Directed Evolution of a Bacterial Sialidase and Characterization of Mechanism Based Inactivation of Glycosidases

by

Saeideh Shamsi Kazem Abadi

M.Sc., Simon Fraser University, 2011

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Molecular Biology and Biochemistry
Faculty of Science

SIMON FRASER UNIVERSITY

Fall 2017

Approval

Name:	**Saeideh Shamsi Kazem Abadi**
Degree:	**Doctor of Philosophy**
Title:	**Directed Evolution of a Bacterial Sialidase and Characterization of Mechanism Based Inactivation of Glycosidases**
Examining Committee:	**Chair: Barry Honda** Professor
	Andrew Bennet Senior Supervisor Professor
	David Baillie Supervisor Professor Emeritus
	William Davidson Supervisor Professor
	David Vocadlo Supervisor Professor
	Margo Moore Internal Examiner Professor Biological Sciences
	Harry Brumer External Examiner Professor Chemistry University of British Columbia
Date Defended/Approved:	December 7, 2017

Abstract

Sialic acids are often found at the terminal positions on the glycan chains that adorn all vertebrate cells and glycoproteins. This prominent position confers an essential role to sialic acid residues in biology, evolution and disease propagation. The most widespread sialic acid family members are *N*-acetylneuraminic acid, *N*-glycolylneuraminic acid and Kdn, which is an abbreviation for 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid. Enzymes that catalyze the removal of carbohydrate linkages from biological molecules are called glycoside hydrolases (GHs). These enzymes have been categorized into more than 130 different families. Glycoside hydrolase family 33 (GH33) contains *exo*-sialidases (E.C. 3.2.1.18, neuraminidases), from both eukaryotes and prokaryotes, which catalyze the hydrolysis of sialic acid from glycoconjugates. Interestingly, subtle differences exist in both the structure of the particular sialic acid and its position of attachment to glycoconjugate chains between humans and other mammals. These differences are indicators of the unique aspects of human evolution, and are relevant to understanding an array of human conditions. The present thesis reports on routes that we explored to further unravel the importance of sialic acids. We developed tools to probe for various sialic acid structures such as Kdn. To this end, we constructed a random mutant library of the neuraminidase from the soil bacterium *Micromonospora viridifaciens* (*Mv*NA*)* and identified a number of recurring mutations in the sialidase gene which lead to a more efficient hydrolysis of synthetic natural substrate analogues such as 8FMU α-Kdn-(2$\rightarrow$6)-β-D-Galp. We also using the available structure of wild type *Mv*NA bound to the natural inhibitor, DANA, to identify amino acids potentially involved in recognition and binding to acetylated sialic acids and generated genetic libraries which we used along with positive and negative evolutionary screens to identify several clones capable of hydrolyzing Kdn glycosides more efficiently than Neu5Ac substrates. Kinetic studies on these clones allowed for determination of enzyme efficiencies and specificities. We also report our study of covalent inhibition of α-glucosidase from *Saccharomyces cerevisiae* (GH13). The measured pH-rate profiles for inhibition and reactivation as well as the corresponding catalytic and inhibitory proficiencies suggested that inhibition results from the formation of carbenium ions in the active site that are trapped rapidly by an enzymatic residue.

Keywords: Alpha glucosidase; Inhibition; Sialidase; Kinetics; Directed Evolution

Dedication

This work is dedicated to my mom, Zahra, who has been my greatest source of strength and support. I love you mom! I would also like to dedicate this thesis to my brothers Sajjad and Saeid, the "best' brothers any sister could hope for.

Acknowledgements

I would like to thank my senior supervisor, Dr. Andrew Bennet for his constant support and dedication to my educational goals. He has been a great mentor with many “interesting” projects that have been a source of great science and personal growth. His patience and calm demeanor have been instrumental in the successful completion of my degree. It has been an honor to work in his laboratory and learn from his wisdom.

I would also to thank my committee members, Dr. David Baillie, Dr. David Vocadlo and Dr. William Davidson for their guidance and advice over the course of my PhD studies. This work would not have been possible without their support.

I also like to thank my examining committee, Dr. Margo Moore and Dr. Harry Brumer for their time and valuable feedback.

Last but most definitely not the least, I would like to extend my sincere thanks to the Bennet Lab members that I have had the pleasure of working with. The friendship and support offered by them has been a great source of joy and one I hope that will last for a lifetime. All the coffee runs, Christmas potlucks, late night youtube videos and inside jokes are memories that I will cherish forever.

Table of Contents

List of Tables

List of Figures

List of Schemes

List of Equations

Chapter 1. Introduction

1.1. Carbohydrates

Carbohydrates are the most abundant group of “biogenic substances”, which are defined as biomolecules, produced by living organisms. Carbohydrates were historically defined as polyhydroxylated aldehydes or ketones because the original molecular formula for a carbohydrate was defined as $C_n(H_2O)_m$. However, many compounds that are recognized as carbohydrates do not possess this formula, for example, acetamido containing sugars including *N*-acetylglucosamine and sialic acid. At present, there are a variety of interrelated classification schemes for sugars, which are in common usage, one of which is based on the degree of polymerization, that is carbohydrates can be subdivided into three major groups; (1) monosaccharides (single sugar molecules); oligosaccharides (2-9 monomers joined through glycosidic linkages) and polysaccharides (polymers that are made up from more than 10 monosaccharide units). Figure 1.1 depicts an example from each subgroup.

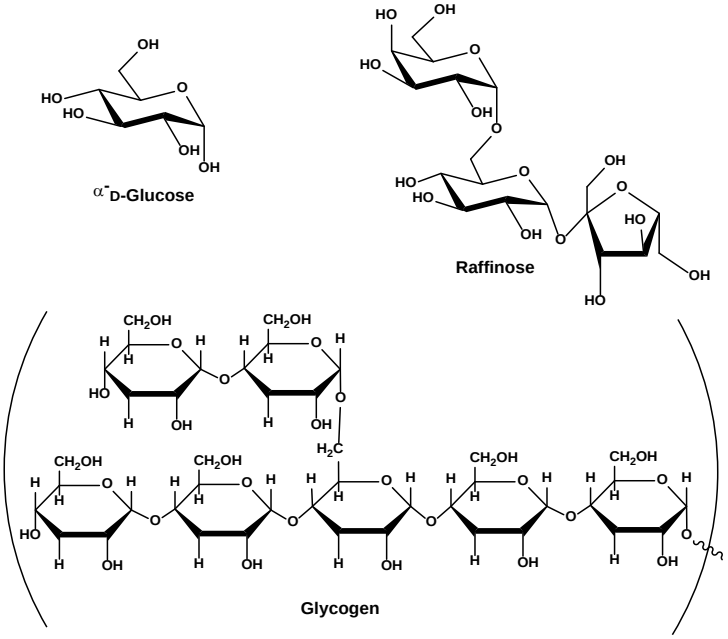


Figure 1.1 Structural representation of a monosaccharide (Glucose), a trisaccharide (Raffinose) and a polysaccharide (Glycogen).

Within this classification scheme, the monosaccharides can be further classified based on criteria such as the ease of the oxidation of the carbonyl carbon (aldose vs. ketose), the number of carbons (trioses, tetroses, pentoses, etc.), the absolute stereoconfiguration (L or D), the ring size (four ring furan vs. five ring pyran) and the diastereomer formed by ring closure to give an anomeric hydroxyl group (α or β). Figure 1.2 depicts the hexoketose α-D-fructofuranose.

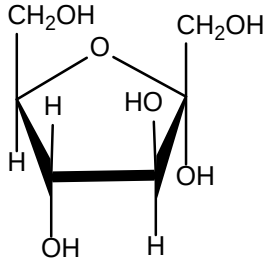


Figure 1.2 Modified Haworth projection of the ketose, α-D-fructose.

The prevalence of sugars in nature means that interactions between carbohydrates and their cognate receptor have evolved to generate a multitude of roles in biological processes. For instance, carbohydrates are used as energy sources (both long and short term), major structural components of cells, and as intermediates in metabolism. In addition, carbohydrates can be covalently conjugated to proteins (glycoproteins) and lipids (glycolipids). The glycan or sugar components of glycoconjugates are often comprised of pentoses or hexoses attached by either an α- or β-glycosidic bond to the underlying carbohydrate, protein or lipid. These glycoconjugates are often displayed on cellular surfaces and are widely implicated in cellular communication and signaling events. Components of cellular signaling pathways, namely glycoconjugates, are often promising targets for development of therapeutics against a wide range of diseases including bacterial and viral infections. To this end, the field of glycobiology encompasses the understanding of the various relationships between the biologically relevant carbohydrates and their interactions with proteins and other biomolecules.

1.2. Sialic Acids

Sialic acids or *N*-acetylneuraminic acids are a family of nine carbon *α*-keto acids predominantly found at the terminal positions on glycan chains that adorn all vertebrate cells and glycoproteins[1]. The two names for this family of sugars stem from their independent discovery by two separate groups: Ernest Klenk in 1941[2] and Gunnar Blix in 1936 [3]. The Klenk group isolated a sugar cleavage product from brain lipids (gangliosides) and named it “neuraminic acid” while the Blix group identified a sugar component from bovine submaxillary mucin and aptly named it “sialic acid”.

Sialic acids, originally discovered within the Deuterostome lineage of animals and associated microbes, are a subset of a family of 9-carbon backbone monosaccharides called nonulosonic acids[4]. It is thought that sialic acids are invented in the common ancestor of protostomes and deuterostomes and while they became essential in deuterostomes, they were partially or completely discarded in some protostome lineages.[4]

Figure 1.3 depicts the structure of the most abundant sialic acid, *N*-acetylneuraminic acid or Neu5Ac (5-acetamido-3,5-dideoxy-D-*glycero*-D-*galacto*-non-2-ulosonic acid). The nine-carbon backbone of a sialic acid incorporates a carboxylic acid at C-1 that is deprotonated at physiological pH and thus sialic acids are negatively charged. The numbering of the sialic acid structure begins at the carboxylate carbon and continues along the chain. The anomeric centre at C2 is assigned a stereochemistry of α- or β- according to the configurational relationship between the C7 carbon and C1 carbonyl groups being in trans (α) or cis (β) configuration (Figure 1.3). In nature, sialic acid is found only in the α-configuration when it is a component of a glycoconjugate, and in the β-configuration in the high-energy donor form, CMP-sialic acid.

