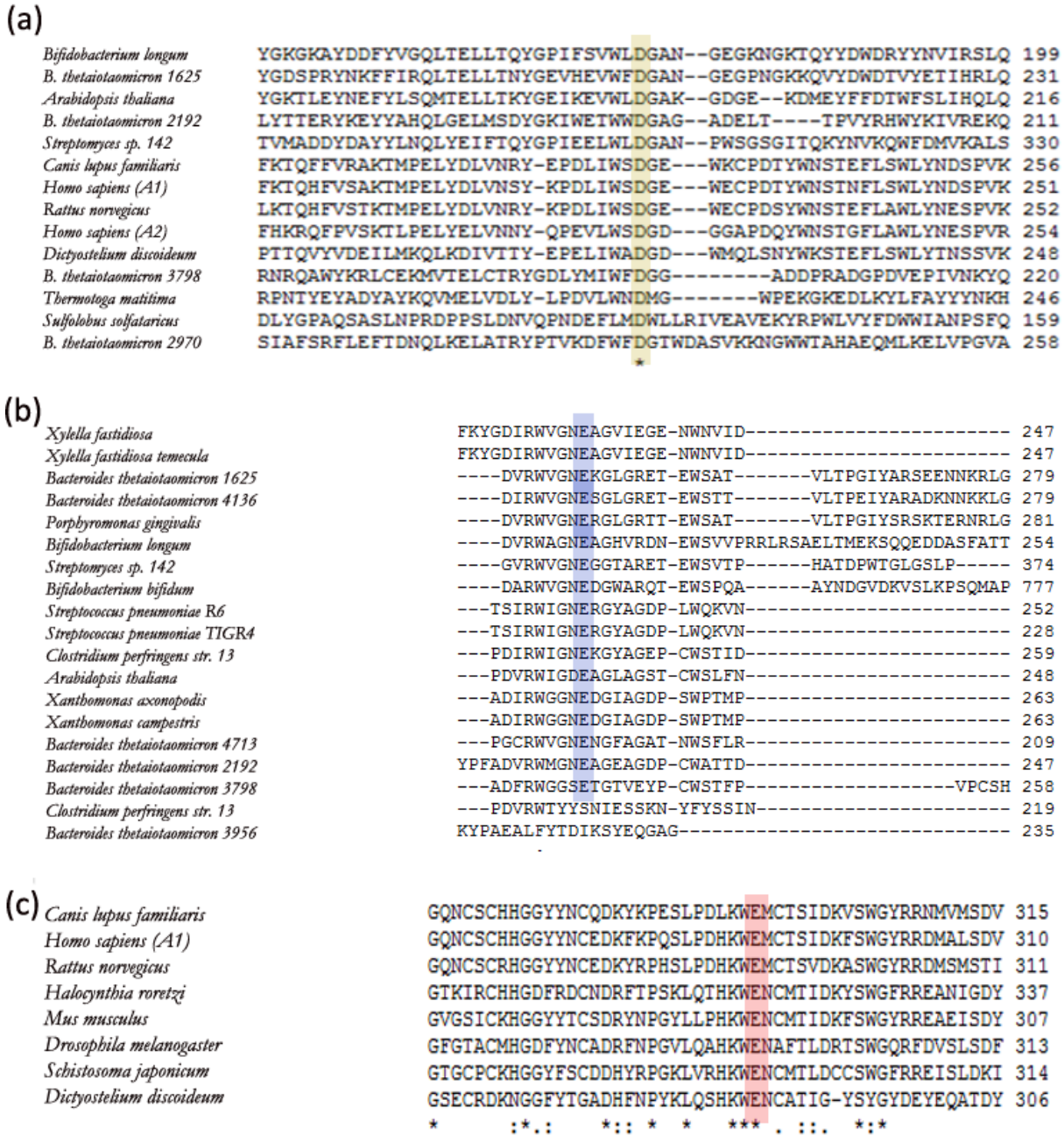



***Figure 4.6*** *Partial multiple sequence alignment of (a) selected GH29 α-L-fucosidases illustrating conservation of the nucleophile residue; (b) alignment of Class B α-L-fucosidases illustrating conservation of the acid/base residue; and (c) alignment of eukaryotic (non-plant) α-L-fucosidases illustrating conservation of the acid/base residue.*

## 4.4 Cloning, site-directed mutagenesis, and expression of Tmfuc and α-L-fucosidases from BT2970, BT4136 and BT1625

### 4.4.1 *T. maritima*

The α-(1,2)-fucosidase from *T. maritima* has been previously characterized in this lab, however the gene for wild type and mutants had not been cloned within the lab (the protein had been supplied by previous collaborators). As a result the gene for the α-(1,2)-fucosidase was successfully cloned into the pET22-b vector from *T. maritima* genomic DNA. The gene in this plasmid consists of a 1350 bp open reading frame encoding a protein consisting of 449 amino acids.

Again, although the nucleophile and acid/base mutants had previously been characterized in the lab, no plasmids containing the mutants were available and thus had to be generated using the wild type plasmid. Site directed mutagenesis was used to generate alanine and glycine mutants of the nucleophile residue, D224, and the acid/base residue identified in Tarling et al's publication, E266, and of the residue E281, which structurally aligns with the acid/base residue identified for Hfuc. The latter residue had not been previously investigated in Tmfuc. Mutagenesis using the QuikChange method failed for all six mutants presumably because of secondary structure formation in the polynucleotide. Instead, the 4-primer method was used to successfully generate the desired mutants. Sequencing of selected colonies showed that they contained the desired mutations.

### 4.4.2 BT2970

Plasmids containing the nucleophile and acid/base mutant for BT2970 (as determined by comparison of the models and sequences (at the time) of these enzymes to the crystal structure of *T. maritima* fucosidase) were made by Dr. Alicia Lammerts van Bueren and sent to us. In order to generate other mutants and compare activity to wild type α-L-fucosidase, site directed mutagenesis was performed to reverse the nucleophile mutant back to wild type. Mutagenesis was performed using the QuikChange method and yielded several colonies. DNA sequencing of selected colonies showed that the wild type sequence was successfully cloned. The wild type fucosidase (along with all mutants) are in a custom pET28a LIC plasmid (used by Lammerts van Bueren). The residue that structurally aligns with the acid/base of Hfuc, E301, was mutated to an alanine residue using site-directed mutagenesis on the wild-type plasmid using the Quikchange method.

### 4.4.3 BT4136 and BT1625

Wild type plasmids for both BT1625 and BT4136 α-fucosidases were supplied by Dr. Alicia Lammerts van Bueren along with primers for the nucleophile (D229G) and acid/base (E288G) mutants. In addition to the primers supplied, longer primers (and primers for the alanine mutants) were designed and obtained. The following mutations: D286A, D294A, E305A, D307A and D249A (numbering is the same for both species), were generated using optimized QuikChange methods and 4-primer methods (on mutants for which secondary structure formation inhibited the use of the QuikChange method). Sequencing of selected colonies showed that they contained the single desired mutations.

### 4.4.4 Expression of wild type and mutant α-L-fucosidases from *T. maritima*, BT2970, BT4136, and BT1625

All α-L-fucosidases were expressed under standard conditions as described in the Methods and Materials. These enzymes contained an N-terminal His-6-tag and were purified by nickel affinity chromatography and their purity was determined by SDS-PAGE. Tables 4.2 and 4.3 list the yields obtained for each protein.

***Table 4.2*** *Yields (per liter) of Tmfuc and BT2970fuc and their nucleophile, acid/base and human acid/base analogue alanine mutants. The corresponding residue numbers are written before the yields.*

| | Tmfuc (mg) | BT2970fuc (mg) |
|---|---|---|
| Wild Type | 33 | 80 |
| Nucleophile | (D224) 16 | (E229) 65 |
| Acid/Base | (E266) 10 | (E288) 40 |
| Human acid/base analogue | (E281) 22 | (E305) 25 |

***Table 4.3*** *Yields (per liter) of (alanine) mutant protein obtained per liter of LB broth for BT1625fuc and BT4136fuc*

| | BT1625fuc (mg) | BT4136fuc (mg) |
|---|---|---|
| D286A | 22 | 20 |
| D294A | Not expressible | Not expressible |
| E305A | 32 | 30 |
| D307A | 20 | 32 |
| D249A | 30 | 25 |
| Wild type | 28 | 34 |

## 4.5 pH stability and dependence

No literature on the kinetic behaviour of *Bacteroides thetaiotaomicron* fucosidases (2970, 1625 and 4136) at various pH values was available at the time these fucosidases were expressed and purified, therefore, the stability of these fucosidases with respect to pH had to be determined. The enzyme stability was determined by measuring residual enzyme activity (at pH 7.0 and 37$^{o}$C) after a 2-minute incubation of the enzyme in buffers with pH values ranging from 4-9. The enzyme was stable, i.e. showed >95% residual activity, between a pH of 4.5-8.5 (data not shown; up to 9.0 for BT4136fuc and BT1625fuc).

Values of $k_{cat}/K_M$ for the hydrolysis of pNPαfuc by wild type BT2970, BT4136, and BT1625 fucosidases were determined as a function of pH within the stability range for each enzyme: 4.5-8.5 for BT2970 and 4.5-9.0 for both BT1625 and BT4136. The pH profile for all three wild type enzymes is a typical bell-shaped curve with optimal activity at ~ pH 6.0 (BT2790 shown in Figure 4.7). The bell-shaped curves obtained suggest that activity is dependent upon two ionizable catalytic groups, which is consistent with the theory that one catalytic nucleophile and one catalytic acid/base residue are present. This is unlike the results obtained for Ssfuc by Cobucci-Ponzano *et al.* in which an irregular shaped curve (multiple humps) for the dependence of $k_{cat}/K_M$ on the pH was obtained indicating that > 2 ionizable catalytic groups were present [183]. The pKa of the acidic limb ($pK_{a1}$) is most likely representative of the $pK_a$ of the nucleophile residue as this residue needs to be fully deprotonated in order to be catalytically active. The basic limb ($pK_{a2}$) likely reflects the ionization of the acid/base catalyst which needs to be protonated for the glycosylation step in order to be catalytically active. A fit of each limb of the

plots of $k_{cat}/K_M$ versus pH in GraFit 4.0 yielded $pK_{a1}$ and $pK_{a2}$ values for the three wild type fucosidases; these are summarized in Table 4.4. Similarly, a pH profile of Tmfuc was made by Sulzenbacher *et al*. which showed a similar characteristic bell-curve.[178] Kinetic parameters for the hydrolysis of pNPαfuc were determined at a pH of 6.8 which is just above the pH optimum of 6.0, because these enzymes were intended for use on red blood cells after kinetic characterization.

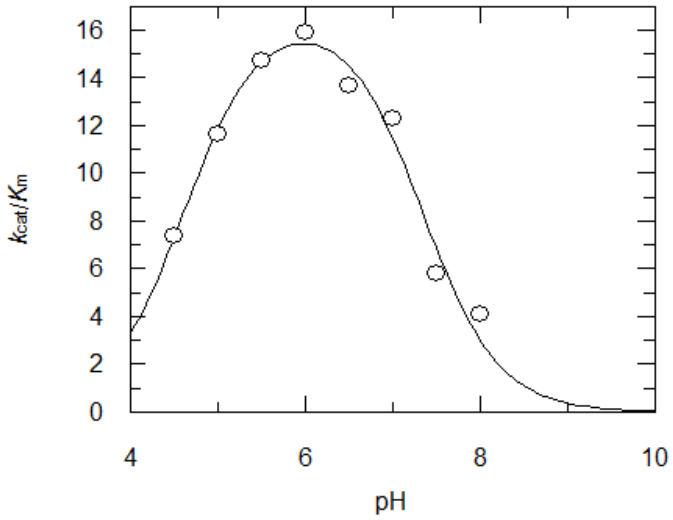



***Figure 4.7*** *The pH dependence of $k_{cat}/K_M$ for the hydrolysis of pNPαfuc by BT2970 on pH*

***Table 4.4*** *$pK_{a1}$ and $pK_{a2}$ values for wild type α-fucosidases from BT2970, BT1625 and BT4136 obtained from plots of $k_{cat}/K_M$ values vs pH*

| | $pK_{a1}$ | $pK_{a2}$ |
|---|---|---|
| BT2970 | 4.6 ± 0.1 | 7.3 ± 0.1 |
| BT1625 | 5.1 ± 0.1 | 7.1 ± 0.1 |
| BT4136 | 4.9 ± 0.2 | 7.2 ± 0.1 |

## 4.6 Kinetic characterization of Tmfuc and BT2970fuc (Group S1, Class A)

The Tmfuc and BT2970fuc enzymes structurally align and are both in Group S1 (and both in Class A). Kinetic parameters for the wild type, nucleophile (D224A) mutant, and acid/base (E266A) mutant (Table 4.5) are closely comparable to those obtained by Sulzenbacher *et al.* [179] The residue that corresponds, structurally, to Hfuc's acid/base residue, E281 has not been previously kinetically characterized, and therefore its role in the hydrolysis of fucosides by Tmfuc was unknown. The E281A mutant hydrolyzed 4-nitrophenyl α-fucoside (pNPαFuc) (at 37$^{o}$C and pH 5.0) as efficiently as wild type Tmfuc, suggesting that this residue does not play a role in catalysis or binding. This result is not surprising considering the 10 Å distance between E281 and the nucleophile residue.

***Table 4.5*** *Kinetic parameters for the hydrolysis of pNPαfuc (at 37$^{o}$C pH 5.0) by Tmfuc (WT and mutants)*

| | $k_{cat}$ ($min^{-1}$) | $K_M$ (μM) | $k_{cat}/K_M$ ($min^{-1}$/μM) |
|---|---|---|---|
| Wild type | 756 ± 28 | 55 ± 3 | 14 ± 4 |
| D224A (nuc) | ND | ND | (1.7 ± 0.1) x $10^{-3}$ |
| E266A (acid/base) | ND | ND | (2.6 ± 0.1) x $10^{-4}$ |
| E281A | 720 ± 19 | 68 ± 4 | 11 ± 2 |

ND = not determined; no saturation observed

Kinetic parameters for wild type BT2970fuc and its mutants are summarized in Table 4.6. Mutation of E288 (acid/base) to an alanine residue gave a 25000-fold decrease in $k_{cat}/K_M$ for pNPαFuc. The $k_{cat}$ values for the hydrolysis of αFucF (leaving group pKa = 3.2) are similar within experimental error to those obtained for the hydrolysis of pNPαFuc (leaving group pKa = 7.2) for both wild type and the alanine mutant of E288. This suggests that in both cases, deglycosylation is the rate limiting step, as one would expect a higher $k_{cat}$ value for the hydrolysis of αFucF if glycosylation was the rate limiting step [47,94]. Additionally, the $K_M$ observed for the hydrolysis of αFucF by the E288A mutant was 50-fold lower than that of wild type; resulting from an accumulation of intermediate, as expected for a substrate which does not require acid catalysis for the glycosylation step. This decrease in $K_M$ is consistent with the formation of the trapped glycosyl-enzyme intermediate obtained upon the reaction of αFucF with the E288N mutant by Lammerts van Bueren *et al* supporting its assignment as the acid/base residue.[185]

***Table 4.6*** *Kinetic characterization of the hydrolysis of pNPαFuc (at 37$^o$C pH 7.0) by BT2970*

| | $k_{cat}$ ($min^{-1}$) | $K_M$ (μM) | $k_{cat}/K_M$ ($min^{-1}$/μM) |
|---|---|---|---|
| WT | 319 ± 16 | 26 ± 5 | 12 ± 3 |
| D229A (nucleophile) | NA | NA | NA |
| E288A (acid/base) | $0.18 \pm 0.02$ | $391 \pm 3$ | $(4.7 \pm 0.6) \times 10^{-4}$ |
| E305A (Hfuc acid/base analogue) | 312 ± 11 | 32 ± 6 | 10 ± 3 |

NA = no activity observed at up to 150 μg protein

The $k_{cat}$/ $K_M$ of the acid/base mutant for the hydrolysis of αFucF is also much higher than the $k_{cat}$/ $K_M$ for the hydrolysis of pNPαFuc. Since the $k_{cat}$/ $K_M$ reflects the first irreversible step (glycosylation) and fluoride is a far better leaving group than para-nitrophenol, removal of the acid/base residue is expected to have a more significant effect on the glycosylation step for a substrate with a relatively poor leaving group.

***Table 4.7*** *Kinetic characterization of the hydrolysis of FucF (at 25$^{o}$C pH 7.0) by BT2970 WT and E288A mutants*

| | $k_{cat}$ ($min^{-1}$) | $K_M$ (μM) | $k_{cat}$/ $K_M$ ($min^{-1}$/μM) |
|---|---|---|---|
| WT | 294 ± 12 | 76 ± 9 | 3.9 ± 0.3 |
| E288A | 0.16 ± 0.02 | 1.5 ± 0.5 | 0.11 ± 0.02 |

The nucleophile mutant, D229A, was completely unable to cleave pNPαFuc even when up to 70 μg of enzyme was used (maximum before precipitation), as would be expected for a mutant at this position. Small molecule rescue studies for both acid/base and nucleophile are described in section 4.8.

## 4.7 Kinetic characterization of BT4136fuc and BT1625fuc (Group S2, Class B)

Both BT4136 and BT1625 are in Class B of the phylogenetic tree and are in Group S2 since both their modeled structures align with BiAfcB's crystal structure (free enzyme). From these observations, it is predicted that these two fucosidases undergo a conformational change upon substrate binding and that E249 (numbering is the same for BT4136 and BT1625) is the acid/base residue. The five aspartate and glutamate residues initially chosen for mutation were

as follows: D286, E294, E298, E305, and E307. Residues, E305, D286, and D307 were chosen because E305 residue aligns structurally with the acid/base residue found for Hfuc and all three residues are also reasonably conserved in sequence alignments. While not well conserved by sequence, residues E294 and E298 were chosen because they are in a region with a string of aromatic residues (FWY) which could likely be involved in hydrophobic stacking interactions with the substrate. Kinetic parameters for hydrolysis of pNPαFuc by the alanine mutants on pNPαfuc are reported in Table 4.8. Mutations of D286A and E298A, and E305 do not significantly affect the activity or binding of the substrate, therefore it is unlikely these residues are directly involved in catalysis or binding. A ten-fold decrease in activity and slight increase in $K_M$ was observed upon mutation of D307 to an alanine residue; however, no rescue was observed using azide or formate anions (up to 2 M azide and 1 M formate). Also, since this residue appears to be located in the active site, mutation of the residue might lead to slight disruptions in structure or hydrogen bonding between residues or binding of the substrate.

***Table 4.8*** *Hydrolysis of pNPαFuc by mutants of BT4136 and BT1625 (at 37°C and pH 7.0)*

| enzyme | $k_{cat}$ ($min^{-1}$) | $K_M$ (μM) | $k_{cat}/K_M$ ($min^{-1}$/μM) |
|---|---|---|---|
| BT4136WT | 265 ± 12 | 45 ± 4 | 5.9 ± 0.7 |
| D286A | 258 ± 13 | 56 ± 6 | 4.6 ± 0.5 |
| E305A | 253 ± 10 | 26 ± 3 | 9.7 ± 0.6 |
| E298A | 272 ± 16 | 84 ± 7 | 3.3 ± 0.2 |
| D307A | 36 ± 4 | 33 ± 2 | 1.1 ± 0.1 |
| E294A | NE | NE | -- |
| D204A[a] | NA | NA | NA |
| E249A[b] | 0.41 ± 0.04 | 230 ± 13 | (1.8 ± 0.4) x $10^{-4}$ |
| BT1625WT | 217 ± 11 | 32 ± 7 | 6.8 ± 0.2 |
| D286A | 115 ± 9 | 64 ± 6 | 1.8 ± 0.2 |
| E305A | 104 ± 10 | 27 ± 3 | 3.8 ± 0.3 |
| E298A | 116 ± 12 | 76 ± 8 | 1.5 ± 0.1 |
| D307A | 15.8 ± 3 | 41 ± 3 | 0.40 ± 0.05 |
| E294A | NE | NE | -- |
| D204A[a] | NA | NA | NA |
| E249A[b] | 0.23 ± 0.03 | 359 ± 24 | (6.4 ± 0.1) x $10^{-4}$ |

NE = not expressible; protein does not express under a variety of expression conditions; NA = no activity observed at up to 150 μg protein; [a] nucleophile residue; [b]proposed acid/base residue

Mutation of the nucleophile residue, D204, to an alanine gave rise to a completely inactive enzyme even when up to 150 μg of enzyme was used (maximum before precipitation), consistent with its role; small molecule rescue studies for both the acid/base and nucleophile are described in section 4.8. Mutation of the proposed acid/base residue, E249, to alanine gave rise to an enzyme with a $k_{cat}/K_M$ value 32000-fold (for BT4136) and 10000-fold (for BT1625) lower than that observed with wild type for the hydrolysis of pNPαfuc. Similar to BT2970,

comparison of the $k_{cat}$ values for the hydrolysis of pNPαFuc and αFucF by BT4136fuc and BT1625fuc show that deglycosylation is likely rate limiting. Again for BT4136fuc and BT1625fuc, the acid/base mutant cleaves αFucF more efficiently than it cleaves pNPαFuc (comparison of $k_{cat}/K_M$), and a low $K_M$ is seen for αFucF due to intermediate accumulation (Table 4.9).

***Table 4.9*** *Hydrolysis of FucF by BT4136fuc (WT and E249A) and by BT1625fuc (WT and E249A)*

| | $k_{cat}$ ($min^{-1}$) | $K_M$ (μM) | $k_{cat}/K_M$ ($min^{-1}$/μM) |
|---|---|---|---|
| BT4136fuc WT | 242 ± 13 | 62 ± 5 | 3.9 ± 0.2 |
| BT4136fuc E249A | 0.39 ± 0.02 | 1.7 ± 0.2 | 0.22 ± 0.03 |
| BT1625fuc WT | 208 ± 6 | 64 ± 8 | 3.3 ± 0.1 |
| BT1625fuc E249A | 0.21 ± 0.01 | 0.92 ± 0.06 | 0.22 ± 0.03 |

## 4.8 Small molecule rescue of acid/base (and nucleophile) mutants

A common diagnostic tool for identification of the nucleophile and acid/base residue of retaining glycosidases is the rescue of the activity of mutants (at those positions) by exogenous nucleophilic anions. Azide rescue has been observed for the nucleophile and acid/base mutants of other characterized GH29 α-L-fucosidases. Investigation of the roles of the putative acid/base residues, E288 (for BT2970) and E249 (for BT4136 and BT1625), in catalysis were performed through small molecule rescue experiments of each of their alanine mutants in the hydrolysis of pNPαFuc. Rescue of the nucleophile mutants, while not critical to this study, is summarized in Table 4.10 and showed that activity could be rescued as expected. More importantly, a 3-fold, 4.5-fold and 4.5-fold increase in $k_{cat}$ for the hydrolysis of pNPαfuc by the E288A, BT4136fuc E249A, and BT1625 E249A mutants respectively, was seen at 200 mM

sodium azide (Table 4.9). The small amount of rescue observed is not unusual for α-glycosidases but could be explained by glycosylation becoming rate limiting in the presence of azide for substrates with relatively poor leaving groups such as pNP and supports deglycosylation being at least partially rate determining for pNPαFuc.[179,180] Azide "rescue" was not seen for any of the other mutants and no increase (or decrease) in wild type activity was observed in the presence of azide. The presence of azide rescue provides strong support for the assignment of these proposed acid/base residues. TLC analysis of the product(s) obtained from the hydrolysis of pNPαfuc by the acid/base mutants (E249A of BT4136 and E288A of BT2970) show that a new product (RF = 0.71) is indeed formed upon addition of azide, which runs at a different $R_F$ than L-fucose ($R_F$ ~0.46) and pNPαFuc ($R_F$ = 0.81). After purification, mass spectral analysis confirmed that a product which has a molecular formula consistent with fucosyl azide was formed: Calculated for $C_6H_{11}N_3O_4Na^+$: 212.0593 Found: 212.0596.

**Table 4.9** Azide rescue of the nucleophile and acid/base residue in the hydrolysis of pNPαFuc (37$^{o}$, pH 7.0, 200mM azide)

| Enzyme | Exogenous nucleophile | $k_{cat}$ ($min^{-1}$) | $K_M$ (μM) | $k_{cat}/K_M$ ($\mu M^{-1}$ $min^{-1}$) |
|---|---|---|---|---|
| BT2970 E288A[b] | None | 0.18 ± 0.02 | 391 ± 3 | (4.7 ± 0.6) x $10^{-4}$ |
| BT2970 E288A | 200 mM Azide | 0.63 ± 0.04 | 719 ± 2 | (8.7 ± 0.3) x $10^{-4}$ |
| BT2970 D229A[a] | None | ND | ND | ND |
| BT2970 D229A | 200 mM Azide | 0.16 ± 0.04 | 31 ± 4 | (5.2 ± 0.6) x $10^{-3}$ |
| | | | | |
| BT4136 E249A[b] | None | 0.41 ± 0.04 | 230 ± 13 | (1.8 ± 0.4) x $10^{-3}$ |
| BT4136 E249A | 200 mM Azide | 1.8 ± 0.2 | 639 ± 3 | (2.8 ± 0.2) x $10^{-3}$ |
| BT4136 D204A[a] | None | ND | ND | ND |
| BT4136 D204A | 200 mM Azide | 0.25 ± 0.02 | 52 ± 2 | (4.8 ± 0.4) x $10^{-3}$ |
| | | | | |
| BT1625 E249A[b] | None | 0.23 ± 0.03 | 359 ± 24 | (6.4 ± 0.1) x $10^{-4}$ |
| BT1625 E249A | 200 mM Azide | 1.0 ± 0.1 | 741 ± 3 | (1.3 ± 0.2) x $10^{-3}$ |
| BT1625 D204A[a] | None | ND | ND | ND |
| BT1625 D204A | 200 mM Azide | 0.06 ± 0.01 | 52 ± 3 | (1.1 ± 0.2) x $10^{-3}$ |

ND = not detected

[a]nucleophile mutants; [b]acid/base mutants

## 4.9 Conclusions

GH29 α-L-fucosidases can be placed into three structural groups; group S1: a group in which no major conformational change occurs upon substrate binding (complete structural overlap of the acid/base residue with Tmfuc's acid/base residue); Group S2: a group in which a conformational

change similar to that observed for BiAfcB upon ligand binding exists, bringing the acid/base residue from a distance of >15 Å to within ~5.5 Å of the nucleophile (the acid/base then aligns with Tmfuc); and Group S3: a group in which a different conformational change upon binding is predicted—this group consists of only the eukaryotic fucosidases and no structural overlap of the acid/base residue with Tmfuc is observed. The structural classifications above do not correlate completely with the phylogenetic classification, though the eukaryotic fucosidases show alignment both structurally and by sequence for the acid/base residue. Class B fucosidases also show sequence alignment for the acid/base residue (barring the most divergent, i.e. last two, fucosidases in the group). The comparison between sequences and structures helped to identify the acid/base residue for Class B GH29 fucosidases. From the results the acid/base residue for GH29 can be easily predicted by SSM-superimposing either the crystal structure or PHYRE2 modeled structure with Tmfuc or BiAfcB and finding the overlapping acid/base residue. All eukaryotic (except plant) fucosidase acid/base residues can likely be predicted by comparison to the model of Hfuc. For fucosidases that are either in Class B of the phylogenetic tree (with the exception of the most divergent members) or Class C of the phylogenetic tree, a simple sequence comparison shows conservation of the acid/base residue within each group. Experimental verification for the acid/base residue was shown through mutagenesis of the catalytic glutamate to alanine and kinetic characterization of these mutants. The alanine mutants of the acid/base residue showed a much lower kcat (or $k_{cat}/K_M$) than wild type and also showed a decrease in $K_M$ (for substrates with a good leaving group), suggesting the accumulation of intermediate. Azide rescue of the mutants of the acid/base and nucleophile further support their assignment. Additional support for the assignment of

acid/base residue in these fucosidases is given by the crystal structures of BT2970fuc and BiAfcB (with bound deoxyfuconojirimycin) which show the corresponding acid/base residues within 5.5 Å of the nucleophile and the latter structure, which confirms the hypothesis of a conformational change (induced-fit). Additionally, accumulation of the covalent fucosyl-enzyme intermediate was observed for BT2970 with the E288Q mutant as would be expected for the mutation of the acid/base residue to a non-catalytically competent residue. The combination of modeling studies and structural analysis and verification of the results through kinetic characterization provides a greater understanding of the GH29 catalytic mechanism; which also allows the manipulation of these fucosidases as glycosynthases or for hydrolysis of fucose-containing oligosaccharides. Finally, the use of structural information and PHYRE2 models to aid in the successful prediction of catalytic residues is extremely useful in cases where sequence alignments fail and also helps visualize their mechanism.

# Chapter 5: Materials and Methods

## 5.1 Protein chemistry

### 5.1.1 Materials, bacterial strains, plasmids, and media

Primers were synthesized by Integrated DNA Technologies (IDT) (Iowa State). DNA sequencing was performed by the Nucleic Acid and Protein Service (NAPS) Unit at the University of British Columbia using an Applied Biosystems PRISM 377 automated sequencer. Bacterial transformations (in either electrocompetent or chemically competent cells) were performed according to standard protocols. All DNA fragments were purified using agarose gel with SYBR® Safe (Invitrogen) DNA gel stain added to visualize DNA. The DNA fragments were extracted from the gel using QIAquick DNA gel purification kits (Qiagen, CA). Plasmids were prepared using single-colony overnight cultures and purified using Qiaprep Spin Miniprep kits (Qiagen, CA). Restriction endonucleases were supplied by Fermentas and used with the 1X Tango buffer (Fermentas) optimal for double digest with NheI, NdeI and XhoI.

### 5.1.2 Cloning and site-directed mutagenesis

#### *5.1.2.1 Cloning and site-directed mutagenesis of YesZ (GH42 β-galactosidase from B. subtilis)*

The gene coding for YesZ was amplified by PCR from genomic B. subtilis DNA (ATCC 23857D-5) using primers Bsu_yes_fw (C ACC ATG AGA AAA CTG TAT CAT GGC GCT TG) and Bsu_yes_rv (GAG ATT GTC AAA TTG AAT CAC ACG GTA TTC). Pwo polymerase was used with standard PCR

conditions for 30 cycles and the resulting fragment was cloned into pET101 using the Champion pET Directional TOPO Expression Kit (Invitrogen, Burlington, ON) and named ‘pET Bsu YesZ wt’.

The mutation at the acid/base position (E145A) was introduced using a four-primer methodology. Outside primers were the same as above, the mutational primers were Bsu_yes_E145Afw (GGC TGA TCG GAT GGC AGC TAG ACA ATG CGT TCA A) and Bsu_yes_E145rv (ATT GTC TAG CTG CCA TCC GAT CAG CC). Two fragments, each containing the desired mutation, were amplified in 15 cycles using the primer combinations Bsu_yes_fw/Bsu_yes_E145Arv and su_yes_rv/Bsu_yes_E145Afw and ’pET Bsu YesZ wt’ as template. These fragments were purified and in a second step the derived fragments were combined and amplified using the primer combination Bsu_yes_fw/Bsu_yes_rv and the resulting mutated fragment cloned into pET101 using the above described technology and named ‘pET Bsu YesZ E145A’. The insert of the derived plasmid was completely sequenced and contained the desired mutation. The constructs encoding the β-galactosidase gene and the E145A (acid/base) mutant in pET101 vector were transformed into Escherichia coli BL21 DE3 electrocompetent cells via electroporation. This work was performed by Dr. Johannes Müllegger.

*5.1.2.2 Cloning and site directed mutagenesis of EABase (GH 98 endo-β-galactosidase)*

The gene coding for EABase was amplified by PCR from genomic *C. perfringens* DNA (ATCC 10543) using primers containing restriction sites NheI and XhoI as follows: EABase-fw (5'-GGT ATG GAA GTT TAT GCT AGC TTG GAA GAA AGC AG-3') and EABase-rv (5'-CCG CTC GAG CTT AAT TAC AAT ATC AAA ATC-3'). Standard PCR conditions were used for 30 cycles with EXPAND high-fidelity polymerase, and the 2.3 kbp insert was digested (NheI/XhoI) and ligated into pET-21a vector similarly digested and also treated with shrimp alkaline phosphatase. Plasmids obtained from selected observed colonies were submitted for sequencing to the NAPS unit at UBC, and plasmids containing the full length gene with no errors were used in subsequent experiments.

Two methods, the Quikchange method and the two step, 4-primer method, were used to generate mutants. The only method that was able to generate D453A and E506Amutants was the two-step, four primer method. This is most likely due to the self-complementarity of the primers used and possible hairpin formation of primers that had a Tm of 65-68 $^{o}$C. For the E506A mutant, pET-21aWT was used as the template and two separate PCR reactions containing the following primer combinations in the first amplification: E506Afw, 5'-ATA TAA TTT TGC ACA TCC AGC AT-3' with T7term and E506Arev, 5'-ATGCTG GAT GTG CAA AAT TAT ATA CAC-3'. The amplified fragments from this were used as templates in the second PCR reaction using the T7 and the T7term primers. This generated a 2.5 kbp full length gene, which was then subcloned into pET-21a vector using the NheI and XhoI restriction sites. A similar protocol was followed with D453A. The E354A mutant was generated using the QuikChange (Stratagene)

method with the following primers (E354Afw, 5'-GGT GTA TTT AGT ACA GCA AAT TAT TGG GTT TGG-3' and E354Arev, 5'-CTG TCC AAA CCC AAT AAT TTG CTG TAC TAA ATA CAC C-3') using pET-21aWT as the template. In the QuikChange method, a complementary pair of oligodeoxyribonucleotides containing the desired mutation is used as the primers in a primer extension of the wild type plasmid where the generated plasmid contains the mutation of interest. DpnI digestion of the wild type template followed by transformation into *E. coli* R1360 yielded colonies containing the mutation.

### *5.1.2.3 Cloning and site-directed mutagenesis of GH 29 α-L-fucosidases from Thermotoga maritima and Bacteroides thetaiotaomicron (2790, 4136, and 1625)*

The gene coding for Tmfuc was amplified by PCR from genomic *T. maritima* DNA (ATCC 43589) using primers containing restriction sites NdeI and XhoI as follows: Tmfuc-fw (5'-GCC CAT ATG ATT TCT ATG AAA CCC CG -3') and Tmfuc-rv (5'-GCC TCG AGT TAC TTA ATT ACA ATA TC -3'). Standard PCR conditions were used for 35 cycles with Pfu polymerase, and the 1.3 kbp insert was digested (NdeI/XhoI) and ligated into pET-22b vector similarly digested and also treated with shrimp alkaline phosphatase. Plasmids obtained from selected observed colonies were submitted for sequencing to the NAPS unit at UBC, and plasmids containing the full length gene with no errors were used in subsequent experiments.