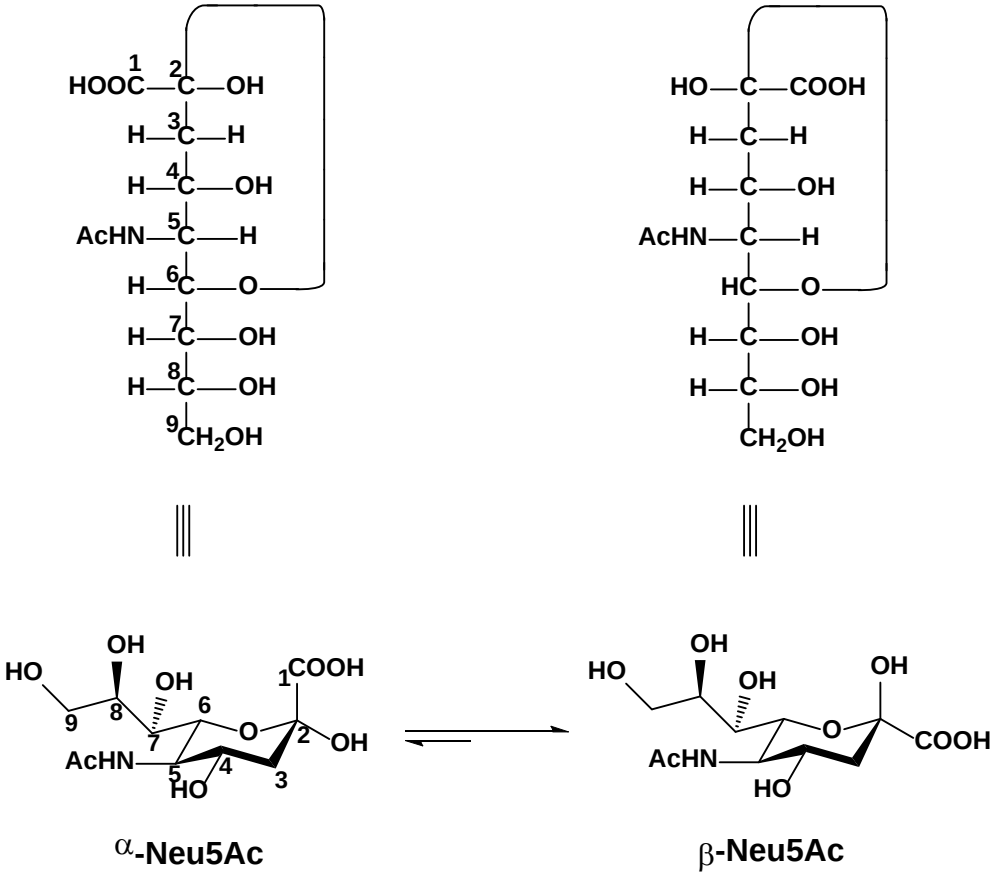


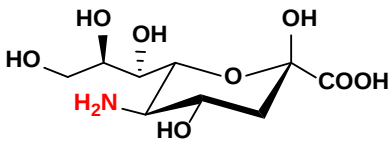
Figure 1.3 Structure of α- and β-Neu5Ac, the most common sialic acid shown in Fisher projection and its corresponding pyranose chair conformations.

In the process of cell apoptosis, degradation or recycling of cell surface glycans that is preceded or induced by biological processes such as differentiation or cancer cell

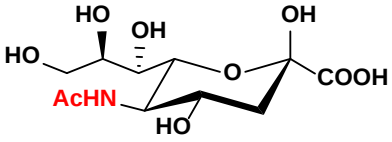
metastasis, sialic acids attached to glycoconjugates must first be removed. Desialyation of glycoconjugates occurs in both the endosome/lysosomal compartments through the action of a group of enzymes called sialidases that catalyze the cleavage of the glycosidic linkages between sialic acid residues and the underlying glycan chain [5].

1.2.1. Diversity of Sialic Acids

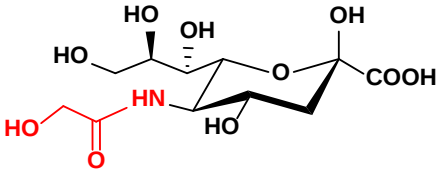
The sialic acid family shows remarkable structural diversity, with the current number of members being over 50 that occur naturally [6]. The first level of sialic acid diversity results from natural modifications at various carbon atoms of the "parent molecule" neuraminic acid (Neu). The largest structural variations are at carbon 5, where substitution with an acetamido, hydroxyacetamido or hydroxyl moiety leads to the following sugars: 5-*N*-acetylneuraminic acid (Neu5Ac), 5-*N*-glycolylneuraminic acid (Neu5Gc) or deaminoneuraminic acid (Kdn), which along with non-acetylated neuraminic acid (Neu) make up the four major core types of sialic acids (Figure 1.4)



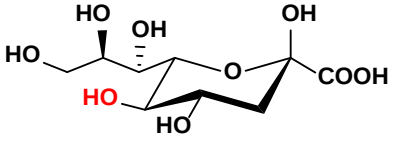
Neuraminic acid (Neu)



N-acetylneuraminic acid (Neu5Ac)



N-Glycolylneuraminic acid (Neu5Gc)



Deaminated neuraminic acid (Kdn)

Figure 1.4 Structures of natural sialic acids with substitutions at C-5 position.

Further structural diversity is achieved through a combination of the above-mentioned variations at carbon 5 with substitution of hydroxyl groups at carbons 4, 7, 8 and 9 with *O*-

acetyl – the esterification of hydroxyl groups with acetic acid, *O*-lactyl, *O*-methyl, *O*-sulfate or phosphate groups (Figure 1.5)[6].

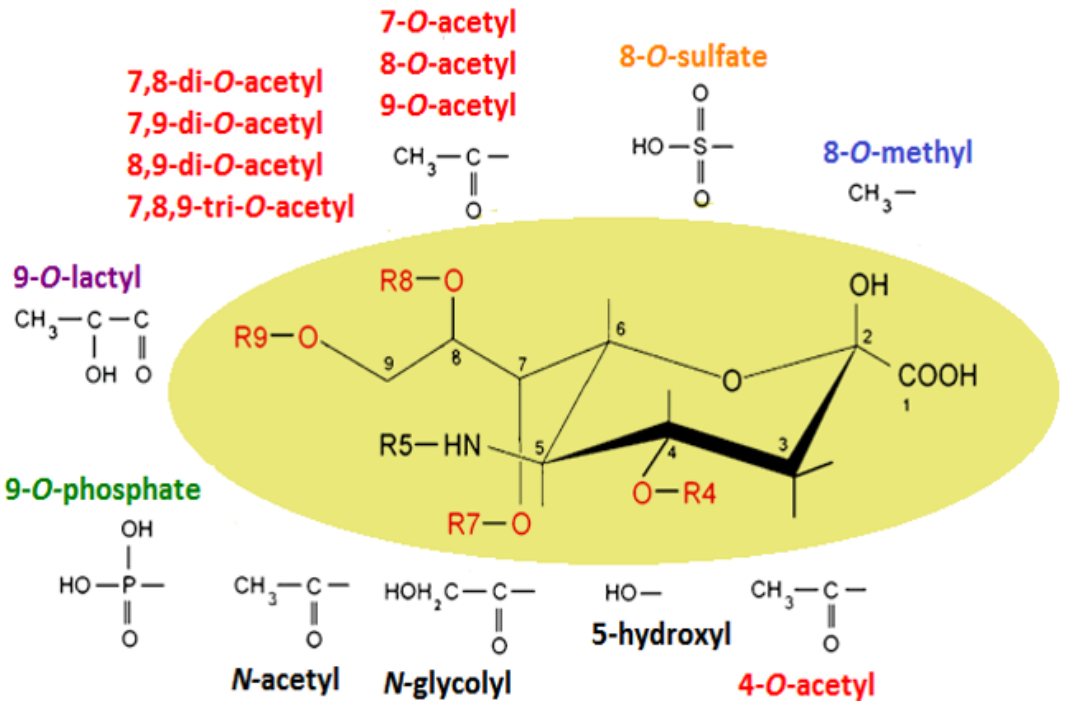


Figure 1.5 Structural diversity of sialic acid family.

Sialic acids can occur in nature as free monosaccharides but they are generally covalently linked to glycan chains, and are called sialosides. This introduces a second level of diversity as the result of the different α-sialoside linkages formed between C-2 of the sialic acid and the underlying sugar by specific sialyltransferases that catalyze the reaction. The most common linkages are between C2 of the sialic acid to either the 3- or 6-hydroxyl group of galactose (Gal) residues, the 6-hydroxyl group of *N*-acetylglucosamine (GlcNAc) or *N*-acetylgalactosamine (GalNAc) residues and to the C-8 hydroxyl of another sialic acid [5,6]. There are some less common linkages that occur between the C-2 of sialic acid to C-3 of *N*-acetylgalactosamine, C-6 of *N*-acetylglucosamine and glucose, C-4 of galactose and *N*-acetylgalactosamine, and C-9 linked to another sialic acid [6].

Certain sialoside linkages, which can include the sialic acid modifications noted above, typically show tissue-specific and developmentally regulated expression. Furthermore, certain modifications such as *O*-acetylation may be restricted to certain sialic acid residues at particular positions within the glycan chain of the glycoconjugate[1]. Such findings suggest specific roles for these linkages and modifications. The tissue and organ specificity of sialic acid modifications, is evident in the difficulty of cross transmission of avian influenza A to humans and conversely of human influenza A to birds. Unlike the human influenza virus A which preferentially recognizes and binds to α-2,6-linked sialic acid found on the surfaces of epithelial cells, typically in the nose, throat, and lungs of mammals, the avian influenza virus A receptors bind the isomeric α-2,3-linked sialosides selectively. On the human's non-ciliated epithelial cells in the respiratory tract, including larynx and trachea, α-2,6-linked sialosides are expressed more abundantly than the α-2,3-isomers, and this difference is why avian influenza A viruses are less infectious towards humans [7].

The diverse combinations of different glycosidic linkages with the multitude of possible natural modifications at various carbons generate a huge diversity of sialic acid containing biomolecules. This degree of variety along with their prominent terminal positions on glycan chains contributes to the enormous variety of glycan structures on cell surfaces and confers an essential role to sialic acid residues in biology, evolution and disease propagation. The specificity of sialidases for cleavage of linkage specific glycosidic bonds introduces an added layer of complexity to the biological roles of sialic acid glycans and this is an area of in depth investigation[1].

Interestingly, subtle differences exist in both the structure of the particular sialic acids and its position of attachment to glycoconjugate chains between humans and other mammals. These differences are indicators of the unique aspects of human evolution, and are relevant to understanding an array of human conditions.

1.2.2. Sialic Acid Metabolism

1.2.2.1 Biosynthesis of Sialic Acid

The general pathway for the biosynthesis of Neu5Ac in bacteria and vertebrates has been well characterized and is depicted in Figure 1.6.

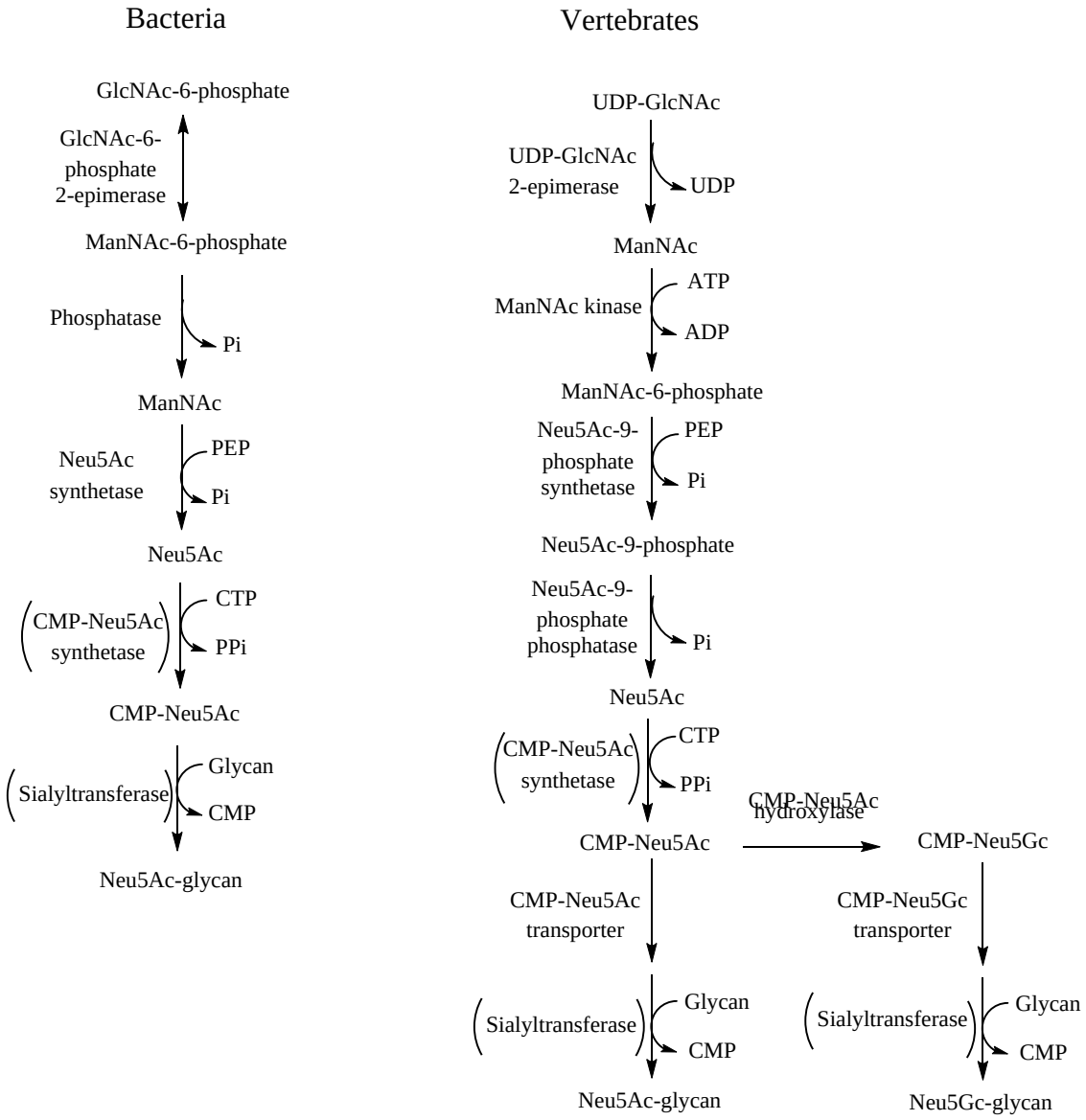


Figure 1.6 Biosynthesis pathway of Neu5Ac and Neu5Gc in bacteria and vertebrates. The enzymes that are common between bacteria and vertebrates are enclosed in parentheses.

The epimerase catalyzed conversion of a GlcNAc precursor to ManNAc, which is common between vertebrates and bacteria, is unique to the sialic acid synthesis pathway, and as such pathways are generally are only shown from this step onwards[6]. In vertebrates, the generated ManNAc is further phosphorylated by ManNAc kinase using ATP to give ManNAc-6-phosphate [8]. The next step which takes place primarily in the cytosol involves the condensation of ManNAc-6-P with phosphoenolpyruvate (PEP) to give Neu5Ac-9-phosphatethat is then dephosphorylated to produce Neu5Ac. In bacteria, in contrast, the epimerase generated ManNAc-6-phosphate is initially dephosphorylated to generate ManNAc which is then condensed with PEP to give Neu5Ac directly [9]. The activation of Neu5Ac to CMP-β-Neu5Ac by the enzyme CMP-Neu5Ac synthetase, which catalyzes the addition of a cytidine 5'-monophosphate (CMP) group from cytidine 5'-triphosphate (CTP) to Neu5Ac is common between bacteria and vertebrates [10]. In vertebrates, these steps take place in the nucleus followed by transportation of CMP-β-Neu5Ac to the Golgi body by a CMP-sialic acid transporter [11]. The CMP-β-Neu5Ac donor is used to add Neu5Ac onto acceptors by the sialyltransferases in both bacteria and vertebrates [12].

1.2.2.2 Degradation of Sialic acid

There are multiple enzymatic pathways for the break down and degradation of sialic acids and these include: sialidases, esterases, and sialic acid aldolases. Sialidases, catalyze the removal of sialic acid from glycoconjugates by hydrolyzing the glycosidic bond. This removal however, must often be preceded by the hydrolysis of any acetyl groups present on sialoside by esterase action. Following the removal of sialic acid, it is transported to the cytosol, where it can be recycled to ManNAc and pyruvate by sialic acid aldolase.

1.2.3. Biological Significance of Sialic Acid

As the terminal sugar residue on cell surface glycan chains, sialic acids play important roles in cellular communication and are indicated in multitude of biologically significant roles.

The negative charge caused by the carboxylate group is critical for electrostatic interactions that are fundamental to cellular surfaces adorned with sialic acid [13]. An example of this is the observed increase in sialic acid coverage of metastatic tumour cells, where the repulsion generated from the close interactions of negatively charged sialic acids leads to repulsion of adjacent cells and the release of them from the primary tumour to the blood stream, a process that results in establishment of secondary tumour sites [14].

Sialic acids also serve as recognition moieties for a variety of sialic acid binding proteins of intrinsic and extrinsic origin, including antibodies and parasitic enzymes [15]. An example of this is the influenza virus which possesses a sialic acid-recognizing receptor called haemagglutinin that recognizes and binds to host cell surface sialic acids and allows for attachment of the virus to the host's cell surface. This recognition event and the subsequent multivalent interactions are crucial for entry of the virus into the host cell, which is the initiation event for infection [16].

Additionally, the terminal location of sialic acid residues contributes to its function of masking the penultimate saccharide residue. The classic example involves the mammalian regulation of glycoproteins where terminal sialic acids prevent recognition of galactose residues by the Ashwell receptors in the liver, an interaction that mediates glycoprotein clearance [17]. Therefore, sialic acids play a vital role in maintaining the serum half-lives of glycoproteins.

1.2.4. Kdn

Deaminoneuraminic acid, (Kdn), which has the IUPAC name 3-deoxy-D-*glycero*-D-*galacto*-non-2-ulosonic acid, was first detected and isolated from the cortical alveolar polysialoglycoproteins of rainbow trout (*Salmo gairdneri*) in 1986 by the Yasuo Inoue group [18]. It was speculated that biosynthesis of Kdn might occur via deacylation and deamination of Neu5Ac, however, no evidence exists for this process [19]. De novo synthesis of Kdn has been hypothesized to involve the condensation of PEP and mannose 6-phosphate (Man-6-P), giving rise to Kdn 9-phosphate (Kdn-9-P) which is dephosphorylated to Kdn, activated to CMP-Kdn, and ultimately transferred to an acceptor sugar. While some of the enzymes indicated in this pathway, preferably recognize Kdn,

many others are common to the Neu5Ac synthesis [19]. Figure 1.7 is a depiction of this proposed pathway.

Although structurally Kdn is very similar to Neu5Ac, as shown in Figure 1.4, this substitution of the aminoacyl group on C-5 on Neu5Ac or Neu5Gc by a hydroxyl group makes glycoproteins adorned with this sugar nearly completely refractory to bacterial *N*-acetylneuraminidases that are commonly used for structure-function studies[20,21]. This intrinsic resistance to hydrolysis by bacterial enzymes is of biological importance as it may be a mechanism of protecting Kdn-containing glycoconjugates from bacterial enzymes. Kdn-containing glycoconjugates have been shown to be widely distributed in many cell types, including mammalian tissues, human lung and ovarian cancer cells, and human red blood cells [22,23]. Further research into the metabolism and importance of Kdn is ongoing.

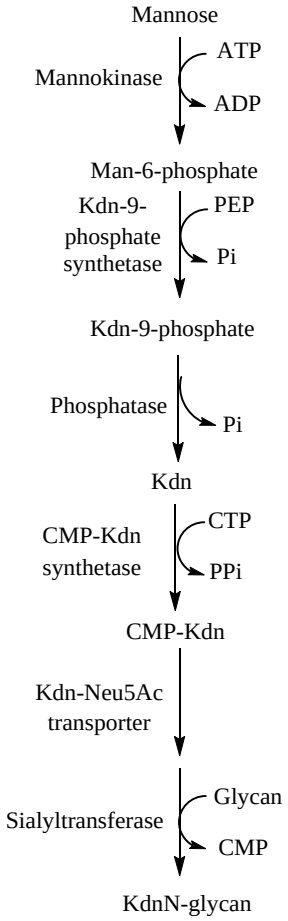


Figure 1.7 Proposed pathway for Kdn biosynthesis.

1.3. Glycoside Hydrolases

Glycoside hydrolases (GHs) or glycosidases (E.C.3.2.1.X) are a diverse group of enzymes that have been categorized, based on their primary protein sequence, into ~132 families or GHs. These enzymes are found ubiquitously in nature and are a subgroup of glycan degrading enzymes. These GH family enzymes catalyze the hydrolysis of -O, -N and -S glycosidic linkages (Figure 1.8) are classified as EC.3.2.1.X in the IUPAC nomenclature system and compiled on the CAZy (Carbohydrate-Active enZymes) website (www.cazy.org). Glycosidases can be classified into several distinct subgroups based on their specific properties and hydrolysis outcome. They are most commonly referred to according to their substrate specificity namely, galactosidases, glucosidases and sialidases as enzymes that catalyze the hydrolysis of galactosides, glucosides and sialosides, respectively. The stereochemistry of their substrate at the anomeric centre allows for further grouping of these enzymes into α- and β-glycosidases, with each group specifically hydrolyzing substrates that possess the preferred anomeric stereochemistry. Of more importance is the stereochemical outcome of the hydrolysis reaction, a distinction which allows for division of these enzymes into retaining and inverting subgroups. In retaining enzymes, as the name implies, the stereochemistry of the substrate and product are the same, while in inverting enzymes, the product has an inverted anomeric stereochemistry compared to that of the substrate.