The four-primer method was used to generate all Tmfuc mutants (D224A, E266A, and E281A) (as the QuikChange method failed, again likely due to secondary structure formation). The For the D224A mutant, pET-22bWT was used as the template and two separate PCR reactions containing the following primer combinations in the first amplification: D224Afw, 5'-CCC GAC

GTT CTC TGG AAC GCC ATG -3’ with T7term and D224Arev, 5’-CCT TCC TCC GGC CAG CCC ATG GCG -3’. The amplified fragments from this were used as templates in the second PCR reaction using the T7 and the T7term primers. This generated a 1.3 kbp full length gene, which was then subcloned into pET-22b vector using the NdeI and XhoI restriction sites. A similar protocol was followed with D453A. Table 5.1 lists all the primers used for generating the Tmfuc mutants (along with all BT mutants).

Plasmids containing the D229G mutation and the E288Q mutation were made and sent to us by Dr. Alicia Lammerts van Bueren. In order to generate other mutants and compare activity to wild type α-fucosidase, site directed mutagenesis was performed to reverse the nucleophile mutant back to wild type. Mutagenesis was performed using the QuikChange method using the following primer combination with the D229G mutant plasmid: BT2970WTfw, 5’- CCG ACC GTT AAG GAC TTC TGG -3’ and BT2970WTrev, 5’- CCA GAA GTC CTT AAC GGT CGG -3’. Dpn1 digestion of the purified product followed by transformation into R1330 cells and subsequent sequencing of the purified plasmid showed that the wild type sequence was successfully cloned and no undigested nucleophile mutant colonies were present. The wild type fucosidase (along with all mutants) are in a custom pET28a LIC plasmid (used by Lammerts van Bueren). This wild type BT2970 plasmid was then used to generate the E288A and E306A mutants through the QuikChange method using the forward and reverse primers listed in Table 5.1.

*Table 5.1: Listing of all the forward and reverse primers used for generating mutants using the Quikchange method*

| Primer name | Sequence (5'→ 3') |
|---|---|
| Tmfuc E266A forward (fw) | GAT TTC AAA ACG GCG TAC CAC |
| Tmfuc E266A reverse (rev) | CCC CGG ATA GTT CAC GTG GTA CGC |
| Tmfuc E281A fw | GCC GGG CTA CAA ATG GGC GTT TAC |
| Tmfuc E281A rev | CCC TAT TCC CCT CGT AAA CGC CCA |
| BT2970 D224A fw | CCG ACC GTT AAG CAC TTC TGG |
| BT2970 D224A rev | CCA GAA GTG CTT AAC GGT CGG |
| BT2970 E288A fw | CCG GCT ACG CAC GCC GCT TGC C |
| BT2970 E288A rev | GGC AAG CGG CGT GCG TAG CCG G |
| BT2970 E306A fw | GTG GGA CTG GGC AGC CTG CAT G |
| BT2970 E306A rev | CAT GCA GGC TGC CCA GTC CCA C |
| BT4136 D204A fw | GTG TGG TTT GCC GGG GCA AAC GG |
| BT4136 D204A rev | CCG TTT GCC CCG GCA AAC CAC AC |
| BT4136 D286A fw | GGA CAG TCA AAT GCT TTA GGA AGC CGC |
| BT4136 D286A rev | GCG GCT TCC TAA AGC ATT TGA CTG TCC |
| BT4136 E305A fw | GGT ATC CCT CGG CAG TAG ATG TTT CC |
| BT4136 E305A rev | GGA AAC ATC TAC TGC CGA GGG ATA CC |
| BT4136 E298A fw | GGA GAA AGC GAC AGC GTT ATT TTG GTA TCC C |
| BT4136 E298A rev | GGG ATA CCA AAA TAA CGC TGT CGC TTT CTC C |
| BT4136 D307A fw | CCC TCG GAA GTA GCT GTT TCC ATA CGC CC |
| BT4136 D307A rev | GGG CGT ATG GAA ACA GCT ACT TCC GAG GG |
| BT4136 E294A fw | CCG CAA GAT GTT GGC GAA AGC GAC AGA G |
| BT4136 E294A rev | CTC TGT CGC TTT CGC CAA CAT CTT GCG G |
| BT4136 E249A fw | CTG GGT AGG CAA CGC AAG CGG ATT AGG TCG |
| BT4136 E249A rev | CGA CCT AAT CCG CTT GCG TTG CCT ACC CAG |
| BT1625 D204A fw | GTG TGG TTT GCA GGC GCA AAC GG |
| BT1625D204A rev | CCG TTT GCG CCT GCA AAC CAC AC |
| BT1625D286A fw | GCA AAG CCG AAG CTT TAG GAA GCC G |
| BT1625D286A rev | CGG CTT CCT AAA GCT TCG GCT TTG C |
| BT1625E305A fw | GGT ATC CTT CGG CAG TAG ACG TGT C |
| BT1625E305A rev | GAC ACG TCT ACT GCC GAA GGA TAC C |
| BT1625E298A fw | GGA AAA GGC AAC GGC GTT ATT CTG G |
| BT1625E298A rev | CCA GAA TAA CGC CGT TGC CTT TTC C |
| BT1625D307A fw | CGG AAG TAG CAG TGT CTA TCC |
| BT1625D307A rev | GGA TAG ACA CTG CTA CTT CCG |
| BT1625E294A fw | GCC GTG CGA TGC TGG CAA AGG CAA CGG AG |
| BT1625E294A rev | CTC CGT TGC CTT TGC CAG CAT CGC ACG GC |
| BT1625E249A fw | GGG TAG GTA ATG CAA AAG GAT TGG G |
| BT1625E249A rev | CCC AAT CCT TTT GCA TTA CCT ACC C |

Wild type plasmid for both BT1625 and BT4136 α-L-fucosidases were supplied by Dr. Alicia Lammerts van Bueren along with primers for the nucleophile (D204A) mutants; the numbering is the same for both BT1625 and BT4136. These enzymes are again, all in custom pET28a LIC plasmids. The QuikChange method was used to generate the E286A, E294A, E305A, D307A and E249A mutants (acid/base candidates) using the forward and reverse primer combinations found in Table 5.1. Dpn1 digestion of the purified product followed by transformation and sequencing as mentioned above showed that the desired mutation was present in each case.

### 5.1.3 Expression of wild type and mutant glycosidases

The general protocol for the expression of glycosidases studied in this thesis is described below. No major deviations from this method occurred in any of the expressions though in several cases small-scale expression was performed to determine the optimal expression temperature (which turned out to be 37$^{o}$C for all enzymes studied in this thesis). All plasmids containing the gene for the enzymes of interest were transformed into *E. coli* BL21(DE3) cells by electroporation using the Genepulser II (Bio-Rad Inc.). For the production of each enzyme, a single colony was inoculated into 50 mL of LB medium containing 100 µg/mL ampicillin (or kanamycin) and grown overnight. From this overnight culture, 5 mL was inoculated into 1 L of LB medium containing 100 µg/mL ampicillin (or kanamycin). Cells were induced with 0.5 mM IPTG once they reached an $OD_{600}$ of 0.7 and centrifuged at 5000 g for 20 minutes after 4 hours (or overnight in some cases) of induction.

### 5.1.4 Purification of wild type and mutant glycosidases

The cell pellets were resuspended in 50 mL of lysis buffer (20 mM Tris at pH 7.9, 0.5 M NaCl, and 20 mM imidazole) and lysed using a French press. EDTA-free protease inhibitor was added after the lysis step, and the lysed cell mixture was centrifuged at 15000 g for 30 min. The filtered supernatant was added to a 1.0 mL (or 5.0 mL) HisTrapFF column and eluted using a gradient of 0.05-0.30 M imidazole on the Fast Protein Liquid Chromatography (FPLC) equipment (Pharmacia). Unbound proteins were washed using washing buffer (containing 0.05 M imidazole) and the elution of the desired protein was visualized as a defined peak at $A_{280}$. Fractions containing the enzyme of interest were concentrated using a 10-30K molecular weight cut-off centrifugal filter (Amicon Ultra; Millipore) and dialyzed against a suitable buffer overnight. For EABase, the dialyzed protein was applied to a MonoQ column and eluted with a NaCl gradient of 0-300 mM NaCl. A broader peak containing EABase (eluted at 0.1 M NaCl) was concentrated and run on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) to afford >95% pure protein. All enzymes were concentrated and run on SDS-PAGE gels to determine approximate purity and the approximate molecular weights were determined by comparison to pre-stained molecular weight markers (Bio-Rad). Repeat column-purification was performed if the enzymes were not >95% pure. Fresh columns were used for each mutant to avoid cross-contamination. The enzymes were concentrated and buffer exchanged into appropriate buffers and stored at $4^{o}$C. The Bradford assay was used for the determination of concentration of proteins.

### 5.1.5 CD spectroscopy

CD spectra were recorded on a JASCO spectropolarimeter in a 0.1 cm cuvette. Enzyme concentrations were standardized to be in the range of 0.2-0.3 µg/µL in 20mM sodium acetate buffer (pH 6). Samples were analyzed in scanning mode from 230 to 197 nm. Sample cuvettes were cleaned extensively with water and buffer between runs and a blank run consisting of 20 mM sodium acetate (pH 6.0) was performed after cuvettes were washed between different enzyme samples to ensure the CD spectra obtained were free of contamination.

### 5.1.6 Labelling and proteolysis (YesZ)

A sample of the E145A (acid–base) mutant of YesZ (100 µl of 0.2 mg/mL in 50 mM sodium phosphate, pH 7.0) was incubated with 5 µL of DNP2FGal (3 mM in water) for 2 h at 37 $^{o}$C. Complete inactivation was evidenced by assay of a 20 µl aliquot of the enzyme/inhibitor mixture with 3 mM PNPGal. The intact protein and control sample (as above with 5 µL H2O instead of inhibitor) were subjected to proteolysis by diluting the samples with 10 µl of a pepsin solution (0.1 mg/ml in 50 mM phosphate buffer, pH 2.0) and 60 µL of 500 mM phosphate buffer, pH 2.0 and incubated for 3 h at room temperature. The sample was then frozen until ready for analysis and was used immediately upon thawing.

### 5.1.7 Electrospray ionization mass spectrometry

Mass spectra were recorded using an ABI MDS-SCIEX API QSTAR Pulsar i mass spectrometer (Sciex, Thornhill, ON). Peptides were separated by reverse phase C18 column on an Ultimate HPLC system (LC Packings, Amsterdam, Netherlands) interfaced with the mass spectrometer. For LC/MS experiments, proteolytic digests of proteins were loaded onto a C18 column (LC Packings, 75 µm i.d. · 150 mm PepMap) and eluted with a gradient of 2–40% solvent B over the course of 60 min at a flow rate of 0.2 µL/min (Solvent A, 0.1% formic acid and 2% acetonitrile in water; solvent B, 0.1% formic acid and 85% acetonitrile in water). The TOF data were acquired over a mass-to-charge ratio range of 300–2000 amu, with a step size of 0.1 amu and with a scan time of 1 s. The ion source potential was set at 2.2 kV; the rifice energy was 50 V. Intact proteins were run on a C4 column (PLRP-S 0.75 lm i.d. · 50 mm, 8 lm, 4000A, Michrom BioResources, Inc., Auburn, CA, USA) on an Ultimate HPLC system and interfaced to the mass spectrometer as above. They were eluted with a gradient of 15% solvent B to 80% solvent B over 15 min, 80% solvent B for 5 min and finally 15% solvent B for 15 min. To determine the amino acid sequence, the mass spectrometer was operated in an IDA (Information Dependent Acquisition) MS/MS mode. These experiments were performed by Dr. Shouming He.

## 5.2 Enzyme kinetics

### 5.2.1 Michaelis-Menten kinetics: spectrophotometric measurements

#### *5.2.1.2 YesZ and its mutants*

All experiments were carried out at 37 $^{o}$C in 50 mM phosphate buffer, pH 7.0 containing 0.1% BSA. Kinetic parameters were determined by measuring the initial linear increase in absorbance at 400 nm upon addition of the enzyme (final concentration 5.7 nM) to substrate paranitrophenyl β-D-galactopyranoside (PNPGal) at a range of concentrations (typically 0.05–3 mM). Cuvettes had a path length of 1 cm and were used in a Cary 4000 UV/visible spectrophotometer connected to a circulating water bath. The rates were calculated using an extinction coefficient for p-nitrophenol at pH 7.0, 37 $^{o}$C of 9050 $M^{-1}$ $cm^{-1}$ via a direct fit of the data to the Michaelis–Menten equation using the program GraFit 4.0 (Erithacus software). For the chemical rescue experiments, 10 mM – 2 M of sodium azide was added to the assays and the pH of the solution was checked to ensure that it remained at 7.0.

#### *5.2.1.3 EABase and its mutants*

The reaction progress of the EABase-catalyzed hydrolysis of DNP-A-trisaccharide was monitored at 37 $^{o}$C using a continuous spectrophotometric assay on a Cary-4000 spectrometer (Varian Inc.) connected to a circulating water bath. Unless otherwise indicated, assays were performed in 20 mM sodium acetate buffer at pH 6.0 in cells of 1 cm path length. Prewarmed 200 µL

cuvettes were loaded with an appropriate volume of water, buffer, and substrate at concentrations from 0.1 to 5 times $K_M$ and incubated for 6 min before data were acquired. Spontaneous hydrolysis of substrates was monitored prior to the addition of enzyme. The change in absorption was monitored at 405 nm and apparent kinetic parameters were generated by GraFit 4.0 (Erithacus Software). An absorption coefficient of 8584 $M^{-1}$ $cm^{-1}$ was determined for 2,4-dinitrophenolate in 20 mM sodium acetate buffer at pH 6.0 at 37 $^{o}$C. Similar procedures were used for the kinetic analysis of mutants. Attempts at azide rescue were performed under similar conditions to those described earlier.

#### *5.2.1.4 α-L-Fucosidases from T. maritima and Bacteroides thetaiotaomicron (2970, 4136, and 1625)*

All experiments were carried out at 37 $^{o}$C using a continuous spectrophotometric assay on a Cary-4000 spectrometer (Varian Inc.) connected to a circulating water bath. Prewarmed 200 µL cuvettes (1 cm path length) were loaded with an appropriate volume of water, buffer, and substrate at concentrations from 0.1 to 5 times $K_M$ and incubated for 5 min before data were acquired. Spontaneous hydrolysis of substrates was monitored prior to the addition of enzyme. Michaelis-Menten parameters for hydrolysis of pNPαfuc was measured by monitoring the release of the phenol at 400 nm (or 360 nm for Tmfuc). For Tmfuc, kinetic parameters were obtained in the following buffer: 50 mM sodium citrate containing 150 mM sodium chloride and 0.1% BSA (pH 5.0). An extinction coefficient ($A_{360}$) of 1816 $M^{-1}$ $cm^{-1}$ was determined for p-nitrophenolate under these conditions. The buffer used for all *B. thetaiotaomicron* fucosidases was 50 mM sodium phosphate buffer containing 150 mM sodium chloride and 0.1% BSA. An

extinction coefficient of 7450 $M^{-1}$ $cm^{-1}$ was determined for p-nitrophenolate under these conditions. The apparent kinetic parameters were generated by GraFit 4.0 (Erithacus Software). Azide rescue was performed under similar conditions to those described earlier.

### 5.2.2 Michaelis-Menten kinetics: fucosyl fluorides

Determination of the rate of hydrolysis of αFucF by Tmfuc, BT2970fuc, BT4136fuc, BT1625fuc and mutants (for which kinetic data is listed) was performed by following the increase in fluoride ion concentration using the Orion 96-09 combination fluoride ion electrode connected to a computer running the LoggerPro software (Vernier Software Ltd). Glass vials containing various concentrations of αFucF were incubated at 25$^{o}$C or 37$^{o}$C to establish a steady-state spontaneous hydrolysis rate before the enzyme was added (to a final concentration of 750 µL). Background-hydrolysis corrected initial rates were used to determine the kinetic parameters which were obtained by direct fit of the data in GraFit 4.0 (Erithacus Software).

### 5.2.3 Michaelis-Menten kinetics: reducing sugar assay (EABase)

The reaction progress of the EABase-catalyzed hydrolysis of A-pentasaccharide was monitored using a stopped reducing sugar assay. The stop (quenching) solution (HBAH solution) (50 mL) consisted of 4% NaOH, 0.1 M p-hydroxybenzoic acid hydrazide (HBAH), 0.1 M sodium sulfite, 50 mM trisodium citrate dihydrate, and 20 mM calcium chloride. To reaction mixtures (0.5 mL) containing buffer and substrate at concentrations varying from 0.1 to 5 times $K_M$ was added

wild type EABase, and the reactions were incubated at 37 $^{o}$C. Controls which did not contain EABase were also run. Aliquots (50 µL) were removed at 15 min intervals and added to 100 µL of the HBAH solution. These samples were then boiled for 12 min, cooled to room temperature, and their absorbance at 420 nm was recorded by adding 100 µL of this solution to 900 µL of water in a 1 cm plastic cuvette. The measured absorbance was plotted as a function of time, and the initial rates were used to generate the apparent kinetic parameters using GraFit 4.0 (Erithacus Software).