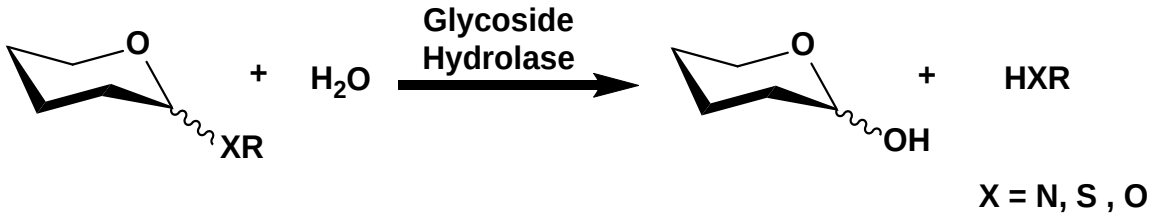


Figure 1.8 General scheme of glycoside hydrolysis leading to the formation of a hemiacetal and the corresponding free aglycone (RXH).

Given the widespread presence of glycosidases in nature and their involvement in a variety of important biological processes, it is not surprising that they have been targeted in many therapeutic efforts for the design of effective treatments for a variety of conditions such as pneumococcal infection [24] and Alzheimer's disease [25].

1.3.1. Catalytic mechanism of glycoside hydrolases

The catalytic mechanism of retaining or inverting glycosidases was first postulated by Koshland [26] The categorization of typical inverting and retaining enzymes is shown below (Figure 1.9).

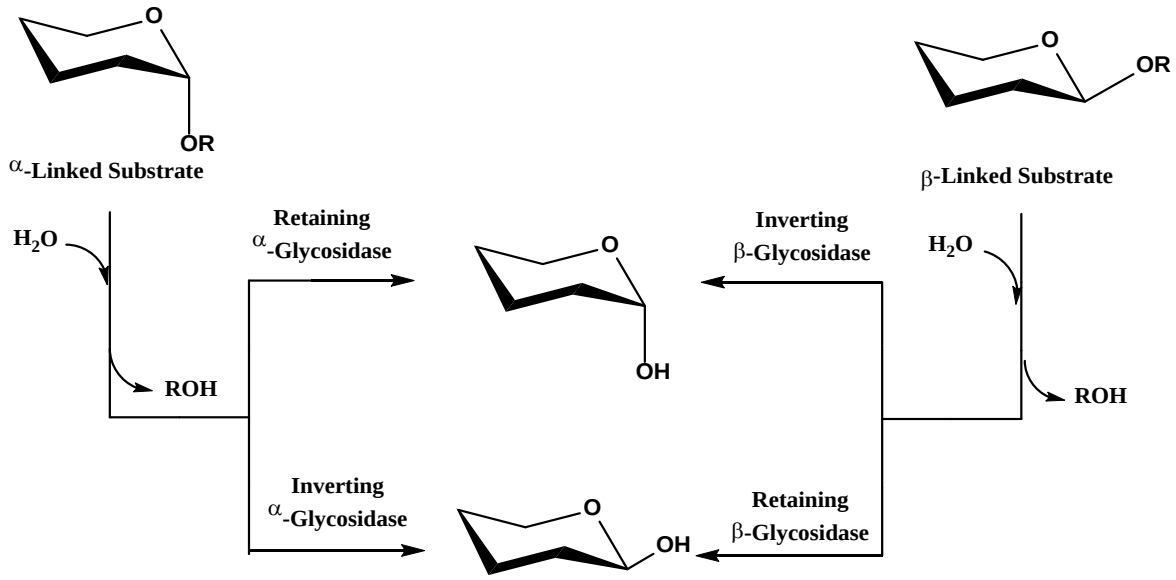


Figure 1.9 Categorization of stereochemical outcome during hydrolysis of glycosidases.

1.3.2. Mechanism of retaining glycoside hydrolases

These enzymes hydrolyze the glycosidic bonds with retention of configuration at the anomeric centre. For most retaining glycosidases, catalysis involves two active site carboxylic acid residues namely aspartate and/or glutamate, one of which acts as a general acid/base residue and the other as a catalytic nucleophile. Moreover, these two amino acid residues are located 5.5-6.5 Å apart [27,28]. This mechanism entails a double displacement or “ping pong” and involves glycosylation and deglycosylation steps that flank a glycosyl-enzyme intermediate. During glycosylation, nucleophilic attack by one of the catalytic carboxylate groups at the anomeric centre leads to the formation of a covalently bound glycosyl-enzyme intermediate while the other carboxylate acts as a general-acid residue to assist the departure of the aglycone by the way of a pyranosylium ion-like transition state. In the second step of the mechanism, the acid/base carboxylate residue now acts as a general-base residue to deprotonate a water molecule as it attacks

the glycosyl-enzyme intermediate to give the final hydrolyzed product [29] This mechanism is depicted in Figure 1.10.

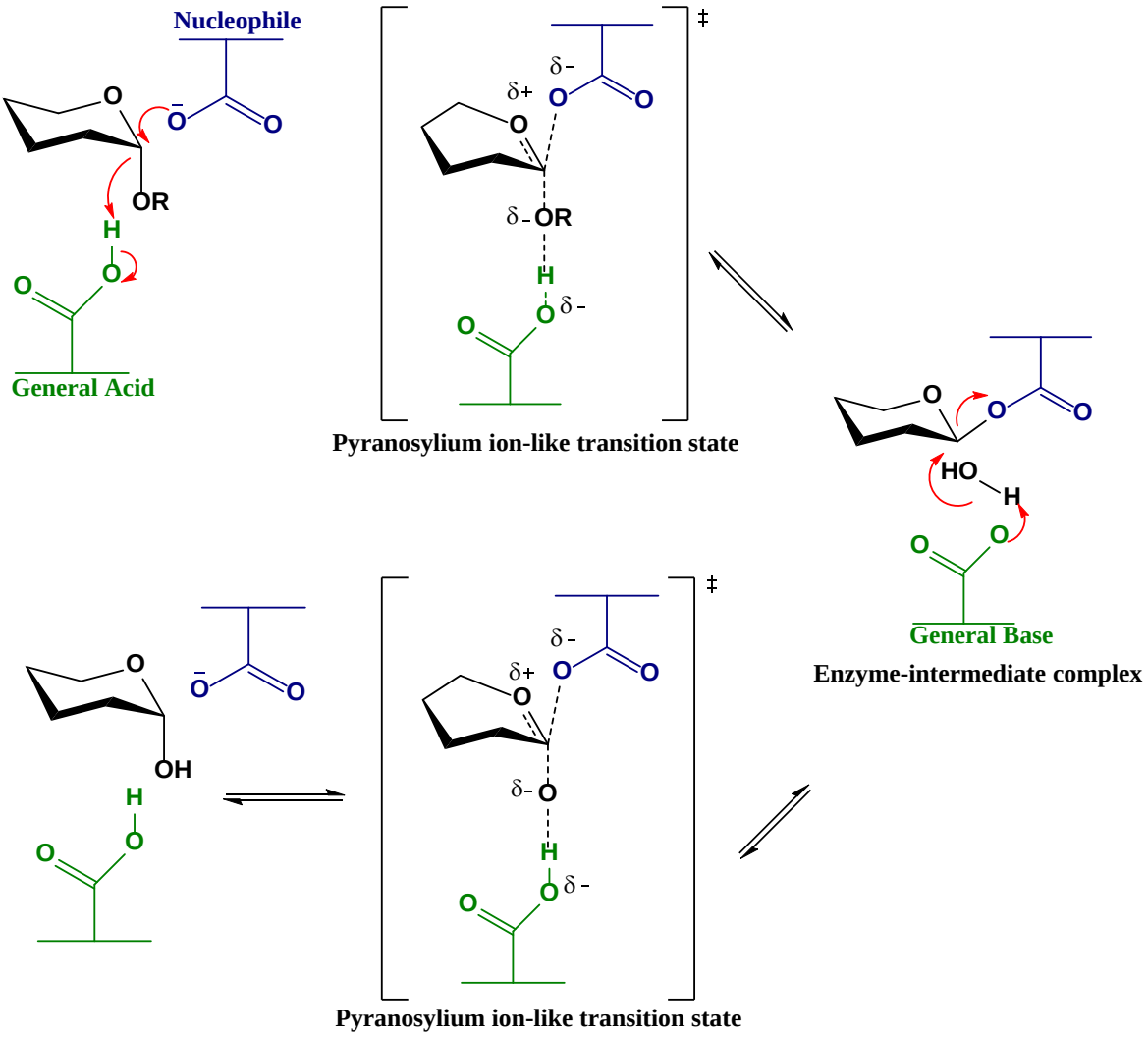


Figure 1.10 Double-displacement mechanism of retaining glycosidases. All non-anomeric substituents are removed for clarity.

1.3.3. Mechanism of inverting glycoside hydrolases

In contrast to retaining glycoside hydrolases, catalysis by inverting enzymes proceeds via a single nucleophilic displacement of the aglycone by a water molecule that occurs directly at the anomeric centre [26]. This single step reaction occurs in a similar fashion to retaining enzymes. That is, the hydrolysis reaction is catalyzed by two carboxylate residues (carboxylic acid side chains of either aspartic acid and/or glutamic

acid) acting in unison as a pair of general catalysts, that promote the reaction via a pyranosylium ion-like transition state [27] (Figure 1.11). In the generally accepted mechanism, the general-base residue deprotonates a water molecule, a process that is simultaneous with nucleophilic attack at the anomeric centre and protonation of the aglycone by the general-acid catalyst. Consequently, the product has an inverted stereocentre at the anomeric centre.