### 5.2.4 pH profiles

The stability of the enzyme at each pH value over the assay time was first studied by incubating the enzyme in buffers of different pH values from 4.0-9.0. After 15 min, 20 µL of the mixture was removed for assay by adding to another preincubated solution containing the substrate in the enzyme's standard assay buffer (pH 7.0 for YesZ, 6.0 for EABase, 5.0 for Tmfuc and 6.5 for BT fucosidases). Subsequent measurements of activity were only performed at those pH values where the enzyme retained more than 95% activity over the 10-min period. The buffers used are as follows: pH 4.0-6.0, 20 mM sodium citrate; pH 6.5-8.0, 20 mM phosphate; pH 8.0-9.0, 20 mM Tris-HCl; pH 9.5-10, glycine/NaOH.

The $k_{cat}/K_M$ values for the hydrolysis of DNP-A-trisaccharide by EABase and pNP-β-galactoside by YesZ at each pH were determined by the substrate depletion method, following reaction time-courses at low (≥0.2 x $K_M$) substrate concentrations. A final concentration of 1-5 µg of

enzyme was added to each cuvette, and the release of dinitrophenolate was monitored continuously at 405 nm for 5-10 min at which time 5-7 half-lives had passed. After the reaction was judged to be complete, the pH of the resulting mixture was checked to ensure that no significant change in pH had occurred. The resulting change in absorbance as a function of time was fitted to a first order rate equation using GraFit 4.0, yielding values for the pseudo-first order rate constant at each pH value as follows.

At low substrate concentrations ([S] << $K_M$), the reaction rates are given by the following equation:

$$v = \frac{k_{cat}[E]_o[S]}{K_M}$$

The $k_{obs}$ values correspond to $k_{cat}$[E]o/$K_M$. Thus, division of the obtained rate constants by the enzyme concentration gives the $k_{cat}/K_M$ values. By analyzing the bell-shaped $k_{cat}/K_M$ versus pH plots using GraFit 4.0, two apparent pKa values of ionizable groups were assigned.

For the GH 29 α-L-fucosidases, since the $K_M$ was already low for the hydrolysis of pNPαfuc the substrate depletion method described above could not be used. Instead, the pH dependence of these enzymes was determined by performing standard Michaelis-Menten analysis at pH increments of 0.5 units.

### 5.2.5 Inactivation kinetics

Studies of inactivation kinetics were performed by pre-incubating 100 µl of the enzyme at 37 $^{o}$C with DNP2FGal at a range of concentrations from 0.005 mM to 0.3 mM in a total volume of 140 µl. The concentrations of enzyme in these inactivation mixtures were: WT, 41.0 nM; E145A, 14.1 µM. Aliquots (20 µl) were withdrawn at regular time intervals and added to 180 µL of pNPGal that had been pre-incubated in 50 mM phosphate buffer, pH 7.0 at 37 $^{o}$C. The initial rates at each time point were plotted as a function of time to obtain time dependent exponential decay curves from which $k_{obs}$ could be obtained for each inactivator concentration using the first-order rate equation:

$$rate = A_0 e^{-kt} + offset$$

where (k = $k_{i,obs}$), in GraFit 4.0. Since the rates did not decay to zero, the single exponential decay with offset was used. Values of the inactivation parameters $K_i$ and $k_i$ were obtained from the fits of a plot of $k_{obs}$ versus inactivator concentration to the equation: $k_{obs} = \frac{k_i \times [\mathrm{I}]}{(K_I + [\mathrm{I}]}$

### 5.2.6 Protection studies

Protection studies were performed by measuring the initial reaction velocity of the hydrolysis of pNP-β-Gal by YesZ at five different concentrations (from 0.75-15 mM) of pNP-β-Gal in the presence and absence of IPTG. A Lineweaver-Burk plot of these initial reaction velocities in the presence and absence of IPTG was constructed (Appendix B). Conditions of this assay are similar to those for the kinetic characterization of YesZ. The two lines intersect on the y-axis, which is characteristic for competitive inhibitors. The apparent $K_i$ value was calculated using the following equation with the slopes obtained from the Lineweaver-Burk plots:

$$K_i = \frac{[\mathrm{I}]}{\frac{slope_i}{slope_o} - 1}$$

### 5.2.7 Reactivation kinetics

YesZ which was 85% inactivated by 2FDNPβGal was freed of excess inactivator by 10-fold dilution with buffer and subsequent concentration at 4$^{o}$C using a 10 kDa molecular weight cut-off centrifugal concentrator (Amicon). This procedure was repeated 15 times to ensure excess inactivator was washed away. The resultant solution was diluted to 500 µL and kept on ice (or at 4$^{o}$C) until reactivation studies began. Reactivation was initiated by incubation of the enzyme in buffer at 37$^{o}$C and aliquots (20 µL) were removed at appropriate time intervals and assayed

using 3 mM pNPβGal as the substrate at pH 7.0 and 37$^{o}$C (conditions used for kinetic characterization of YesZ).

## 5.3 Stereochemistry determination by $^{1}$H NMR

$^{1}$H NMR spectrometry was performed on a Bruker 600AV at 600 MHz using 5 mm tubes. Experiments were conducted at ambient temperature using a water suppression protocol. EABase was buffer-exchanged with 10 centrifugations into 99.9% $D_2O$ using an Amicon centrifugal ultrafilter with a molecular weight cutoff of 10 000 K (Amicon), and DNP-A-tri (3.4 mg) was lyophilized twice from 99.9% D2O. Sodium acetate buffer (20 mM, pH 6.0) was lyophilized three times in 99.9% $D_2O$. The NMR tube was filled with 0.3 mL of 20 mM sodium acetate buffer at pH 6.0 containing 16 mM DNP-β-A-trisaccharide. EABase (0.20 mg) was added after initial spectra of DNP-β-A-trisaccharide were obtained. After the addition of EABase, spectra were acquired approximately every 4 min, with a final
$^{1}$H NMR of the same reaction mixture being measured the following morning after overnight equilibration at 20 $^{o}$C.

## 5.4 Synthesis

### 5.4.1 Materials

#### *5.4.1.1 Synthetic carbohydrates*

A-trisaccharide and A-pentassacharide were synthesized by M. Riandriantsoa from Dr. Eric Samain's group and sent to us.

#### *5.4.1.2 Commercially available substrates, buffers and oligonucleotides*

pNPαFuc and pNPβGal were purchased from Sigma-Aldrich. All buffer salts and reagents were purchased from commercial chemical suppliers (Sigma-Aldrich, Fluka and Cambridge Isotope Laboratories) and were used without further purification. All oligonucleotides were synthesized by IDTOligos Inc.

### 5.4.2 General methods

#### *5.4.2.1 General technical methods*

Reactions were monitored by thin layer chromatography and visualized using UV light (for choromogenic materials) and by exposure to 10% ammonium molybdate in 2 M $H_2SO_4$ followed by charring. Flash column chromatography was performed on 230-400 mesh gel. $^{1}H$ and $^{13}C$ NMR spectra were recorded on 300, 400, or 600 MHz spectrometers while $^{19}F$ NMR spectra were obtained at 282 MHz on a 300 MHz spectrometer. High resolution mass spectra were

measured in the mass spectrometry laboratory in the Department of Chemistry at UBC. All anhydrous solvents were dried and distilled prior to use and acetyl chloride (used for deacetylation) was also distilled prior to use.

#### *5.4.2.2 Acetylation*

Unprotected sugars were dissolved in dry pyridine in an ice bath (cooled to $0^{o}$C). Acetic anhydride was slowly added while stirring under a positive nitrogen atmosphere (ratio of volumes of acetic acid : pyridine is 2:3). The reaction mixture was left to slowly warm up to room temperature and left stirring until TLC showed that the reactions were complete. The reaction mixture was then evaporated and re-dissolved in either ethyl acetate of $CHCl_3$ and washed with 1 M HCl, $H_2O$, saturated $NaHCO_3$ and brine. The organic layer was then dried over anhydrous $MgSO_4$, filtered and then evaporated *in vacuo* to dryness before flash chromatography.

#### *5.4.2.3 Ammonia-promoted deacetylation*

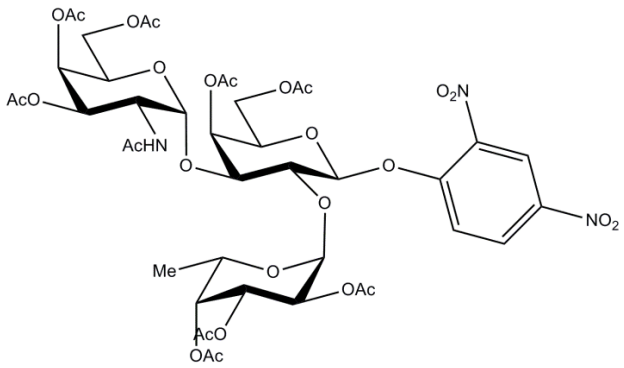
The protected sugar was dissolved in dry methanol (approximately 10 mg sugar to 5 mL methanol), cooled to $0^{o}$C in an ice water bath, and anhydrous ammonia gas was bubbled through until the solution was saturated with ammonia (~5 minutes). The reaction mixture was then left to stir at room temperature until the reaction was complete (by TLC). The reaction mixture was then concentrated *in vacuo* and the product was purified by flash chromatography.

*5.4.2.4 Acetyl-chloride promoted deacetylation*

Dry, distilled acetyl chloride (5% v/v of MeoH) was added to a solution of the protected sugar in dry methanol (approximately 10 mg sugar to 5 mL methanol) and the reaction was stirred at 4$^{o}$C until the reaction was complete. The reaction was then concentrated *in vacuo* and purified by flash chromatography.

## 5.4.3 Synthesis of DNP-A-trisaccharide

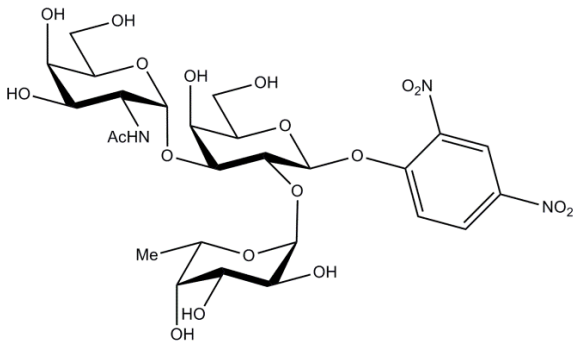
*Synthesis of fully protected DNP-β-A-trisaccharide (2,4-dinitrophenyl 2,3,4,5-tetra-O-acetyl-2-acetamido-2-deoxy-α-D-galactopyranosyl-(1→3)-[α-L-fucopyranosyl-(1→2)]- β-D-galactopyranose)*




A-trisaccharide was converted to its 2,4-dinitrophenyl glycoside using approaches that were previously described for the synthesis of other glycosides containing the DNP leaving group. The A-trisaccharide (100 mg, 0.2 mmol) was acetylated by stirring it in 25 mL of a 20% v/v acetic anhydride/pyridine mixture at 0 °C for 1 hour and subsequently stirring it overnight at 20 °C.

The solvents were evaporated and the resulting residue was dissolved in ethyl acetate (30 mL) and washed with 1 M HCl (1×30 mL), saturated $NaHCO_3$ (2×30 mL), brine (30 mL), dried over $MgSO_4$ and concentrated. The anomeric acetate was removed by treating the fully acetylated A-trisaccharide (145 mg, 0.16 mmol) with 20 mg (1.2 equivalents) of hydrazine acetate in 15 mL dry DMF at 55 °C for 3 hours (until the starting material was no longer visible by TLC). The reaction mixture was cooled to room temperature, concentrated, dissolved in ethyl acetate (50 mL) and washed with saturated $NaHCO_3$ (2×30 mL), brine (30 mL), dried over $MgSO_4$ and concentrated. In the next step, 30 mg (1.2 equivalents) of dinitrofluorobenzene (DNFB) and 20 mg (1.5 equivalents) of 1,4-diazabicyclo[2.2.2]octane (DABCO) in dry DMF (10 mL) were added to 110 mg (0.13 mmol) of the protected trisaccharide hemiacetal and the reaction mixture stirred overnight at 20 °C. A mixture of α and β products were obtained in an approximate ratio of 25% : 75% (α:β). The solvent was evaporated under reduced pressure and the β-product was purified by flash column chromatography (petroleum ether/ethyl acetate) to give the protected DNP-β-A-trisaccharide product (84 mg, 0.08 mmol, 45% from the starting material A-trisaccharide). $^1$H NMR: (400 MHz, $CDCl_3$): δ=8.90 (d, *J*=2.3 Hz, 1 H, Ar-H), 8.49 (dd, 1 H, Ar-H), 7.58 (d, *J*=9.0 Hz, 1 H, Ar-H); 5.19 (d, $J_{1,2}$=7.9 Hz, 1 H, H1'), MS: Calcd for $C_{20}H_{35}NO_{15}$+$Na^+$: 552.4769. Found: 552.4780.

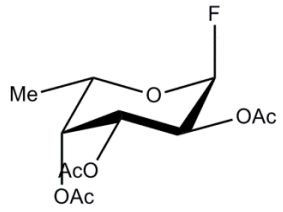
*Synthesis of DNP-β-A-trisaccharide (2,4-dinitrophenyl -2-acetamido-2-deoxy-α-D-galactopyranosyl-(1→3)-[α-L-fucopyranosyl-(1→2)]- β-D-galactopyranose)*




Deprotection was achieved by suspending 50 mg of the fully acetylated DNP-β-A-trisaccharide (49 µmol) in dry methanol (3 mL) which was cooled to 0 °C, under nitrogen. Distilled acetyl chloride (0.6 mL) was added drop-wise to the stirring solution and the mixture was stirred at 0 °C and allowed to stand at 4 °C for three days, with monitoring by TLC. The solvent was evaporated under reduced pressure and the residue was purified over a reverse phase C18 Sep-Pak column (gradient of 5-100% methanol : water) and subsequently lyophilized to yield the deprotected DNPβ-A-tri a white powder (21 mg, 30 µmol, 67% yield). $^{1}$H NMR: (600 MHz, $D_2O$): δ=8.79 (d, *J*=2.3 Hz, 1 H, Ar-H), 8.43 (dd, 1 H, Ar-H), 7.62 (d, *J*=9.2 Hz, 1 H, Ar-H); 5.30 (d, $J_{1,2}$=7.73 Hz, 1 H, H1'), 5.01 (d, $J_{1,2}$=3.67 Hz, 1 H, H1''), 5.02 (d, $J_{1,2}$=3.72 Hz, 1 H, H1''') ESI-MS: calcd for $[C_{26}H_{37}N_3O_{19}+Na]^+$ 718.2012; found: 718.2018.

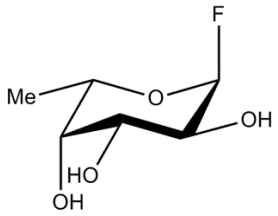
### 5.4.4 Synthesis of α-fucosyl fluoride

*Synthesis of 2,3,4-tri-O-acetyl-α-L-fucopyranosyl fluoride*




70% hydrogen fluoride (3 mL) in pyridine was added to a stirred solution of 1,2,3,4-tetra-O-acetyl-α-L-fucopyranose (1.0 g, 3.5 mmol) in dichloromethane (1 mL) at room temperature and stirred for 4 h. The reaction mixture was subsequently diluted with ethyl acetate (30 mL) and saturated $NaHCO_3$ (30 mL) and then the organic phase was washed multiple times with 30 mL each of saturated $NaHCO_3$, water and brine and dried over anhydrous $MgSO_4$ then filtered. The organic layer was concentrated *in vacuo* and purified by flash-column chromatography (petroleum ether : ethyl acetate, gradient from 4:1 to 1:1) to give the protected α-L-fucosyl fluoride product (0.52 g, 1.8 mmol , 59%. $^1$H NMR data (300 MHz, $CDCl_3$): δ 1.23 (d, 3H, H6a, H6b, H6c), δ 4.30 (q, 1H $J_{5,6}$=6.6 Hz, H5), δ 5.17 (ddd, 1H, H2, $J_{2,3}$=11.2, $J_{1,2}$ = *2.6*), δ 5.35 (m, 2H, H3, H4), δ 5.76 (dd, 1H, H1, $J_{1,2}$=2.6) $^{19}$F NMR ($CDCl_3$ ,282 MHz) δ 63.18 (dd, $J_{1,F}$=53.6, $J_{2,F}$= 22.9, 1F, F1) MS: Calcd for $C_{12}H_{17}FO_7$+$Na^+$: 315.0869. Found: 315.0910.