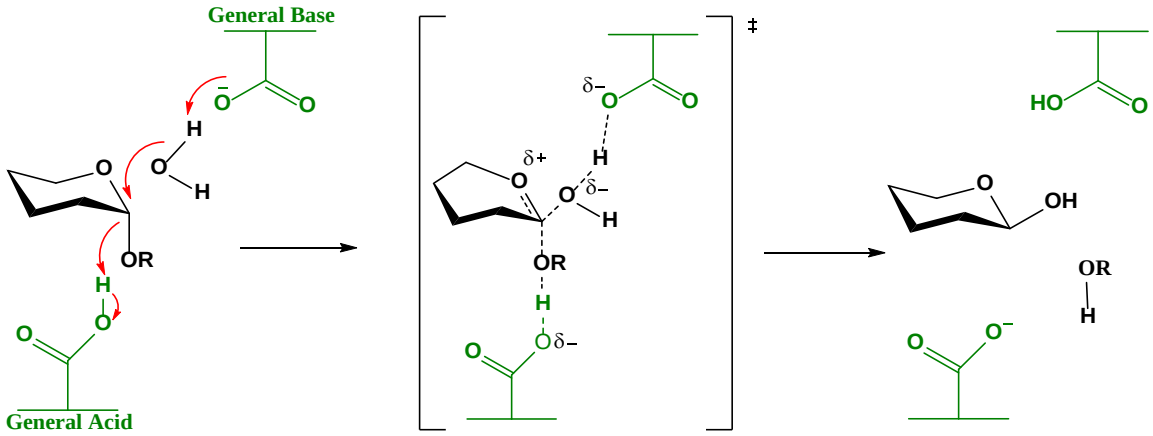


Figure 1.11 Single displacement mechanism of inverting glycosidases. All non-anomeric substituents are removed for clarity.

1.4. Sialidases

Sialoside hydrolases which are also known as neuraminidases or more commonly as sialidases (EC 3.2.1.18), are a subgroup of glycoside hydrolases that catalyze the hydrolysis of sialic acids α-ketosidically linked to glycoproteins, glycolipids or oligosaccharides. The rate of cleavage of specific substrates is dependent on the origin of the sialidase, the penultimate sugar, the type of linkage as well as the nature of the substitutions on the sialic acid itself [30]. These enzymes are categorized into two groups on the CAZY database: 1) *exo*-sialidases (EC 3.2.1.18) which include GH families 33, 34 and 83); and 2) *endo*-sialidases (EC 3.2.1.129) found in GH 58. Notably, the *exo*-sialidases are grouped into a glycoside hydrolase clan (GH-E), which possess a six-fold β-propeller as their defining structural fold [31]. Of interest to us are the *exo*-sialidases that catalyze the hydrolysis of terminally linked sialic acids from sialoglycoconjugates and will be referred to as sialidases throughout this thesis. The general hydrolysis reaction catalyzed by sialidases is depicted in Figure 1.12.

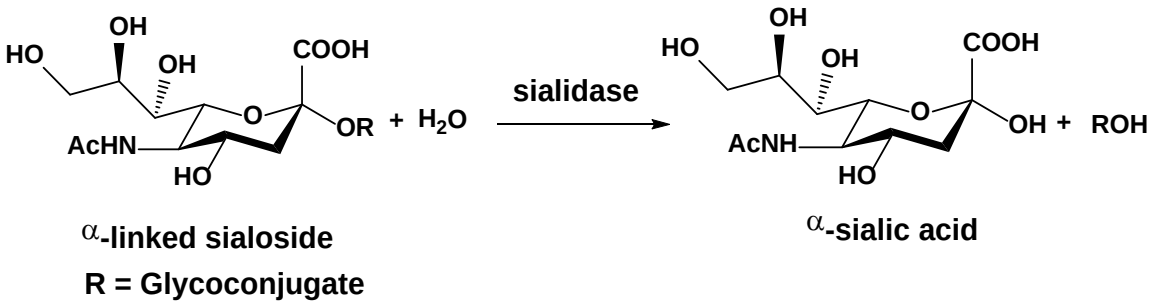


Figure 1.12 Sialidase catalyzed hydrolysis of α-sialosides gives α-sialic acid as the initially formed reaction product.

1.4.1. Sialidase Sequence Homology and Families

Sialidases are found in a wide range of organisms including viruses, bacteria and mammals. The fist x-ray crystal structure of a sialidase, an influenza type A sialidase was published by Varghese, Laver and Colman in 1983 and it is shown in Figure 1.13. This tetrameric enzyme has four identical monomers, each composed of six topologically identical beta-sheets that are arranged in a propeller formation. Soaking of the crystal structure with sialic acid revealed a deep pocket, shown from the top of each monomer, that is the catalytic site [32].

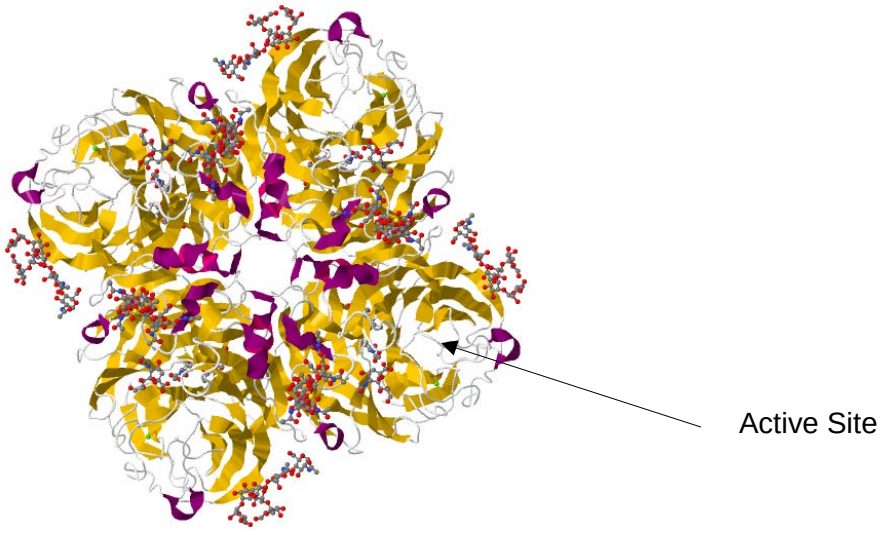


Figure 1.13 Crystal structure of influenza sialidase (PDB: 2BAT)[33] showing its tetrameric structure. Also labelled is an active site pocket.

There are literally thousands of sialidase genes and protein sequences cited in GenBank and Protein Databases, and there have been numerous solved X-ray diffraction experimental structures. These include sialidases, *trans*-sialidases and dual function haemagglutinin-neuraminidases. While these three types possess sialidase activities, there are distinct differences between them. Sialidase activity generally refers to transfer of a sialic acid residue from a glycoconjugate to a water molecule. *Trans*-sialidases, which are often found in parasites such as *Trypanosoma cruzi*, have the ability to transfer the sialic acid moiety from one carbohydrate to another. The haemagglutinin-neuraminidases as the name implies are dual function enzymes with both sialidase and haemagglutinin activities.

Viral sialidases are assigned to glycoside hydrolase family 34 (GH34) and due to the well-known rapid evolution of many viruses, there is only an estimated 40% sequence homology between different influenza strain sialidases, which includes the conserved active site residues and key structural component [34]. This sequence homology is reduced to less than 30% when comparing influenza sialidases to bacterial enzymes [35]. Despite their low sequence homology, all sialidase share several common features and can be grouped into large and small enzymes. Small enzymes have molecular weights in the range of 40-50 kDa and contain only the minimal components for activity while large sialidases, up to 150 kDa contain additional binding domains such as lectin domains and signal sequences [36].

Non-pathogenic bacterial sialidases commonly contain an 8-12 residue sequence called the “Asp box” (Ser/Thr-X-Asp-X-Gly-X-Thr-Trp/Phe) which can be repeated up to 5 times throughout the structure. The function of this motif remains unknown. Another conserved sequence is a RIP motif (Arg-Ile/Leu-Pro) that is located upstream of the Asp-box and contains one of the arginines that interact directly with sialic acid in the active site. Such interactions serve to facilitate binding of the negative charged carboxylate of the sialoside hence promoting hydrolysis [37]. The structures of several sialidases are shown in Figure 1.14.

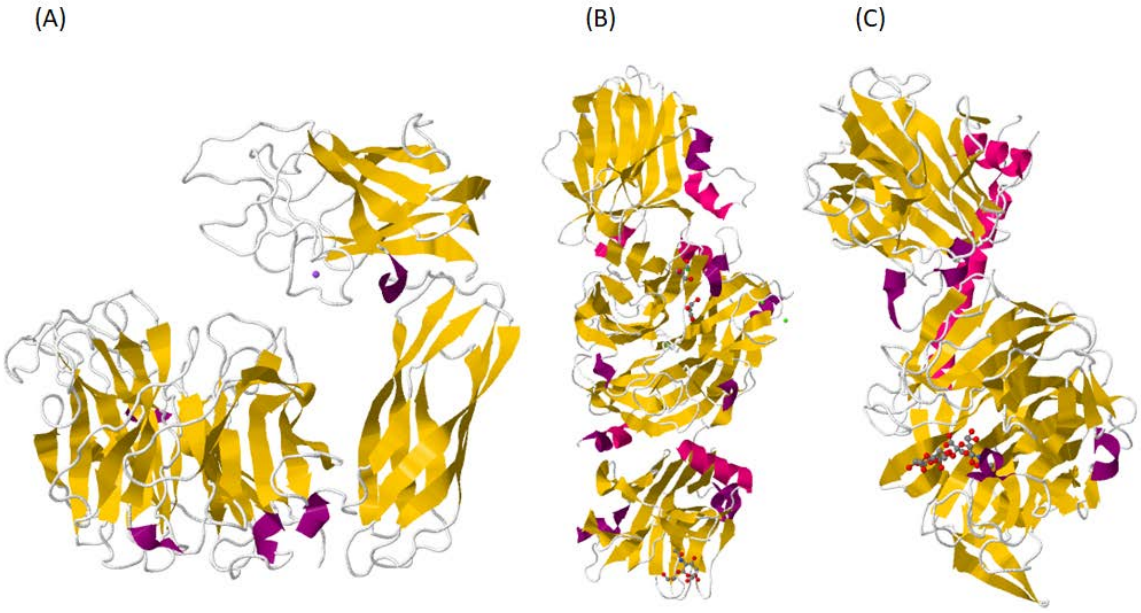


Figure 1.14 Selected protein structures from the sialidase superfamily: (A) *Micromonspora viridifaciens* sialidase (PDB:1EUT)[38], (B) *Vibrio cholerae* sialidase with alpha-2,6-sialyllactose (PDB:1W0P)[39] , (C) *Trypanosoma cruzi* trans-sialidase in complex with sialyl-lactose (PDB:1S0I)[40]

1.4.2. Conserved Active Site Residues and their Catalytic Role in Sialidases

Detailed sequence analysis of bacterial and viral sialidase in addition to X-ray crystallographic studies have allowed the identification of seven conserved active site residues across sialidases of various origin (Figure 1.15).

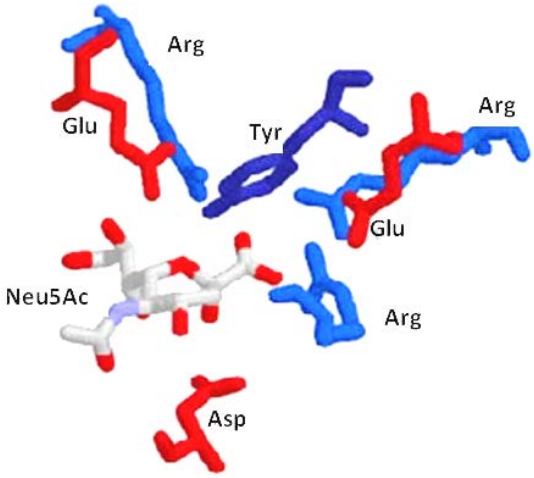


Figure 1.15 Conserved active site residues of sialidases. Seven conserved sialidase active site residues are shown with the product Neu5Ac bound in the active site.

A triad of arginine residues binds the anionic carboxylate of the sialoside substrate via strong electrostatic interactions. Three other residues are involved in catalysis and these are a glutamic acid and tyrosine pair located under the sialic acid ring as well as an aspartic acid positioned above the ring. The seventh conserved residue is another glutamic acid which forms a salt bridge with an arginine in the above-mentioned arginine triad in order to stabilize the triad's position and aid in the hydrogen bonding of the triad with carboxylate group of the sialoside [41].

To date, numerous kinetics and structural studies on various exo-sialidase have shown that all sialidase are retaining glycosidases that give α-sialic acid as product from α-sialoside hydrolysis. However, once the α-sialic acid product is released into solution, it undergoes mutarotation [42] to the more thermodynamically favoured β-configuration so that at equilibrium 95% of the available sialic acid is in the β form [43].

Unlike glycosidases that hydrolyze aldopyranoside linkages, sialidases contain a nucleophilic tyrosine residue in their active site rather than a carboxylate nucleophile, as shown by the Withers' group [44]. This study took advantage of *Trypanosoma cruzi trans*-sialidase (*Tc*TS), a key component in Chagas' disease that while belonging to the sialidase family (GH38) transfers sialic acid residues from glycoconjugates on the host cell to terminal galactose moieties on the cell surface of *Trypanosoma cruzi* in a α-(2,3)

linkage specific manner. Building on their previous work of developing fluorinated sugar analogues as trapping agents [45], they used 2,3-difluorosialic acid as a substrate mimic to trap the 3-fluorosialoyl-enzyme intermediate. ES/MS analysis of the enzyme bound intermediate showed a mass increase of 304±12 Da, which is the expected weight gain in the event of covalent attachment of 3-fluorosialyl to the enzyme. LC-MS analysis of labelled and unlabelled enzyme, identified the active site tyrosine (Tyr342) to be the nucleophilic residue [44]. The enzyme-bound intermediate structure also shows that Asp59 is positioned within hydrogen bonding distance indicating its role as a general-acid catalyst for the formation of the glycosyl-enzyme intermediate and a general-base catalyst for the transfer reaction of the acceptor. The role of the Asp59 as the general acid/base catalyst is further confirmed through chemical rescue studies in which activity is restored to the D59A mutant by azide and a sialyl azide product is formed [46]. Furthermore, confirmation of Tyr as the nucleophile comes from data Bennet and co-workers who have shown that mutation of the conserved Tyr370 nucleophile in *M. viridifaciens* sialidase to an alanine, an aspartate or a glycine residue changes the mechanism from retaining to inverting [47]. This was proposed to be due to a newly created cavity stemming from mutagenesis that accommodates a water molecule and allows it to act as the nucleophile. Watson et al [47] further rationalized that the natural trend of selection of Tyr as a nucleophile in sialidases as opposed to the glutamate/aspartate in glycosidases, is in part due to the intrinsic reactivity differences between sialosides and glycosides.

In other words, the presence of a negatively charged carboxylate group at the anomeric centre of the substrate in addition to the lack of an electron withdrawing group on C3, makes the sialosyl-enzyme bound intermediate more reactive than a corresponding glycosyl-bound enzyme intermediate. This increase in intrinsic reactivity is countered, by covalent interaction of the intermediate with a less reactive leaving group (tyrosine) as opposed to the more reactive leaving groups (glutamate/aspartate). Taken together, this data supports the hypothesis that all sialidases facilitate hydrolysis by formation of a covalent sialosyl-intermediate between the nucleophilic tyrosine and the anomeric centre of the substrate to give hydrolysis with an overall retention of configuration.

1.4.3. Mechanism of Sialidase-Catalyzed Hydrolysis

The reaction proceeds through two chemical steps. In the first step, the aspartate residue acts as a general-acid to facilitate cleavage of the glycosidic bond by protonation of the aglycone oxygen. Simultaneously, the nucleophilic tyrosine is deprotonated by the general-base glutamate as it attacks the anomeric centre. This glycosylation steps leads to the formation of an enzyme bound intermediate by the way of a pyranosylium ion-like transition state. Secondly, the intermediate is then cleaved by a water molecule while being deprotonated by the aspartate residue to give α-sialic acid via another pyranosylium ion-like transition state. These glycosylation and deglycosylation steps are essentially microscopic reverse images of each other. This currently accepted mechanism of hydrolysis by sialidases is depicted in Figure 1.16.

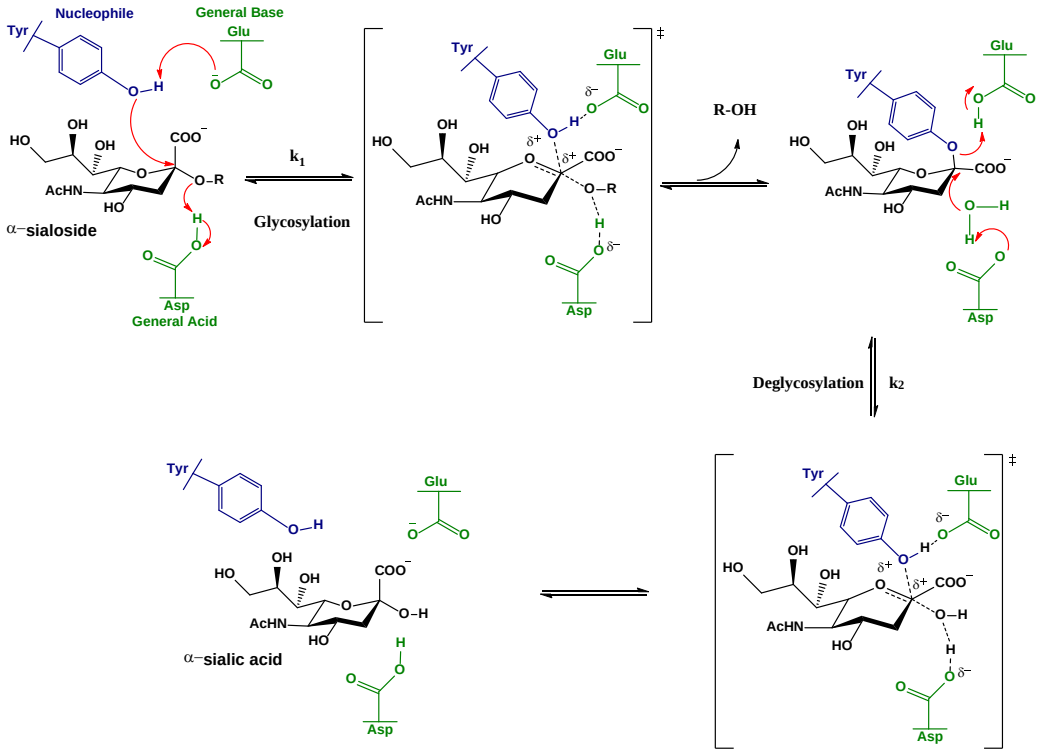


Figure 1.16 The currently accepted mechanism for retaining sialidase catalyzed hydrolysis of α-sialosides.

1.4.4. Substrate Specificity

Depending on the origin and the evolutionary role, sialidases from various species exhibit distinct substrate specificities. The probability and the ultimate rate of the cleavage of sialosides by a specific sialidase is dependent on the type of substrate, the type of linkage to the underlying carbohydrate (2-3, 2-6, 2,8) and the nature of the adjoining glycoconjugate. In general, bacterial sialidases are more promiscuous than viral sialidases in their linkage specificity. Nevertheless, an assay of the rate of hydrolysis of α(2-3) and α(2-6) linked sialoside to galactose residues by sialidases from *Vibrio cholerae, Arthrobacter ureafaciens, Clostridium perfringens*, Newcastle disease virus, fowl plaque virus and influenza A N2 revealed that majority of these sialidases hydrolyzed α(2–3) glycosidic linkages faster than α(2–6) linkages [48].

1.4.5. Biological Significance

Sialidase are found in wide range of organisms in nature including prokaryotes and eukaryotes (GH33) and viruses (GH34, GH83). In general, mammalian sialidases play essential roles in vital cellular process such a proliferation, differentiation, molecular transport and antigen masking [49], while in contrast most bacterial sialidases are indicated in nutritional pathways and disease pathogenesis. For example, the secreted sialidase from the soil bacterium *Micromonospora viridifaciens* cleaves sialic acid from glycoconjugates so that it can be used as carbon source for bacterial growth [38]. Bacterial sialidases can be secreted or cell bound depending on their function. Cell wall bound enzymes due to their locational advantage play an important role in exterior interactions that assist the bacteria in entry into and subsequent infection of the host cells. The secreted forms of sialidase, unlike the cell bound type, can vary in size depending on the availability of a certain sialic acid substrate as well as the final usage of sialic acid [50]. Viral sialidase, such as those expressed by influenza virus play an important role in the infection cycle of the virus. The role of the viral sialidase in transmission of infection is shown to be through cleaving and release of newly replicated and budding virions from the host cell membrane allowing for infection to propagate to neighbouring cells [51].

1.5. Mechanistic Studies of Glycosidases

1.5.1. Enzyme Kinetics

Enzymes by definition are biological catalysts that increase the rate of a reaction while retaining the ability to turn over and not be consumed during the catalytic reaction. The kinetic basis for this phenomenon was outlined by German Biochemist Leonor Michaelis and Canadian Biochemist Maud L. Menten. Their kinetic analysis, aptly termed the Michaelis-Menten equation, was based on previous work by French Chemist Victor Henry and explains the dependence of the initial rate of the enzyme-catalyzed reaction on concentration. Scheme 1.1 shows this relationship in its simplest form.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_{cat}} E + P$$

Scheme 1.1 Michaelis-Menten scheme. (E: enzyme, S: substrate, P: product, ES: enzyme-substrate complex, *k*: rate constants).