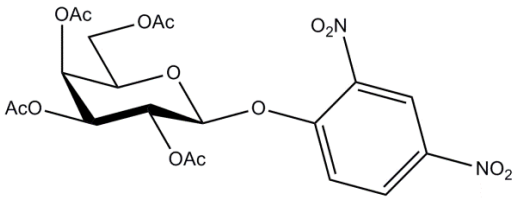
*Synthesis of α-L-fucopyranosyl fluoride*




2,3,4-Tri-O-acetyl-α-L-fucopyranosyl fluoride (0.5 g, 1.8 mmol) was dissolved in 30 mL dry, distilled methanol and deprotected using the ammonia promoted deacetylation method described earlier. The solvent was evaporated under reduced pressure and the residue was purified by flash chromatography (ethyl acetate/methanol/water: 17:2:1 to 7:2:1) and subsequently lyophilized to afford 0.22 g of the final product (1.3 mmol, 55%) . $^{1}$H NMR data (300 MHz, $CD_3OD$): δ 1.23 (d, 3H, H6a, H6b, H6c), δ 3.60 (m, 3H, H2, H3, H4), δ 4.17 (q, 1H, H5, $J_{5,6}$=6.2), δ 5.45 (dd, 1H, H1, $J_{1,2}$=2.58), $^{19}$F NMR ($CDCl_3$ ,282 MHz) δ 64.23 (dd, $J_{1,F}$=54.6, $J_{2,F}$= 21.9, 1F, F1)  MS: Calcd for $C_{12}H_{17}FO_7$+$Na^+$: 189.0612. Found: 189.0711.

### 5.4.5 Synthesis of 2,4-dinitrophenyl galactopyranoside

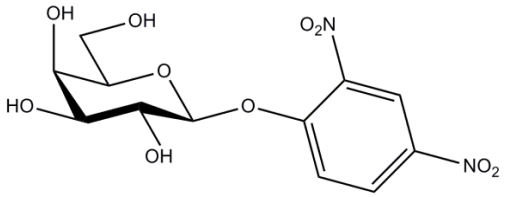
*Synthesis of 2,4-Dinitrophenyl 2,3,4,6-tetra-O-acetyl-β-D-galactopyranoside*




The anomeric acetate of 1,2,3,4,6-penta-O-acetyl-D-galactopyranose (1.0 g, 2.6 mmol) was removed by treatment with 1.2 equivalents (0.28 g) of hydrazine acetate in 50 mL dry DMF at 55 °C for 3 hours (until the starting material was no longer visible by TLC). The reaction mixture was cooled to room temperature, concentrated, dissolved in ethyl acetate (50 mL) and washed with saturated $NaHCO_3$ (2×30 mL), brine (30 mL), dried over $MgSO_4$ and concentrated. In the next step, In the next step, 0.50 g (1.2 equivalents) of dinitrofluorobenzene (DNFB) and 0.38 g (1.5 equivalents) of 1,4-diazabicyclo[2.2.2]octane (DABCO) in 25 mL of dry DMF were added to 0.81 g (2.3 mmol) of the protected hemiacetal and the reaction mixture stirred overnight at 20 °C. The solvent was evaporated under reduced pressure and the product purified by flash column chromatography (petroleum ether/ethyl acetate) to give the protected DNP-β-galactopyranoside product (0.87 g, 1.7 mmol, 66% yield from per-acetylated galactoside). $^1$H NMR: (300 MHz, $CDCl_3$): δ=8.70 (d, $J_{3,5}$=3.3 Hz, 1 H, Ar-H), δ 8.43 (dd, 1 H, Ar-H), δ 7.56 (d, $J$=10.0 Hz, 1 H, Ar-H); δ 5.59 (d, $J_{1,2}$=7.9 , $J_{2,3}$= 10.9 Hz, 1 H, H2'), δ 5.45 (d, 1H, $J_{3,4}$= 3.5 Hz, H4), δ

5.20 (d, 1H, $J_{1,2}$= 8.0 Hz, H1), δ 5.10 (dd, 1H, $J_{2,3}$=10.9 Hz, $J_{3,4}$= 3.6 Hz, H3), MS: Calcd for $C_{20}H_{22}O_{14}N_2$+$Na^+$: 537.1080. Found: 537.1121.

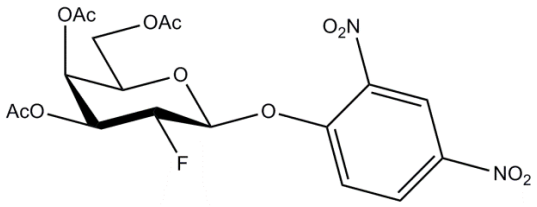
*Synthesis of 2,4-Dinitrophenyl β-D-galactopyranoside*




The tetra-O-acetate produced above (0.50 g, 1.0 mmol) was dissolved in 10 mL dry methanol and deacetylated according to the general acetyl chloride promoted deacetylation method. The product was purified by flash column chromatography (ethyl acetate/methanol/water 17:2:1 to 7:2:1) and lyophilized to give the deprotected product (0.25 g, 0.73 mmol, 68%). $^1$H NMR: (400 MHz, $CD_3OD$): δ=8.70 (d, $J_{3,5}$=3.3 Hz, 1 H, Ar-H), δ 8.48 (dd, 1 H, Ar-H), δ 7.66 (d, $J$=10.0 Hz, 1 H, Ar-H); δ 5.51 (d, $J_{1,2}$=8.1 Hz, 1 H, H1) δ 4.4-3.4 (m, 6H, H2, H3, H4, H6) MS: Calcd for $C_{12}H_{14}O_{10}N_2$+$Na^+$: 369.0141. Found: 369.0186.

### 5.4.6 Synthesis of 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-galactopyranoside

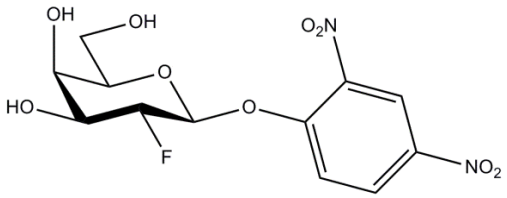
*Synthesis of 2,4-dinitrophenyl 3,4,6-tri-O-acetyl-2-deoxy-2-fluoro-β-D-galactopyranoside*




Per-O-acetylated D-galactal (0.5 g, 1.83 mmol) and 0.77 g (1.2 equivalents) Selectfluor$^{TM}$ were dissolved in 70 mL of a 2:1 acetonitrile: acetic acid mixture and the solution was stirred overnight at 70$^{o}$C. An additional 5mL acetic acid was added the following day and the mixture was stirred at 70$^{o}$C until the starting material had been consumed. The organic solvent was evaporated *in vacuo* and redissolved in 50 mL ethyl acetate, washed with 50 mL each of water, saturated $NaHCO_3$, and brine, dried over anhydrous $MgSO_4$ and filtered and evaporated. In the next step, 0.22 g (0.6 mmol) of the material was dissolved in dry 25 mL $CH_2Cl_2$ and 6 mL HBr (33% in AcOH) was added and the mixture was stirred until all the starting material was consumed. The material was diluted to 30 mL $CH_2Cl_2$ and worked up with 20 mL each of ice cold water, ice cold saturated $NaHCO_3$ and ice cold brine, dried and evaporated. The bromide (0.18 g, 0.5 mmol) was dissolved in an acetone/water mixture (5 mL : 4 drops) and 0.15 g silver carbonate (1:1 molar ratio) was added while keeping the flask contents at 0$^{o}$C and protected from light. The reaction mixture was stirred for an hour at 0$^{o}$C and then overnight at room tempterature. The suspension was filtered and the filtrate was evaporated, and purified by flash column chromatography (ethyl acetate/petroleum ether: gradient of 4:1 to 1:1). This

material (0.15 g, 0.4 mmol) was subsequently converted to the galactopyranoside by reaction with 1.2 equivalents (90 mg, 0.5 mmol) of dinitrofluorobenzene (DNFB) and 1.5 equivalents (70 mg, 0.6 mmol) of 1,4-diazabicyclo[2.2.2]octane (DABCO) in 15 mL dry DMF. Flash column chromatography (ethyl acetate: petroleum ether; 4:1 to 1:1) of the product of the reaction allowed for the separation of the β-anomer of 2,4-dinitrophenyl 3,4,6-tri-O-acetyl-2-deoxy-2-fluoro-β-D-galactopyranoside (76 mg, 0.16 mmol, 10% yield from starting galactal). $^{1}$H NMR: (300 MHz, $CDCl_3$): δ=8.80 (d, *J*=3.0 Hz, 1 H, Ar-H), δ 8.40 (dd, 1 H, Ar-H), δ 7.46 (d, *J*=10.0 Hz, 1 H, Ar-H); δ 5.50 (t, $J_{3,4}$=2.9, 1 H, H4), δ 5.35 (dd, 1H, $J_{1,2}$= 7.5 Hz, H4), δ 5.20 (dt, 1H, $J_{3,4}$= 3.0 Hz, $J_{2,3}$= 10.0 Hz H1), δ 5.00 (ddd, 1H, $J_{1,2}$=7.5 Hz, $J_{2,3}$= 8.8 Hz, H2), $^{19}$F NMR ($CDCl_3$ ,282 MHz) δ 208.23 MS: Calcd for $C_{18}H_{19}O_{12}N_2F+Na^+$: 497.0802. Found: 497.0861.

*Synthesis of 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-galactopyranoside*




50 mg (0.1 mmol) of the protected 2,4-dinitrophenyl 3,4,6-tri-O-acetyl-2-deoxy-2-fluoro-β-D-galactopyranoside was dissolved in 10 mL dry methanol and deacetylated using the acetyl chloride promoted deacetylation method. The product was purified over flash chromatography (ethyl acetate: methanol: water; 17:2:1 to 7:2:1) to yield 21 mg (65 µmol, 60%) of the

deprotected DNP2FGal product. $^{1}$H NMR: (400 MHz, $CD_3OD$): δ=8.70 (d, 1 H, Ar-H), δ 8.48 (dd, 1 H, Ar-H), δ 7.66 (d, 1 H, Ar-H); δ 5.51 (d, $J_{1,2}$=7.6 Hz, 1 H, H1) δ 4.1-3.8 (m, 6H, H2, H3, H4, H6) $^{19}$F NMR ($CDCl_3$ ,282 MHz) δ 218.53 MS: Calcd for $C_{12}H_{14}O_{10}N_2$+$Na^+$: 371.1214. Found: 371.1281.

# References


(1) Gruber, E. *Papier* **1976**, *30*, 133-138.

(2) Zubay, G. *Biochemistry*; 2nd ed.; McMillan Publishing Co.: New York, 1988; p. 131.

(3) Helenius, A.; Aebi, M. *Science* **2001**, *291*, 2364-9.

(4) Lorant, D. E.; McEver, R. P.; McIntyre, T. M.; Moore, K. L.; Prescott, S. M.; Zimmerman, G. A. *The Journal of Clinical Investigation* **1995**, *96*, 171-82.

(5) Hauri, H.; Appenzeller, C.; Kuhn, F.; Nufer, O. *FEBS Letters* **2000**, *476*, 32-7.

(6) Wells, L.; Vosseller, K.; Hart, G. W. *Science* **2001**, *291*, 2376-8.

(7) Lowe, J. *Biochemistry and Biosynthesis of ABH and Lewis Antigens*; Cartron, J.P.R., Rouger, P., Ed.; Plenum Press: Paris, France, 1995; pp. 75-109.

(8) Oriol, R. *ABO, Hh, Lewis and Secretion: Molecular Basis of Human Blood Group Antigens*; Catron, J.P, and Rouger, P., Ed.; Plenum Press: Paris, France, 1995; pp. 37-67.

(9) Williams, S.J., Davies, G. J. *Trends in Biotechnology* **2001**, *19*, 356-361.

(10) Landsteiner, K. *The Central European Journal of Medicine* **2001**, *113*, 768-9.

(11) Watkins, W. M. *Advances in human genetics* **1980**, *10*, 1-136, 379-85.

(12) Clausen, H.; Hakomori, S. *Vox sanguinis* **1989**, *56*, 1-20.

(13) Yamamoto, F.; Clausen, H.; White, T.; Marken, J.; Hakomori, S. *Nature* **1990**, *345*, 229-33.

(14) Ryzhov, I. M.; Korchagina, E. Y.; Popova, I. S.; Bovin, N. V. *Carbohydrate research* **2012**, *351*, 17-25.

(15) Sazama, K. *Current Hematology Reports* **2003**, *2*, 518-21.

(16) Lalueza-Fox, C.; Gigli, E.; de la Rasilla, M.; Fortea, J.; Rosas, A.; Bertranpetit, J.; Krause, J. *BMC Evolutionary Biology* **2008**, *8*, 342.

(17) Barnes, A. *Bibliotheca Haematologica* **1980**, 132-42.

(18) Kelly, R. J.; Ernst, L. K.; Larsen, R. D.; Bryant, J. G.; Robinson, J. S.; Lowe, J. B. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 5843-7.

(19) Roubinet, F.; Despiau, S.; Calafell, F.; Jin, F.; Bertranpetit, J.; Saitou, N.; Blancher, A.; Bertanpetit, J. *Journal of Transfusion* **2004**, *44*, 707-15.

(20) Wolfenden, R., Lu, X. D., and Young, G. *Journal of the American Chemical Society* **1998**, *120*, 6814-6815.

(21) Wolfenden, R., Snider, M., Ridgway, C., Miller, B. *Journal of the American Chemical Society* **1999**, *121*, 7419-7420.

(22) Sinnott, M. L. *Chemical Reviews* **1990**, *90*, 1171-1202.

(23) Davies, G. J.; Wilson, K. S.; Henrissat, B. *The Biochemical Journal* **1997**, *321 ( Pt 2*, 557-9.

(24) Henrissat, B.; Davies, G. *Current Opinion in Structural Biology* **1997**, *7*, 637-44.

(25) Henrissat, B. *Biochemical Society Transactions* **1998**, *26*, 153-6.

(26) Henrissat, B.; Bairoch, A. *The Biochemical Journal* **1993**, *293 ( Pt 3*, 781-8.

(27) Henrissat, B. *The Biochemical Journal* **1991**, *280*, 309-16.

(28) Davies, G. J.; Gloster, T. M.; Henrissat, B. *Current Opinion in Structural Biology* **2005**, *15*, 637-45.

(29) Zechel, D. L.; Reid, S. P.; Stoll, D.; Nashiru, O.; Warren, R. A. J.; Withers, S. G. *Biochemistry* **2003**, *42*, 7195-204.

(30) Koshland, D. E. *Biological Reviews* **1953**, 416-436.

(31) Bourne, Y.; Henrissat, B. *Current Opinion in Structural Biology* **2001**, *11*, 593-600.

(32) Davies, G.; Henrissat, B. *Structure* **1995**, *3*, 853-9.

(33) Kempton, J. B.; Withers, S. G. *Biochemistry* **1992**, *31*, 9961-9.

(34) Li, Y. K.; Chir, J.; Chen, F. Y. *The Biochemical Journal* **2001**, *355*, 835-40.

(35) Vocadlo, D. J.; Wicki, J.; Rupitz, K.; Withers, S. G. *Biochemistry* **2002**, *41*, 9736-46.

(36) Howard, S.; Braun, C.; McCarter, J.; Moremen, K. W.; Liao, Y. F.; Withers, S. G. *Biochemical and Biophysical Research Communications* **1997**, *238*, 896-8.

(37) McCarter, J. D.; Withers, S. G. *The Journal of Biological Chemistry* **1996**, *271*, 6889-94.

(38) Matsui, H.; Blanchard, J. S.; Brewer, C. F.; Hehre, E. J. *The Journal of Biological Chemistry* **1989**, *264*, 8714-6.

(39) van Doorslaer, E., van Opstal, O., Kersters-Hilderson, H. DeBruyne, C. . *Bioorganic Chemistry* **1984**, *12*, 158-169.

(40) Tanaka, Y.; Tao, W.; Blanchard, J. S.; Hehre, E. J. *The Journal of Biological Chemistry* **1994**, *269*, 32306-12.

(41) Wang, Q., Graham, R.W., Trimbur, D., Warren, R.A.J, Withers, S. . *Journal of the American Chemical Society* **1994**, 365-377.

(42) McCarter, J. D.; Withers, S. G. *Current Opinion in Structural Biology* **1994**, *4*, 885-92.

(43) Caines, M. E. C.; Hancock, S. M.; Tarling, C. A.; Wrodnigg, T. M.; Stick, R. V.; Stütz, A. E.; Vasella, A.; Withers, S. G.; Strynadka, N. C. J. *Angewandte Chemie (International ed. in English)* **2007**, *46*, 4474-6.

(44) Vasella, A.; Davies, G. J.; Böhm, M. *Current Opinion in Chemical Biology* **2002**, *6*, 619-29.

(45) Uehara, K.; Mannen, S. *Journal of Biochemistry* **1979**, *85*, 105-13.

(46) McCarter, J. D.; Withers, S. G. *Current Opinion in Structural Biology* **1994**, *4*, 885-892.

(47) Withers, S. G.; Rupitz, K.; Trimbur, D.; Warren, R. A. *Biochemistry* **1992**, *31*, 9979-85.

(48) Zechel, D. L.; Withers, S. G. *Current Opinion in Chemical Biology* **2001**, *5*, 643-9.

(49) Viladot, J. L.; de Ramon, E.; Durany, O.; Planas, A. *Biochemistry* **1998**, *37*, 11332-42.

(50) Monegal, A.; Planas, A. *Journal of the American Chemical Society* **2006**, *128*, 16030-1.

(51) Vallmitjana, M.; Ferrer-Navarro, M.; Planell, R.; Abel, M.; Ausín, C.; Querol, E.; Planas, A.; Pérez-Pons, J. A. *Biochemistry* **2001**, *40*, 5975-82.

(52) MacLeod, A. M.; Tull, D.; Rupitz, K.; Warren, R. A.; Withers, S. G. *Biochemistry* **1996**, *35*, 13165-72.