The above scheme can be used to derive a rate expression for enzyme-catalyzed reactions (Equation 1.1). To this end a number of assumptions are made, including the initial substrate concentration is much larger than the enzyme concentration and thus its concentration is not perturbed by formation of the Michaelis-Menten complex (ES), also that under initial rate conditions, the ES concentration is constant. This “steady state” assumption gives a M-M expression as shown in Equation 1.1.

$$v = \frac{k_{cat}\left[E\right]_0\left[S\right]}{K_m + \left[S\right]} \qquad \text{Such that,} \quad v_{max} = k_{cat}\left[E\right]_0$$

Equation 1.1 Michaelis-Menten equation under steady state where $[E]_0$ is << [S]

Given the complexity and the number of individual reaction steps that occur during catalysis by sialosidases (Scheme 1.2) Equation 1.1. has to be modified in order to take these additional steps into account [52,53].

$$E + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} E{:}S_C \underset{k_4}{\overset{k_3}{\rightleftharpoons}} E{:}S_B \xrightarrow[\text{glycosylation}]{k_5} E\text{-Inter} \xrightarrow[\text{de-glycosylation}]{k_7} E + P$$

Scheme 1.2 Glycosidase Michaelis-Menten scheme. The $E{:}S_C$ refers to Michaelis complex in which the substrate has a chair conformation while the $E{:}S_B$ complex has a boat confirmation.

The catalytic constant, k_{cat}, or the turnover number for the enzyme, is defined as the number of substrate molecules that are converted to product per unit of time when the enzyme is fully saturated with substrate. This first order constant, can be expressed as an equation that takes into account the possibility of the rate limiting step in the catalysis reaction and is a combination of multiple stages within the catalyzed reaction. Namely, the rate limiting step could be the glycosylation (k_5), deglycosylation (k_7), the conformational change of substrate upon binding to the enzyme (k_3) or a combination of all three. This is expressed in Equation 1.2.

$$k_{cat} = \frac{k_3\ k_5\ k_7}{k_3\ k_5 + k_3\ k_7 + k_4\ k_7 + k_5\ k_7}$$

Equation 1.2 Expression of the k_{cat} or turnover number of glycosidases.

The Michaelis constant (K_m), which is defined as the substrate concentration at which the rate of reaction is half the maximal value, varies from one enzyme to another, and also with different substrates for the same enzyme. The K_m is also referred to as an apparent dissociation constant of the enzyme-substrate complex (ES) and can be derived and expressed as shown in Equation 1.3.

$$K_m = \frac{k_2k_4k_7 + k_2k_5k_7 + k_3k_5k_7}{k_1k_3k_5 + k_1k_3k_7 + k_1k_4k_7 + k_1k_5k_7}$$

Equation 1.3 Expression of K_m or apparent dissociation constant of ES complex.

The catalytic efficiency of an enzyme, which is a second-order constant, is expressed as a ratio of k_{cat}/K_m and has the units of $M^{-1}\ s^{-1}$. All the reversible rates from the

initial introduction of substrate and enzyme until the first irreversible step, which is assumed to be glycosylation, play a role in determination of the value of this rate constant as shown in Equation 1.4.

$$\frac{k_{cat}}{K_m} = \frac{k_1 k_3 k_5}{k_2 k_4 + k_2 k_5 + k_3 k_5}$$

Equation 1.4 Kinetic expression for an enzyme's efficiency (k_{cat}/K_m).

1.5.2. Brønsted Analysis

The mechanisms of enzymatic reactions, including those of glycoside hydrolases, can be investigated by use of a number of different kinetic based analytical methods. A major tool for defining bond making and breaking at the reaction transition state involves determination of linear free energy relationships (LFER), or structure-reactivity correlations. An example of a LFER is a Brønsted analysis, which involves correlating a thermodynamic quantity (pK_a) with a kinetic parameter [$\log_{10}(k)$] to give a sensitivity value (α or β)[54].

Brønsted acid/base catalysis is one of the most common types of catalysis in organic chemistry. For a general catalyzed reaction, a chemical step involving proton transfer is part of the rate-limiting step. Thus, variations in the acid/base concentration and the acidity/basicity of the catalyst have direct effects on the reaction rate. Classically, this correlation is called a Brønsted plot, in which the catalyst is systematically varied and the dependence of reaction rate constant, $\log_{10}(k)$ is plotted against the pK_a of the general acid/base catalyst[54].

Similar to general acid/base catalysis, the same correlation can be extended to reactions where the departure of a leaving group occurs during the rate-limiting step. Plotting the logarithm of the measured values versus the pK_a of the conjugate acid of the various leaving groups should afford a straight line with the slope being $\beta_{lg.}$ For a typical retaining glycoside hydrolase, the aglycone (leaving group) is cleaved during the glycosylation step and as such the nature of the leaving group has a direct effect on the

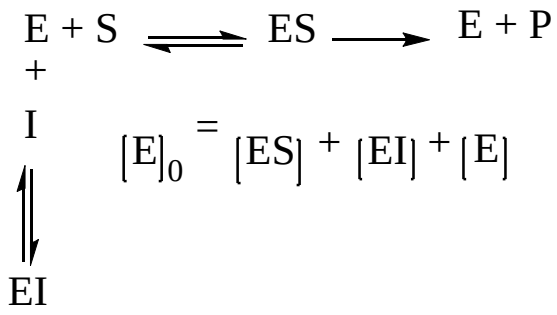
rate constant for this step. To determine if departure is either fully or partially rate-limiting, a series of substrates with varying leaving groups are synthesized and the values of β_{lg} on the two kinetic parameters (k_{cat} and k_{cat}/K_m) can provide valuable insight with regard to the degree of glycosidic bond cleavage at the various enzymatic transition state(s).

1.5.3. Reversible Inhibition

Reversible inhibition, as implied by the name, occurs when the enzyme activity is reduced through a reversible interaction of the inhibitor with the enzyme. This interaction which may not be at the active site of the enzyme, is often achieved through electrostatic interactions including hydrogen and ionic bonds [55]. Reversible inhibitors can be further divided into three major groups: competitive, non-competitive and uncompetitive.

Competitive Inhibitors

In this mode of inhibition, the inhibitor essentially competes with the substrate for the enzyme in a mutually exclusive manner so that any given time, both enzyme-inhibitor (EI) and enzyme-substrate (ES) complexes are present in solution. But no enzyme-inhibitor-substrate (EIS) complex can be formed. The nature of this competition also implies that the degree of inhibition, which is generally thought of as competition for the active site, can be lessened in the presence of saturating levels of substrate, thus reducing inhibitor access to the enzyme active site. The binding of the inhibitor to the enzyme, generates the EI complex, hence reducing the level of free enzyme available for binding to the substrate and formation of the ES complex. Hence the inhibitor increases the apparent K_m of the substrate but has no effect on the maximum rate of the reaction which is dependent on the ES complex. The kinetic scheme for competitive inhibition is depicted in Scheme 1.3.



Scheme 1.3 Schematic representation of competitive inhibition.

The Michaelis-Menten equation for competitive inhibition has to be modified slightly to take into account the dependence of the apparent K_m on the concentration of the inhibitor and its dissociation constant K_i. This is shown in Equation 1.5.

$$v = \frac{v_{max}[S]}{[S] + K_m\left(1 + \frac{[I]}{K_i}\right)}$$

Equation 1.5 Michaelis-Menten equation for competitive inhibition.

Non-competitive Inhibition

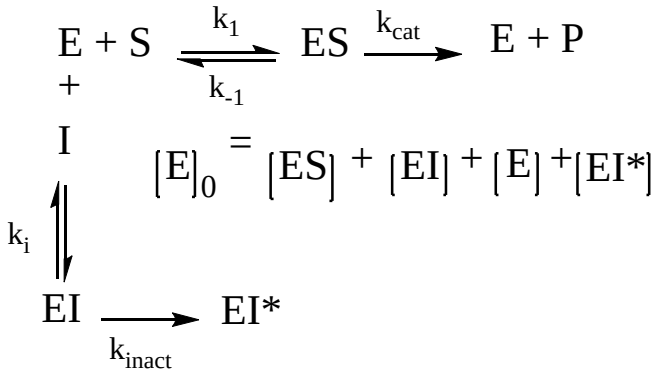
In this type of inhibition, the inhibitor can bind to both the enzyme to form an EI complex, and to ES to give an enzyme-substrate-inhibitor complex (ESI). This change in the binding capabilities of the inhibitor and its interaction with the ES complex, has the overall effect of lowering of the maximum velocity rate (v_{max}) and increasing the apparent dissociation constant (K_m) [55].

Un-competitive Inhibition

In this type of inhibition, the inhibitor only binds to the enzyme-substrate (ES) complex and this results in a lowering of the v_{max} while the K_m remains unchanged.

1.5.4. Irreversible Inhibition

Irreversible inhibition commonly occurs when a covalent bond is formed between the inhibitor and enzyme, often within the active site, where loss of enzyme activity is time dependent. A general scheme for irreversible inhibition is shown in Scheme 1.4.

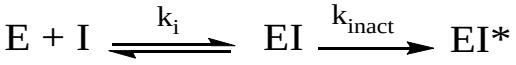


Scheme 1.4 General representation of irreversible inhibition.

In these cases, following binding of the inhibitor to the enzyme and formation of the EI complex, the inhibitor can then react, with an associated a first order rate constant of k_{inact} (maximal rate of enzyme inactivation), to give the covalently modified EI* complex. Irreversible inhibition can, based on their mechanism of action, be conceptually divided into two groups that are affinity labeling agents and mechanism-based inactivators.

Affinity Labeling Agents

In this class of inhibitors, as depicted in Scheme 1.5, an inhibitor that contains an intrinsically reactive group forms a complex with the enzyme and the reactive functional group labels the enzyme without the enzyme providing transient state stabilization of the labelling process. Affinity labels are prominently used to tag enzymes. However, due their lack of selectivity, these groups of inhibitors are poor choice in terms of therapeutic approaches to enzyme inactivation.



Scheme 1.5 A representation of affinity labeling inhibition.

Mechanism-based Inactivation

Mechanism-based inactivators often referred to as “suicide inhibitors”, are inherently inactive compounds that upon binding to the enzyme are activated by the enzyme’s catalytic machinery by either i) irreversibly modify a catalytic group or ii) modify an essential cofactor required for enzymatic activity. The consequence of either route is enzymatic “death”.

In order for a compound to be classified as a mechanism-based inactivator there are a number of additional key criteria that need to be met, namely: i) time-dependent inactivation of enzyme, ii) saturation kinetics that follow the first order kinetics, iii) 1:1 stoichiometry of inhibitor and enzyme, and iv) enzyme must catalyze its own inactivation.