(53) MacLeod, A.M., Lindhorst, T., Withers, S.G. and Warren, R. A. *Biochemistry* **1994**, *33*, 6371.

(54) Knapp, S.; Yang, C.; Pabbaraja, S.; Rempel, B.; Reid, S.; Withers, S. G. *The Journal of Organic Chemistry* **2005**, *70*, 7715-20.

(55) Mark, B. L.; Vocadlo, D. J.; Knapp, S.; Triggs-Raine, B. L.; Withers, S. G.; James, M. N. *The Journal of Biological Chemistry* **2001**, *276*, 10330-7.

(56) Mark, B. L.; Vocadlo, D. J.; Zhao, D.; Knapp, S.; Withers, S. G.; James, M. N. *The Journal of Biological Chemistry* **2001**, *276*, 42131-7.

(57) Vocadlo, D. J.; Withers, S. G. *Biochemistry* **2005**, *44*, 12809-18.

(58) Vocadlo, D. J.; Davies, G. J. *Current Opinion in Chemical Biology* **2008**, *12*, 539-55.

(59) Watts, A. G.; Damager, I.; Amaya, M. L.; Buschiazzo, A.; Alzari, P.; Frasch, A. C.; Withers, S. G. *Journal of the American Chemical Society* **2003**, *125*, 7532-7533.

(60) Watts, A. G.; Oppezzo, P.; Withers, S. G.; Alzari, P. M.; Buschiazzo, A. *The Journal of Biological Chemistry* **2006**, *281*, 4149-55.

(61) Varrot, A.; Yip, V. L. Y.; Li, Y.; Rajan, S. S.; Yang, X.; Anderson, W. F.; Thompson, J.; Withers, S. G.; Davies, G. J. *Journal of Molecular Biology* **2005**, *346*, 423-35.

(62) Rajan, S. S.; Yang, X.; Collart, F.; Yip, V. L. Y.; Withers, S. G.; Varrot, A.; Thompson, J.; Davies, G. J.; Anderson, W. F. *Structure* **2004**, *12*, 1619-29.

(63) Yip, V. L.; Varrot, A.; Davies, G. J.; Rajan, S. S.; Yang, X.; Thompson, J.; Anderson, W. F.; Withers, S. G. *Journal of the American Society for Mass Spectrometry* **2004**, *126*, 8354-8355.

(64) Sinnott, M. L. *FEBS Letters* **1978**, *94*, 1-9.

(65) Umezurike, G. M. *Biochemical Journal* **1988**, *254*, 73-76.

(66) Tull, D.; Withers, S. G. *Biochemistry* **1994**, *33*, 6363-70.

(67) Zechel, D. L.; Withers, S. G. *Accounts of Chemical Research* **2000**, *33*, 11-18.

(68) Legler, G. *Advanced Carbohydrate Chemistry and Biochemistry* **1990**, *48*, 319-385.

(69) Heightman, TD Vasella, A. *Angewandte Chemie (International ed. in English)* **1999**, *38*, 750-770.

(70) Withers, S. G.; Rupitz, K.; Street, I. P. *Journal of Biological Chemistry* **1988**, *263*, 7929-7932.

(71) Withers, S.G., Street, I. . *Journal of the American Chemical Society* **1988**, *110*, 8551-8553.

(72) Namchuk, M. N.; Withers, S. G. *Biochemistry* **1995**, *34*, 16194-202.

(73) McCarter, J. D.; Adam, M. J.; Withers, S. G. *The Biochemical Journal* **1992**, *286 ( Pt 3*, 721-7.

(74) Vocadlo, D. J.; Davies, G. J.; Laine, R.; Withers, S. G. *Nature* **2001**, *412*, 835-8.

(75) Tull, D.; Miao, S.; Withers, S. G.; Aebersold, R. *Analytical Biochemistry* **1995**, *224*, 509-514.

(76) Miao, S.; McCarter, J. D.; Grace, M. E.; Grabowski, G. A.; Aebersold, R.; Withers, S. G. *The Journal of Biological Chemistry* **1994**, *269*, 10975-8.

(77) Miao, S.; Ziser, L.; Aebersold, R.; Withers, S. G. *Biochemistry* **1994**, *33*, 7027-32.

(78) Mosi, R.; He, S.; Uitdehaag, J.; Dijkstra, B. W.; Withers, S. G. *Biochemistry* **1997**, *36*, 9927-34.

(79) Wicki, J.; Rose, D. R.; Withers, S. G. *Methods in Enzymology* **2002**, *354*, 84-105.

(80) Withers, S. G.; Aebersold, R. *Protein Science* **1995**, *4*, 361-72.

(81) Ly, H. D.; Withers, S. G. *Annual Review of Biochemistry* **1999**, *68*, 487-522.

(82) MacLeod, A. M.; Lindhorst, T.; Withers, S. G.; Warren, R. A. *Biochemistry* **1994**, *33*, 6371-6.

(83) Burmeister, W. P.; Cottaz, S.; Rollin, P.; Vasella, A.; Henrissat, B. *The Journal of Biological Chemistry* **2000**, *275*, 39385-93.

(84) McIntosh, L. P.; Hand, G.; Johnson, P. E.; Joshi, M. D.; Körner, M.; Plesniak, L. A.; Ziser, L.; Wakarchuk, W. W.; Withers, S. G. *Biochemistry* **1996**, *35*, 9958-66.

(85) Lawson, S. L.; Wakarchuk, W. W.; Withers, S. G. *Biochemistry* **1997**, *36*, 2257-65.

(86) Lawson, S. L.; Wakarchuk, W. W.; Withers, S. G. *Biochemistry* **1996**, *35*, 10110-8.

(87) Lawson, S. L.; Warren, R. A.; Withers, S. G. *The Biochemical Journal* **1998**, *330 ( Pt 1*, 203-9.

(88) Notenboom, V.; Williams, S. J.; Hoos, R.; Withers, S. G.; Rose, D. R. *Biochemistry* **2000**, *39*, 11553-63.

(89) Varrot, A.; Hastrup, S.; Schülein, M.; Davies, G. J. *The Biochemical Journal* **1999**, *337 ( Pt 2*, 297-304.

(90) Davies, G. J.; Brzozowski, A. M.; Dauter, M.; Varrot, A.; Schülein, M. *The Biochemical Journal* **2000**, *348 Pt 1*, 201-7.

(91) Varrot, A.; Schülein, M.; Fruchard, S.; Driguez, H.; Davies, G. J. *Acta crystallographica. Section D, Biological crystallography* **2001**, *57*, 1739-42.

(92) Heightman, T D., Vasella, A. T. *Angewandte Chemie (International ed. in English)* **1999**, *38*, 750-770.

(93) Williams, S. J., Hoos, R., Withers, S. G. *Journal of the American Chemical Society* **2000**, *122*, 2223-2225.

(94) Wang, Q.; Trimbur, D.; Graham, R.; Warren, R. A.; Withers, S. G. *Biochemistry* **1995**, *34*, 14554-62.

(95) Pauling, L. *Chemical Engineering News* **1946**, *24*, 1375.

(96) Goldstein, J.; Siviglia, G.; Hurst, R.; Lenny, L.; Reich, L. *Science* **1982**, *215*, 168-170.

(97) Liu, Q. P.; Sulzenbacher, G.; Yuan, H.; Bennett, E. P.; Pietz, G.; Saunders, K.; Spence, J.; Nudelman, E.; Levery, S. B.; White, T.; Neveu, J. M.; Lane, W. S.; Bourne, Y.; Olsson, M. L.; Henrissat, B.; Clausen, H. *Nature Biotechnology* **2007**, *25*, 454-64.

(98) Anderson, K. M.; Ashida, H.; Maskos, K.; Dell, A.; Li, S.-C.; Li, Y.-T. *The Journal of Biological Chemistry* **2005**, *280*, 7720-8.

(99) Chapanian, R.; Constantinescu, I.; Brooks, D. E.; Scott, M. D.; Kizhakkedathu, J. N. *Biomaterials* **2012**, *33*, 3047-57.

(100) Chang, B. S.; Mahoney, R. R. *Journal of Dairy Research* **1989**, *56*, 117-127.

(101) Pastore, M.; Morisi, F. *Methods in Enzymology* **1976**, *44*, 822-830.

(102) Hidaka, M.; Fushinobu, S.; Ohtsu, N.; Motoshima, H.; Matsuzawa, H.; Shoun, H.; Wakagi, T. *Journal of Molecular Biology* **2002**, *322*, 79-91.

(103) Di Lauro, B.; Strazzulli, A.; Perugino, G.; La Cara, F.; Bedini, E.; Corsaro, M. M.; Rossi, M.; Moracci, M. *Biochimica et Biophysica Acta* **2008**, *1784*, 292-301.

(104) Hidaka, M.; Fushinobu, S.; Ohtsu, N.; Motoshima, H.; Matsuzawa, H.; Shoun, H.; Wakagi, T. *Journal of Molecular Biology* **2002**, *322*, 79-91.

(105) Withers, S. G. *Carbohydrate Polymers* **2001**, *44*, 325-337.

(106) Wong, A. W.; He, S.; Grubb, J. H.; Sly, W. S.; Withers, S. G. *Journal of Biological Chemistry* **1998**, *273*, 34057-34062.

(107) Blanchard, J. E.; Gal, L.; He, S.; Foisy, J.; Warren, R. A.; Withers, S. G. *Carbohydrate Research* **2001**, *333*, 7-17.

(108) Shipkowski, S.; Brenchley, J. E. *Applied and Environmental Microbiology* **2006**, *72*, 7730-7738.

(109) Hinz, S. W. a; van den Brock, L. a M.; Beldman, G.; Vincken, J.-P.; Voragen, A. G. J. *Applied Microbiology and Biotechnology* **2004**, *66*, 276-84.

(110) Permpoonpattana, P.; Hong, H. A.; Khaneja, R.; Cutting, S. M. *Beneficial microbes* **2012**, 1-9.

(111) Van Laere, K. M.; Abee, T.; Schols, H. a; Beldman, G.; Voragen, a G. *Applied and Environmental Microbiology* **2000**, *66*, 1379-84.

(112) Møller, P. L.; Jørgensen, F.; Hansen, O. L. E. C.; Madsen, S. M.; Stougaard, P. *Applied Environmental Biology* **2001**, *67*, 2276-2283.

(113) Mosi, R. M.; Withers, S. G. *Methods in Enzymology* **2002**, *354*, 64-84.

(114) Leatherbarrow, R. *GraFit Version 7*; 4.0 Ed.; Erithacus Software Ltd.: Horley, UK, 2009.

(115) Street, I. P.; Kempton, J. B.; Withers, S. G. *Biochemistry* **1992**, *31*, 9970-9978.

(116) Huppes, W.; Paulonis, J.; Dijk, H.; Van Rooijen, N.; Van Bekkum, D. W. *European Journal of Immunology* **1993**, *23*, 26-32.

(117) Murphy, S. *Transfusion 28*, 401-2.

(118) Gordon, R. D.; Iwatsuki, S.; Esquivel, C. O.; Todo, S.; Makowka, L.; Tzakis, A.; Marsh, J. W.; Starzl, T. E. *Transplantation Proceedings* **1987**, *19*, 4575-9.

(119) Hendriks, G. F.; Wenting, G. J.; Mochtar, B.; Bos, E.; Simoons, M. L.; Balk, A. H.; Laird-Meeter, K.; Essed, C. E.; Weimar, W. *Transplantation Proceedings* **1989**, *21*, 803.

(120) White, D. J.; Gore, S. M.; Barroso, E.; Calne, R. Y. *Transplantation Proceedings* **1987**, *19*, 4571-4.

(121) Morgan, W. T. J. *Experientia* **1947**, *3*, 257-67.

(122) Awad, M. M.; Bryant, A. E.; Stevens, D. L.; Rood, J. I. *Molecular Microbiology* **1995**, *15*, 191-202.

(123) Petit, L.; Gibert, M.; Popoff, M. R. *Trends in Microbiology* **1999**, *7*, 104-10.

(124) Alape-Girón, A.; Flores-Díaz, M.; Guillouard, I.; Naylor, C. E.; Titball, R. W.; Rucavado, A.; Lomonte, B.; Basak, A. K.; Gutiérrez, J. M.; Cole, S. T.; Thelestam, M. *European Journal of Biochemistry* **2000**, *267*, 5191-7.

(125) Minutti, C. Z.; Immergluck, L. C.; Schmidt, M. L. *Clinical Infectious Diseases* **1999**, *28*, 159-60.

(126) Barth, H.; Aktories, K.; Popoff, M. R.; Stiles, B. G. *Microbiology and Molecular Biology Reviews* **2004**, *68*, 373-402.

(127) Newstead, S. L.; Potter, J. A.; Wilson, J. C.; Xu, G.; Chien, C.-H.; Watts, A. G.; Withers, S. G.; Taylor, G. L. *The Journal of Biological Chemistry* **2008**, *283*, 9080-8.

(128) Yamakawa, Y.; Ohsaka, A. *Journal of Biochemistry* **1977**, *81*, 115-26.

(129) Manji, R. A.; Manji, J. S.; Koshal, A.; Korbutt, G. S.; Rajotte, R. V. *American Journal of Transplantation* **2003**, *3*, 286-93.

(130) Filitti-Wurmser, S. *Biochimie* **1976**, *58*, 1345-53.

(131) Rigden, D. J. *FEBS Letters* **2005**, *579*, 5466-72.

(132) Withers, S. G.; Dombroski, D.; Berven, L. A.; Kilburn, D. G.; Miller, R. C.; Warren, R. A.; Gilkes, N. R. *Biochemical and Biophysical Research Communications* **1986**, *139*, 487-94.

(133) Gastinel, L. N.; Bignon, C.; Misra, A. K.; Hindsgaul, O.; Shaper, J. H.; Joziasse, D. H. *The EMBO journal* **2001**, *20*, 638-49.

(134) Shah, P. S.; Bizik, F.; Dukor, R. K.; Qasba, P. K. *Biochimica et Biophysica Acta* **2000**, *1480*, 222-34.

(135) Drouillard, S.; Driguez, H.; Samain, E. *Angewandte Chemie (International ed. in English)* **2006**, *45*, 1778-80.

(136) Andrés, M. I.; Repetto, G.; Sanz, P.; Repetto, M. *Toxicology* **1996**, *110*, 123-32.

(137) Nunes, C.; Sousa, C.; Ferreira, H.; Lucio, M.; Lima, J. L. F. C.; Tavares, J.; Cordeiro-da-Silva, A.; Reis, S. *Journal of environmental biology / Academy of Environmental Biology, India* **2008**, *29*, 733-8.

(138) Stam, M. R.; Danchin, E. G. J.; Rancurel, C.; Coutinho, P. M.; Henrissat, B. *Protein engineering, design & selection : PEDS* **2006**, *19*, 555-62.

(139) Koeners, H.J., Dekok, A.J., Romers, C., Vanboom, J. H. *Journal of the Royal Netherlands Chemical Society* **1980**, *99*, 355-362.

(140) Strickland, E. H. *Critical Reviews in Biochemistry* **1974**, *2*, 113-75.

(141) Yada, R. Y.; Jackman, R. L.; Nakai, S. *International Journal of Peptide and Protein Research* **1988**, *31*, 98-108.

(142) Nisizawa, K.; Kanda, T.; Shikata, S.; Wakabayashi, K. *Journal of Biochemistry* **1978**, *83*, 1625-30.

(143) Sinnott, M. L.; Souchard, I. J. *The Biochemical Journal* **1973**, *133*, 89-98.

(144) Higgins, M. A.; Whitworth, G. E.; El Warry, N.; Randriantsoa, M.; Samain, E.; Burke, R. D.; Vocadlo, D. J.; Boraston, A. B. *The Journal of biological chemistry* **2009**, *284*, 26161-73.

(145) Damude, H. G.; Withers, S. G.; Kilburn, D. G.; Miller, R. C.; Warren, R. A. *Biochemistry* **1995**, *34*, 2220-4.

(146) Rydberg, E. H.; Li, C.; Maurus, R.; Overall, C. M.; Brayer, G. D.; Withers, S. G. *Biochemistry* **2002**, *41*, 4492-502.

(147) Joshi, M. D.; Sidhu, G.; Pot, I.; Brayer, G. D.; Withers, S. G.; McIntosh, L. P. *Journal of Molecular Biology* **2000**, *299*, 255-79.

(148) Kumar, V. *Carbohydrate Research* **2010**, *345*, 893-8.

(149) Howard, S.; Withers, S. G. *Biochemistry* **1998**, *37*, 3858-64.

(150) Nishimura, T.; Bignon, C.; Allouch, J.; Czjzek, M.; Darbon, H.; Watanabe, T.; Henrissat, B. *FEBS letters* **2001**, *499*, 187-90.

(151) Kostylev, M.; Wilson, D. B. *Applied and Environmental Microbiology* **2011**, *77*, 6274-6.

(152) Fujita, K.; Sato, R.; Toma, K.; Kitahara, K.; Suganuma, T.; Yamamoto, K.; Takegawa, K. *Journal of Biochemistry* **2007**, *142*, 301-6.

(153) Miyake, H.; Otsuka, C.; Nishimura, S.; Nitta, Y. *Journal of Biochemistry* **2002**, *131*, 587-91.

(154) Ohnuma, T.; Fukuda, T.; Dozen, S.; Honda, Y.; Kitaoka, M.; Fukamizo, T. *The Biochemical journal* **2012**.

(155) Wang, L.-X. *Chemistry and Biology* **2009**, *16*, 1026-7.

(156) Hancock, S. M.; Vaughan, M. D.; Withers, S. G. *Current Opinion in Chemical Biology* **2006**, *10*, 509-19.

(157) Shaikh, F. A.; Withers, S. G. *Biochemistry and Cell Biology* **2008**, *86*, 169-77.

(158) Goddard-Borger, E. D.; Fiege, B.; Kwan, E. M.; Withers, S. G. *Chembiochem* **2011**, *12*, 1703-11.

(159) Jahn, M.; Stoll, D.; Warren, R. A. J.; Szabó, L.; Singh, P.; Gilbert, H. J.; Ducros, V. M. A.; Davies, G. J.; Withers, S. G. *Chemical communications (Cambridge, England)* **2003**, 1327-9.

(160) Wang, D.; Kyluik, D. L.; Murad, K. L.; Toyofuku, W. M.; Scott, M. D. *Science China. Life sciences* **2011**, *54*, 589-98.

(161) Wang, D.; Toyofuku, W. M.; Scott, M. D. *Biomaterials* **2012**, *33*, 3002-12.

(162) Rossi, N. A. A.; Constantinescu, I.; Kainthan, R. K.; Brooks, D. E.; Scott, M. D.; Kizhakkedathu, J. N. *Biomaterials* **2010**, *31*, 4167-78.

(163) Rossi, N. A. A.; Mustafa, I.; Jackson, J. K.; Burt, H. M.; Horte, S. A.; Scott, M. D.; Kizhakkedathu, J. N. *Biomaterials* **2009**, *30*, 638-48.

(164) Rossi, N. A. A.; Constantinescu, I.; Brooks, D. E.; Scott, M. D.; Kizhakkedathu, J. N. *Journal of the American Chemical Society* **2010**, *132*, 3423-30.

(165) Nelson, P. *Biological Physics: Energy, Information, Life*; Susan Brennan: Philadelphia, PA, 2003; pp. 250-256.

(166) Listinsky, J. J.; Siegal, G. P.; Listinsky, C. M. *American Journal of Clinical Pathology* **1998**, *110*, 425-40.

(167) Michalski, J. C.; Klein, A. *Biochimica et Biophysica Acta* **1999**, *1455*, 69-84.

(168) Noda, K.; Miyoshi, E.; Kitada, T.; Nakahara, S.; Gao, C.-X.; Honke, K.; Shiratori, Y.; Moriwaki, H.; Sasaki, Y.; Kasahara, A.; Hori, M.; Hayashi, N.; Taniguchi, N. *Tumour Biology 23*, 202-11.

(169) Rapoport, E.; Pendu, J. L. *Glycobiology* **1999**, *9*, 1337-45.

(170) Russell, L.; Waring, P.; Beaver, J. P. *Biochemical and Biophysical Research Communications* **1998**, *250*, 449-53.

(171) Wiese, T. J.; Dunlap, J. A.; Yorek, M. A. *Biochimica et Biophysica Acta* **1997**, *1335*, 61-72.

(172) Mori, E.; Hedrick, J. L.; Wardrip, N. J.; Mori, T.; Takasaki, S. *Glycoconjugate Journal* **1998**, *15*, 447-56.

(173) de La Torre, F.; Sampedro, J.; Zarra, I.; Revilla, G. *Plant Physiology* **2002**, *128*, 247-55.

(174) Vanhooren, P.T., Vandamme, E. J. *Journal of Chemical Technology & Biotechnology* **1999**, *74*, 479-497.

(175) Ashida, H.; Miyake, A.; Kiyohara, M.; Wada, J.; Yoshida, E.; Kumagai, H.; Katayama, T.; Yamamoto, K. *Glycobiology* **2009**, *19*, 1010-7.

(176) Liu, S.-W.; Chen, C.-S.; Chang, S.-S.; Mong, K.-K. T.; Lin, C.-H.; Chang, C.-W.; Tang, C. Y.; Li, Y.-K. *Biochemistry* **2009**, *48*, 110-20.

(177) Eneyskaya, E. V.; Kulminskaya, A. A.; Kalkkinen, N.; Nifantiev, N. E.; Arbatskii, N. P.; Saenko, A. I.; Chepurnaya, O. V.; Arutyunyan, A. V.; Shabalin, K. A.; Neustroev, K. N. *Glycoconjugate Journal* **2001**, *18*, 827-34.

(178) Tarling, C. A.; He, S.; Sulzenbacher, G.; Bignon, C.; Bourne, Y.; Henrissat, B.; Withers, S. G. *The Journal of Biological Chemistry* **2003**, *278*, 47394-9.

(179) Sulzenbacher, G.; Bignon, C.; Nishimura, T.; Tarling, C. A.; Withers, S. G.; Henrissat, B.; Bourne, Y. *The Journal of Biological Chemistry* **2004**, *279*, 13119-28.

(180) Cobucci-Ponzano, B.; Conte, F.; Rossi, M.; Moracci, M. *Extremophiles* **2008**, *12*, 61-8.

(181) Cobucci-Ponzano, B.; Trincone, A.; Giordano, A.; Rossi, M.; Moracci, M. *The Journal of Biological Chemistry* **2003**, *278*, 14622-31.

(182) Cobucci-Ponzano, B.; Mazzone, M.; Rossi, M.; Moracci, M. *Biochemistry* **2005**, *44*, 6331-42.

(183) Cobucci-Ponzano, B.; Trincone, A.; Giordano, A.; Rossi, M.; Moracci, M. *Biochemistry* **2003**, *42*, 9525-31.

(184) Kelley, L. a; Sternberg, M. J. E. *Nature Protocols* **2009**, *4*, 363-71.

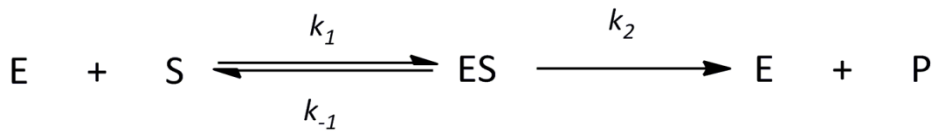
(185) Lammerts van Bueren, A.; Ardèvol, A.; Fayers-Kerr, J.; Luo, B.; Zhang, Y.; Sollogoub, M.; Blériot, Y.; Rovira, C.; Davies, G. J. *Journal of the American Chemical Society* **2010**, *132*, 1804-6.

(186) Sakurama, H., Fushinobu, S., Hikada, M., Yoshinda, E., Honda, Y., Ashida, H, Kumagai, H., Yamamoto, K., Katayama, T. *Journal of Biological Chemistry* **2012**, 287, 16709-19

# Appendix I: Basic Enzyme Kinetics

## Michaelis-Menten Kinetics

In 1913, Michaelis and Menten proposed a simple model to explain the observed kinetic behavior of an enzyme. This model accounted for the relationship between the rate of enzyme catalysis and the concentration of the substrate. This model was expanded upon in 1925 by Briggs and Haldane to include the concept of steady state. A general scheme representing an enzyme catalyzed reaction is shown below. Free enzyme, E, combines with free substrate, S, to form the enzyme-substrate complex, ES, which is then turned over to yield the product, P.




Under steady state conditions:

$$\frac{d[\mathrm{ES}]}{\mathrm{dt}} = \mathrm{k}_1[\mathrm{E}][\mathrm{S}] - \mathrm{k}_{-1}[\mathrm{ES}] - \mathrm{k}_2[\mathrm{ES}] = 0 \qquad [1]$$

The total concentration of enzyme, $[\mathrm{E}]_o$, is equal to the sum of the concentration of free enzyme, [E], plus the concentration of the enzyme bound in the ES complex [ES].

$$\mathrm{E}_o = [E] + [ES] \qquad [2]$$

Solving for [ES] by combining equations [1] and [2] gives us:

$$[\mathrm{ES}] = \frac{[\mathrm{E}]_o[\mathrm{S}]}{[\mathrm{S}] + \frac{\mathrm{k}_{-1} + \mathrm{k}_2}{\mathrm{k}_1}} \qquad [3]$$

Under steady state conditions the initial velocity of the reaction, v, will be equal to the rate of formation of product:

$$v = \frac{\mathrm{dP}}{\mathrm{dt}} = \mathrm{k_2[ES]} \qquad [4]$$

Substituting the expression for [ES] from [3] into [4] one gets the following expression for the initial rate of reaction:

$$v = \frac{\mathrm{k_2[E]_o[S]}}{\mathrm{[S] + \frac{k_{-1} + k_2}{k_1}}} \qquad [5]$$

The ratio of rate constants, $(k_{-1} + k_2)/k_1$, is defined as $K_M$, the Michaelis constant, while $k_2$ is defined as $k_{cat}$, the catalytic constant or turnover number of the enzyme. Equation [5] can be further simplified as equation [6] which is the commonly seen form of the Michaelis-Menten equation:

$$v = \frac{\mathrm{k_{cat}[E]_o[S]}}{\mathrm{[S] + K_M}} \qquad [6]$$

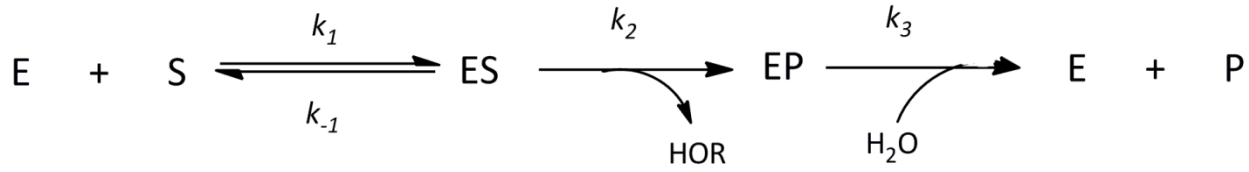
When the initial reaction rate, *v*, is equal to half the maximum rate, $v = V_{max/2}$, then the substrate concentration is equal to $K_M$. The Michaelis constant in its simplest form is a measure of the binding affinity of an enzyme for a particular substrate. Since $k_2$ is included in $K_M$ the rate of enzyme turnover approaches zero when $K_M$ approaches the dissociation constant $K_d$. When $[S] >> K_M$ then *v* approaches $V_{max}$ and the rate becomes largely independent of [S]. In this case the Michaelis-Menten equation can be rewritten as:

$$V_{max} = k_{cat}[\mathrm{E}]_\mathrm{o} \qquad [7]$$

When the reverse is true, [S] << $K_M$, then the initial rate of the enzyme catalyzed reaction is proportional to [S] as shown in equation [8]:

$$v = \frac{k_{cat}[E]_o[S]}{K_M} \qquad [8]$$

The Michaelis-Menten equation can be expanded to represent more complex enzymatic reactions where two distinc reaction steps occur—such as the double-displacement reaction of a retaining glycosidase. The reaction scheme for this mechanism is shown below and is as follows: free enzyme, [E], and substrate, [S], combine to form the enzyme substrate complex, [ES], with a rate constant $k_1$ for the association and $k_{-1}$ for the dissociation. The conversion of ES to EP is the flycosylation step ($k_2$) and the turnover of EP to P is the deglycosylation step, represented by $k_3$.




Assuming that steady state for both ES and EP are reached, then:

$$k_2[ES] = k_3[EP] \qquad [9]$$

and therefore:

$$\frac{d[ES]}{dt} = k_2[E][S] - k_{-1}[ES] + k_2[ES] = 0 \qquad [10]$$

The total enzyme concentration is the sum of the free enzyme plus all of the enzyme bound species:

$$\mathrm{E}_o = [E] + [ES] + [EP] \qquad [11]$$

Substituting equation [9] into equation [11] gives:

$$\mathrm{E}_o = [E] + [ES] + \frac{k_2}{k_3}[EP] \qquad [12]$$

Solving equation [10] for [E] and substituting that into equation [12] followed by rearrangement to isolate [ES] gives:

$$[\mathrm{ES}] = \frac{\frac{k_2 k_3}{k_2 + k_3}[\mathrm{E}]_0[\mathrm{S}]}{\left(\frac{k_3}{k_2 + k_3}\right)\left(\frac{\mathrm{k}_{-1} + \mathrm{k}_2}{\mathrm{k}_1}\right) + [\mathrm{S}]} \qquad [13]$$

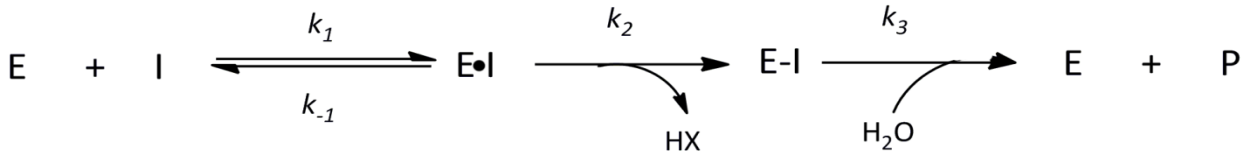
Comparing this to the simplified form of the Michaelis-Menten equation, [6], we can see that

$$k_{cat} = \frac{k_2 k_3}{k_2 + k_3} \qquad [14]$$

$$\mathrm{K_M} = \left(\frac{k_3}{k_2 + k_3}\right)\left(\frac{\mathrm{k}_{-1} + \mathrm{k}_2}{\mathrm{k}_1}\right) \qquad [15]$$

## Enzyme kinetics in the presence of a mechanism-based inactivator

The inactivation of an enzyme by a mechanism based inactivator and its eventual reactivation by water is shown in the following scheme:




This is similar to the kinetic model for the double displacement reaction described above. If $k_3$ << $k_2$, and as $k_3$ approaches zero, an extremely stable glycosyl-enzyme intermediate will accumulate and the enzyme will be inactivated. The kinetic scheme can be simplified to:

This model predicts a time-dependent inactivation of the enzyme in the presence of the mechanism-based inactivator. If [I] >> [E] then [I] can be assumed to be constant during the reaction and pseudo-first order kinetics will be observed with respect to [E]. This equation is similar to the Michaelis-Menten expression seen earlier:

$$v = \frac{\mathrm{k_i}[\mathrm{E}]_\mathrm{o}[\mathrm{I}]}{[\mathrm{I}] + \mathrm{K_I}} \qquad [16]$$

Where $k_i$ is the rate constant for inactivation and $K_I$ ($K_I = k_{-1}/k_1$) is an apparent dissociation constant for all forms of enzyme bound. Assuming $k_{-1} >> k_2$ then [16] can be reqritten as:

$$v = \mathrm{k_{obs}}\,[\mathrm{E}] \qquad [17]$$

where

$$k_{obs} = \frac{k_{\mathrm{i}}[\mathrm{I}]}{[\mathrm{I}] + \mathrm{K_I}} \qquad [18]$$

If $K_I$ >> [I] then equation [20] can be written as

$$k_{obs} = \frac{k_{\mathrm{i}}[\mathrm{I}]}{\mathrm{K_I}} \qquad [19]$$

Where $k_{obs}$ is the observed rate constant for the time-dependent loss of enzyme activity, which is obtained by fitting the values for the residual enzyme activity to an exponential decay equation.

# Appendix II: Graphical Representation of Data

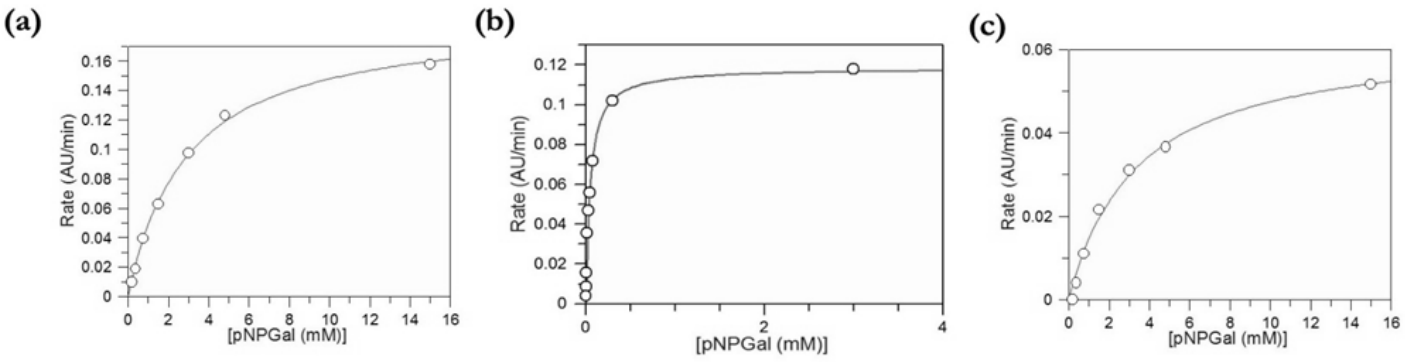



Figure B.1. Michaelis-Menten plots for the hydrolysis of pNP-Gal by (a) WT YesZ, (b) E145A mutant of YesZ, and (c) E298A mutant of YesZ. Fits shown are to the Michaelis-Menten equation.

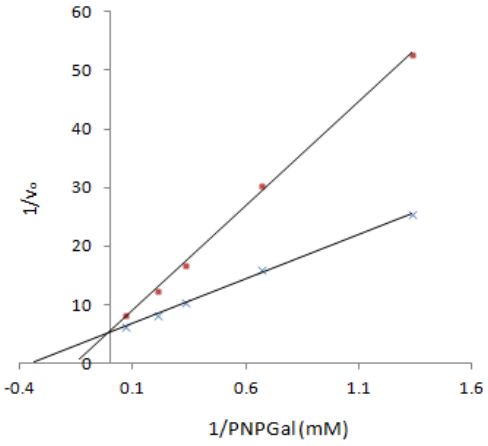



Figure B.2 Lineweaver-Burk plot of initial reaction velocity versus concentration of PNPGal, in the ▪ presence and x absence of 100 µM IPTG

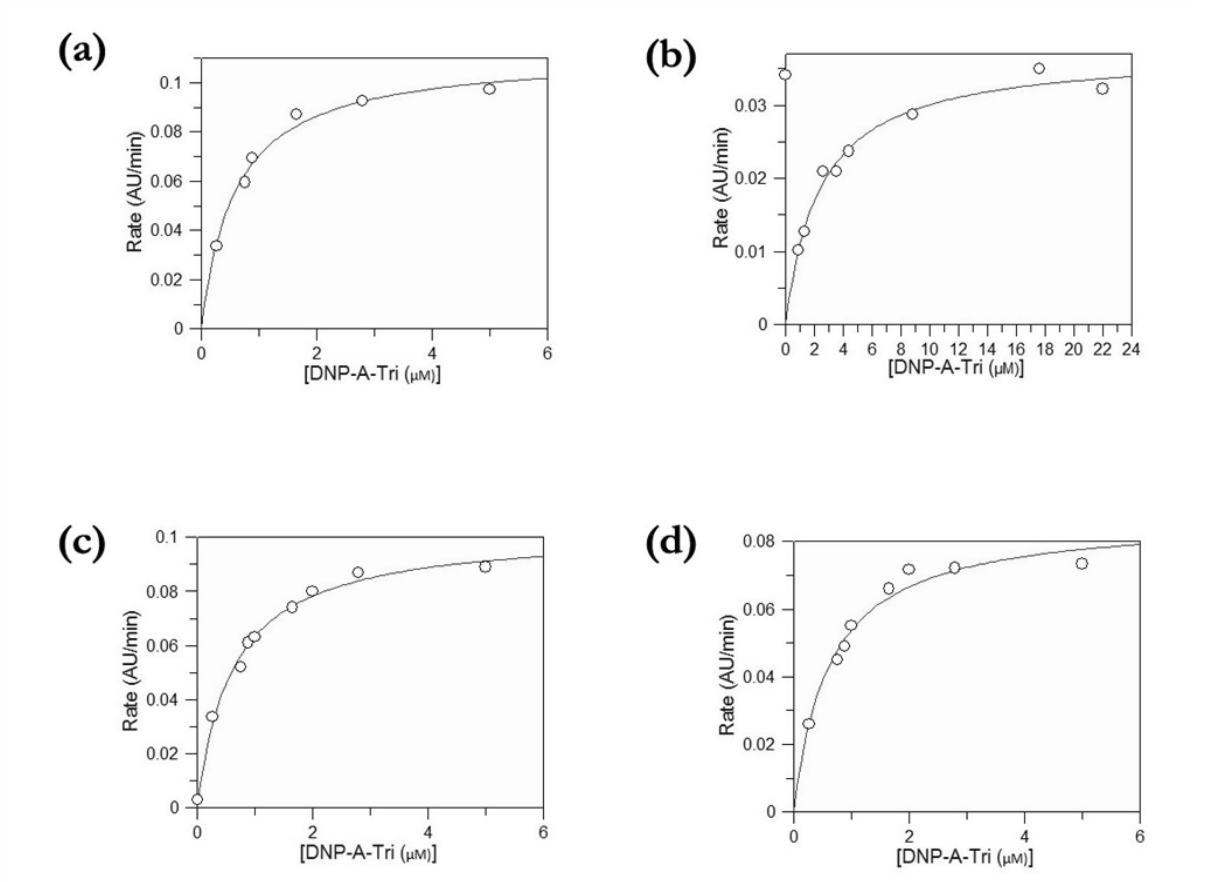



Figure B.3. Michaelis-Menten plots for the hydrolysis of DNP-A-Tri by (a) WT EABase, (b) E354A mutant of EABase, (c) E467A mutant of EABase, and (d) D429A mutant of EABase. Fits shown are to the Michaelis-Menten equation.

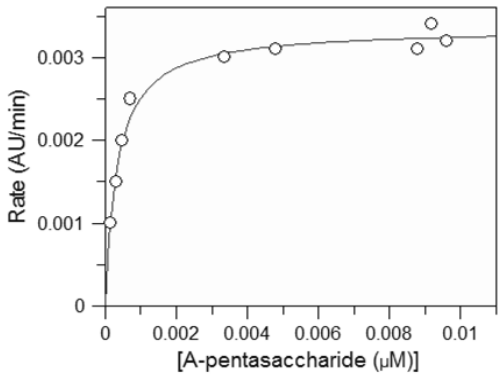



Figure B.4. Michaelis-Menten plots for the hydrolysis of A-pentasaccharide by WT EABase. Fit shown is to the Michaelis-Menten equation.

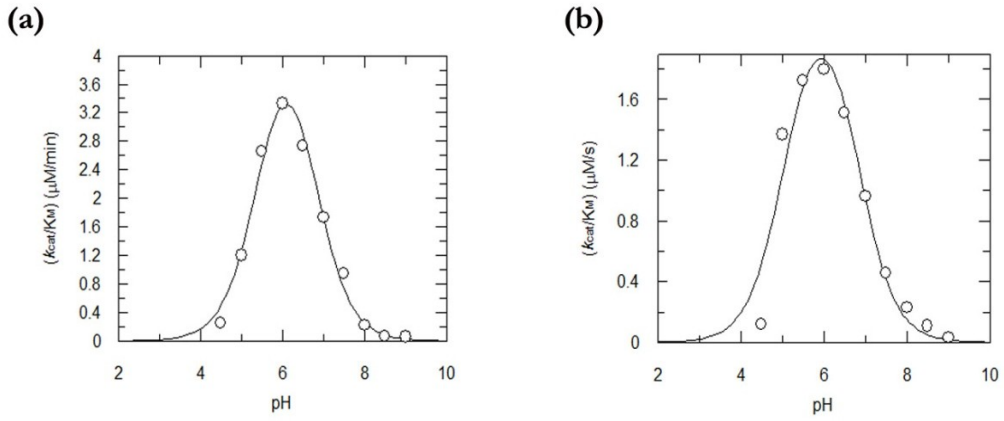



Figure B.5. pH profile of $k_{cat}/K_M$ for BT2970 and BT4136 (not shown in Chapter 4)

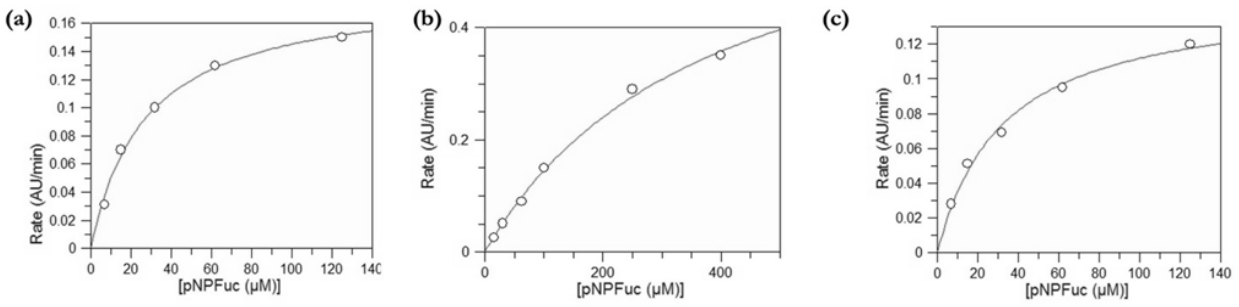



Figure B.6. Michaelis-Menten plots for the hydrolysis of pNPFuc by (a) WT BT2970, (b) E288A mutant of BT2970, and (c) E306A mutant of BT2970. Fits shown are to the Michaelis-Menten equation.

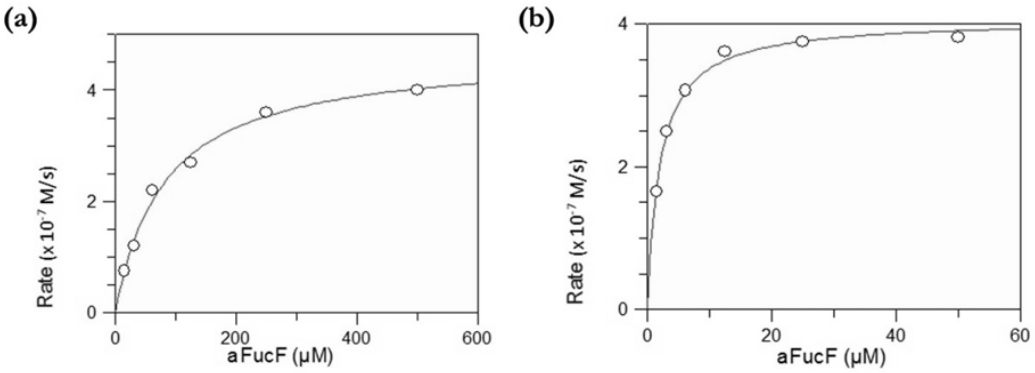



Figure B.7. Michaelis-Menten plots for the hydrolysis of αFucF by (a) WT BT2970, and (b) E288A mutant of BT2970. Fits shown are to the Michaelis-Menten equation.

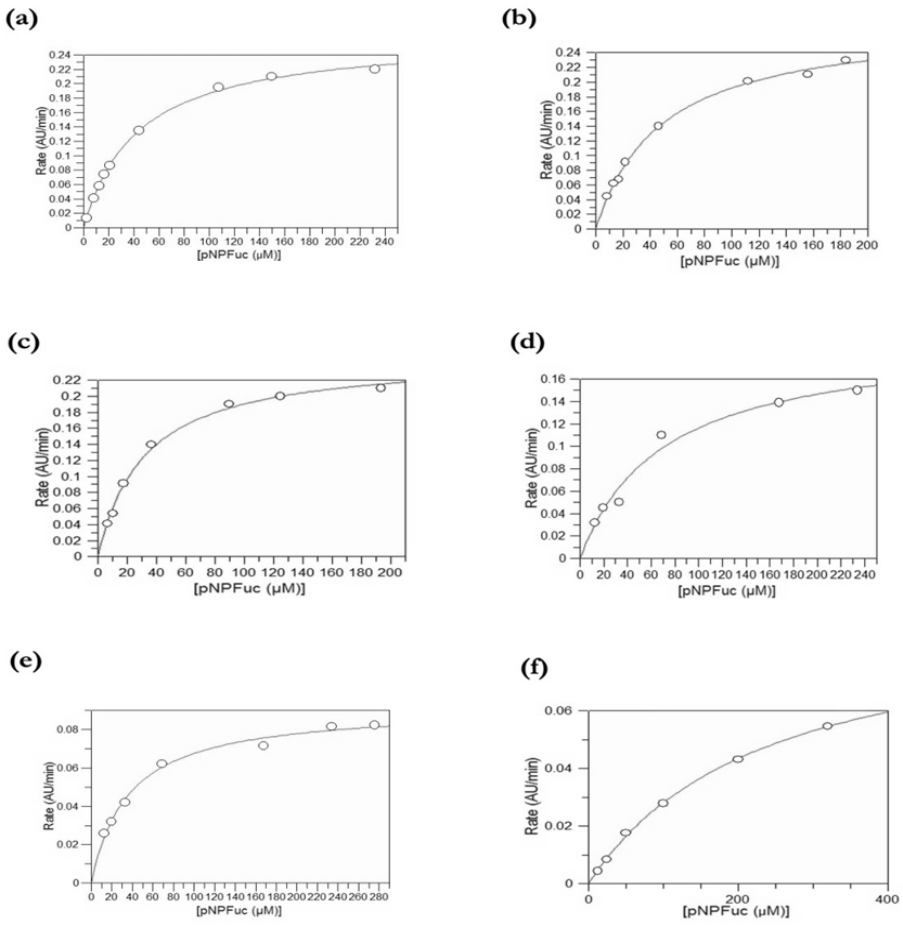



Figure B.8 Michaelis-Menten plots for the hydrolysis of pNPFuc by (a) WT BT4136, (b) D286A mutant of BT4136, (c) E305A mutant of BT4136, (d) E298A mutant of BT4136, (e) D307A mutant of BT4136, and (f) E249A mutant of BT4136. Fits shown are to the Michaelis-Menten equation.

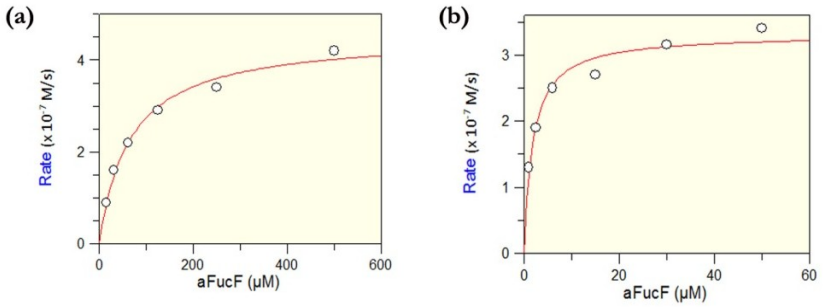



Figure B.9. Michaelis-Menten plots for the hydrolysis of αFucF by (a) WT BT4136, and (b) E249A mutant of BT4136. Fits shown are to the Michaelis-Menten equation.

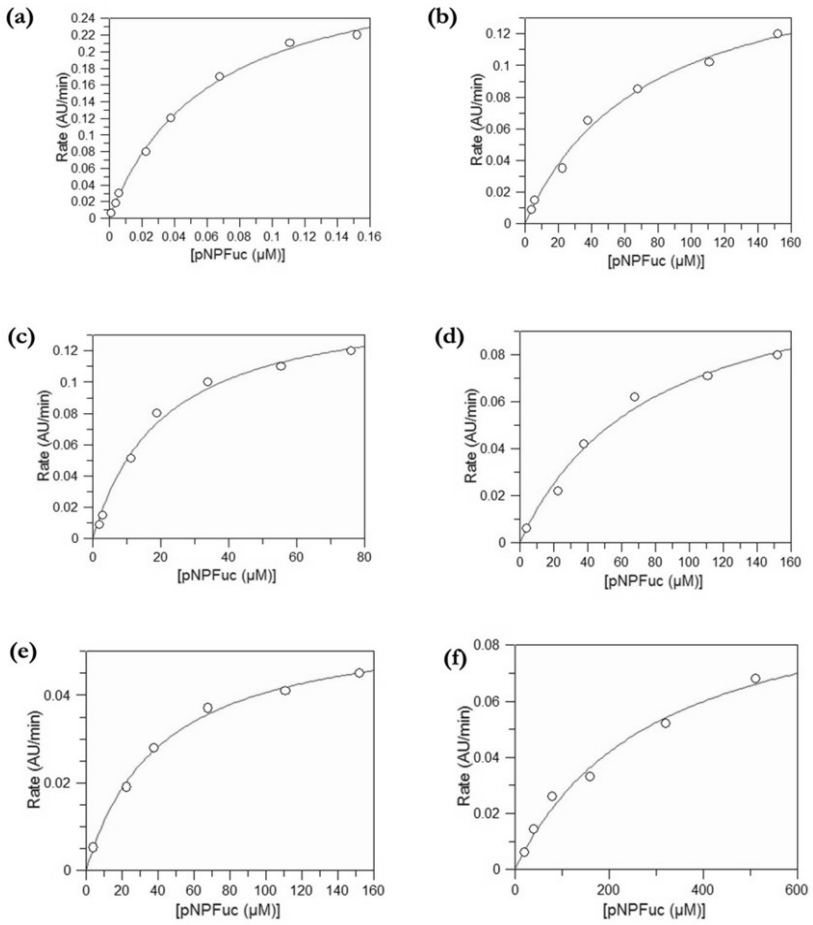



Figure B.10 Michaelis-Menten plots for the hydrolysis of pNPFuc by (a) WT BT1625, (b) D286A mutant of BT1625, (c) E305A mutant of BT1625, (d) E298A mutant of BT1625, (e) D307A mutant of BT1625, and (f) E249A mutant of BT1625. Fits shown are to the Michaelis-Menten equation.

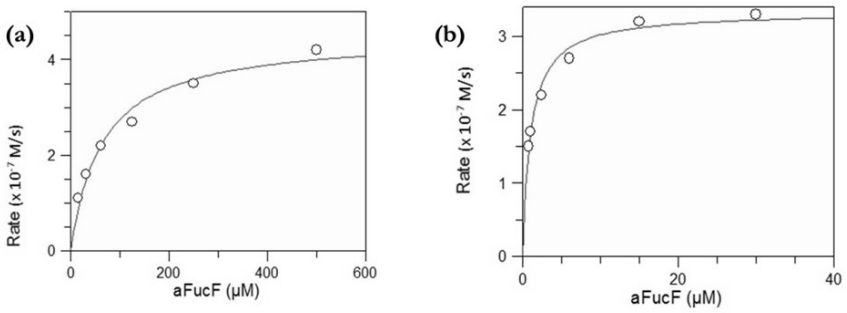



Figure B.11. Michaelis-Menten plots for the hydrolysis of αFucF by (a) WT BT1625, and (b) E249A mutant of BT1625. Fits shown are to the Michaelis-Menten equation.