1.5.5. Glycosidase Inhibitors

The design of potent and specific glycosidase inhibitors is a robust and well established field due to the biological importance of these enzymes and the immense potential of inhibitors for gaining mechanistic information and as therapeutic agents for diseases such as diabetes[56], tumour metastasis[57] and viral infections[58]. There has been great effort in the design and synthesis of both potent[59] and selective inhibitors[60]. The enzymatic proficiency as measured by the rate enhancement of a glycosidase-catalyzed reaction compared to the spontaneous hydrolysis reaction[61], has targeted the search for potent inhibitors toward understanding the details of enzymatic transition state (TS) structures. As shown in Figure 1.10 and 1.11, the majority of glycosidases catalysis occurs through stabilizing the enzymatic transition state by a nucleophilic substitution at the anomeric centre involving positive charge development on the anomeric carbon, which then undergoes delocalization onto the endocyclic oxygen of the sugar ring. To this end, tight binding reversible inhibitors that mimic the transition state charge and ring conformation are often called transition state analogues. Most designs of TS analogue inhibitors include a combination of charge and geometrical distorting to try to mimic the presumed transition state structures. In other words, potent inhibitors often possess a positive charge (at physiological pH), and/or a distorted glycosyl ring in order to try to resemble the pyranosylium ion-like TS conformation.

1.6. Directed Evolution

Natural evolution, which involves changes in DNA and the subsequent selection of improved fitness, has resulted in a huge diversity of proteins, both in structure and activity. These natural polypeptides can potentially be harnessed for numerous applications in biotechnology and pharmaceutical sciences. However, various specific applications and uses of protein catalysts require that the natural protein be tailored by protein engineering methods. In theory, the intrinsic properties of a protein can be altered by rational design, however, in reality this approach is hindered by the complexity of the relationship between protein structure and function. Recent advances in molecular biology techniques coupled with the ability to isolate proteins with desired catalytic properties have shown promising results. That is, the laboratory approach to the natural Darwinian evolution of biological macromolecules, is referred to as directed evolution. The aim of such methods is to generate a protein with a new activity by either screening or selecting for the desired function from a large pool of protein variants. Directed evolution is, in theory, a powerful means to engineer proteins to possess various desired properties under well-defined conditions and in a practical time frame. As such directed evolution of proteins has emerged as one of the most effective and reliable tools to enhance the stability, activity, and substrate selectivity of enzymes that ultimately can lead to the production of useful new proteins for a given application[62]. A few examples include: (i) the enhancement of thermostability for endoglucanase *EngB*, one component of the cellulase complex (cellulosome) from *Clostridium cellulovorans*[63], (ii) the improved practical usefulness of firefly luciferase by increasing its catalytic efficiency[64], and (iii) the generation of the *Discosoma* red fluorescent protein (DsRed) that enables live cell imaging[65,66] Of particular interest to us in terms of our future goals, is the directed evolution of α-l-fucosidase from *Thermotoga maritima* (TmαFuc) into an efficient transfucosidase[67]. The same group has also been successful in development of semi-rational approaches to design an α-l-transfucosidase starting with the α-l-fucosidase from commensal bacteria *Bifidobacterium longum*[68].

John Maynard Smith, in his 1970 paper, fittingly described protein evolution as a walk from one functional protein to another in the space of all possible protein sequences [69]. Each sequence in protein space can be assigned a ‘fitness’ value that while in natural

evolution would be a measure of the survival rate of the host[70], in directed evolution it is defined as the ability of the protein to satisfy the criteria set by the experimenter and imposed through the chosen assay used for screening. A plot of fitness against sequence visually depicts the landscape for evolution. The activity level of each protein variant can thus be visualized as a measure of the elevation of the fitness value on an x–y coordinate, where regions of higher elevation represent desirable function. The goal of directed evolution studies is to take mutational steps within this landscape that climb towards peak activity levels through accumulation of beneficial mutations over many generations (Figure 1.17).

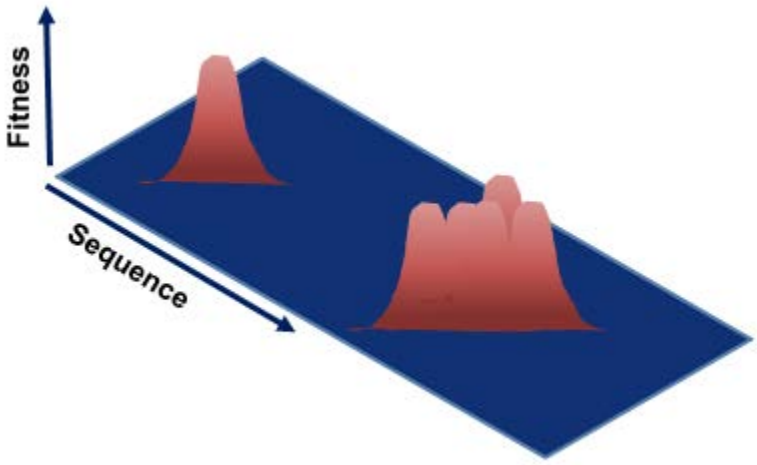


Figure 1.17 Protein fitness landscape.

Much of the appeal of a directed evolution approach for engineering proteins lies in the fact that the genetic/coding information is straightforward to amplify and to manipulate, while the functional molecule, the translated protein, has a rich chemistry that provides a wide range of possible activities. Additionally, as the structural and mechanistic information of the resultant protein is not necessary, the process is fundamentally different from the classical rational design based on site-specific mutagenesis[71]. Essentially like natural evolution, directed evolution in the laboratory comprises of the following steps: (i) the generation of a library of mutated genes, (ii) functional expression of the library to generate the proteins for which the library codes, and (iii) design of a sensitive assay to identify individuals showing the desired properties, either by selection or by screening.

After each round, the genes of improved variants are deciphered and serve as templates for the next round of property optimization. Subsequently, multiple iterations of this three-step process lead to the identification of proteins with desired catalytic, biophysical, and molecular recognition properties (Figure 1.18).

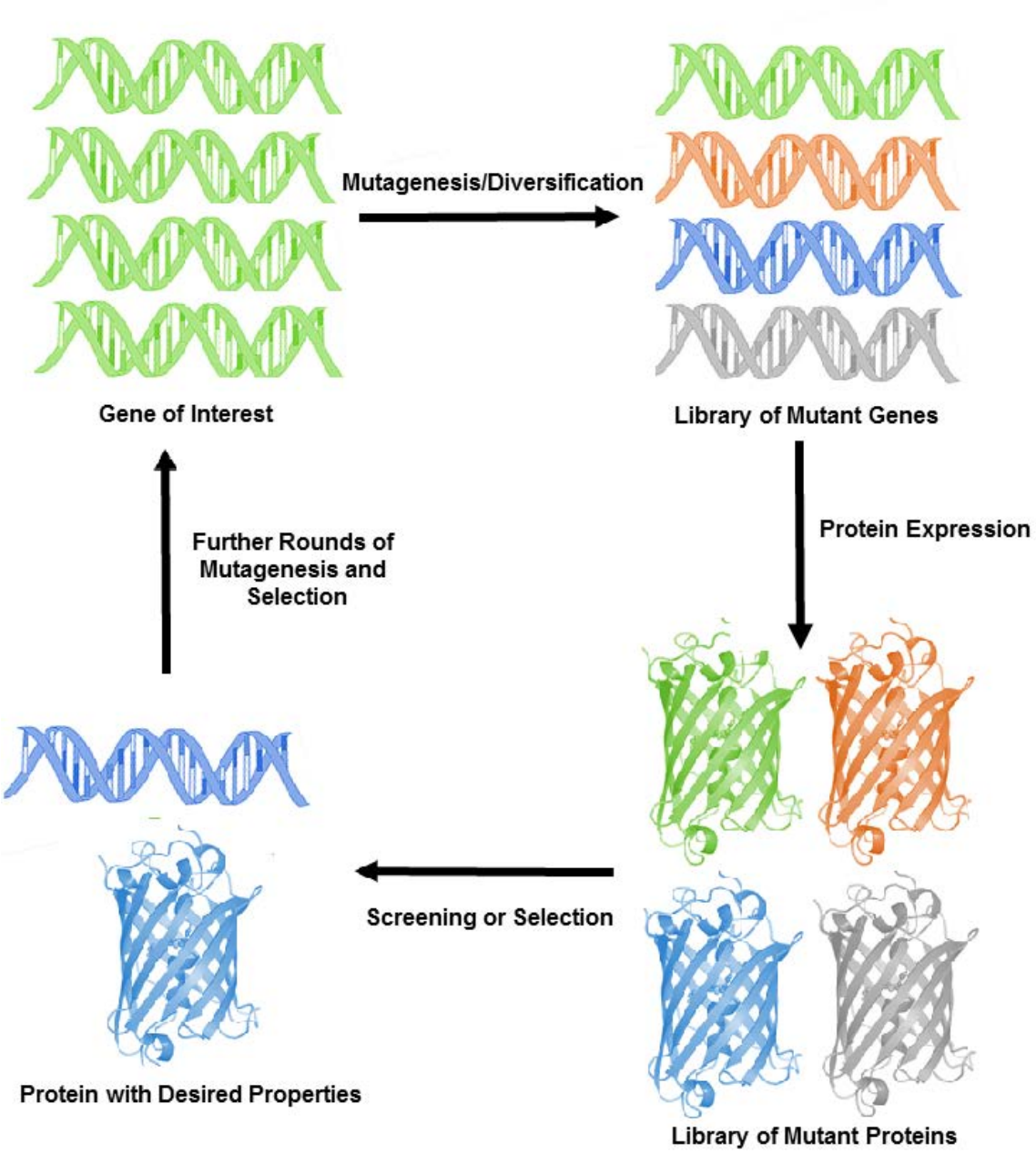


Figure 1.18 General steps of directed enzyme evolution.

There are multiple experimental protocols that can be chosen as the basis for each round of the directed evolution process.

1.6.1. Mutagenesis Library Generation

Screening is often the most difficult and time consuming experimental step, which depends on the screening capabilities available, thus the library is frequently generated to try to enhance the probability of finding improved proteins. A number of methods have been developed for library generation. These can be divided into three broad categories: (i) random mutagenesis; (ii) directed or focused mutagenesis, and (iii) diversification by recombination.

Random Mutagenesis

Random mutagenesis at the nucleotide level is a prevalent strategy which targets whole genes with changes being made, at random, along the whole gene. Random mutagenesis can be further divided into chemically and non-chemically induced mutagenesis depending on the choice of the mutagenesis agent.

The chemical approach involves treatment of the DNA sequence with various chemical and physical agents to damage the DNA in a random fashion. These agents include alkylating compounds such as ethyl methanesulfonate (EMS)[72], deaminating compounds such as nitrous acid[73] and ultraviolet irradiation[74]. While chemical mutagenesis introduces random mutations, due to the intrinsic reactivity of the reagents, it also deactivates genes at a high frequency and as such is less commonly used in directed evolution studies[73].

Non-chemical methods that randomly introduce mutations into the DNA strand generally involve enhancement of the natural rate of errors introduced by DNA polymerase during DNA replication. In a technique labelled error-prone PCR (epPCR)[75] the low fidelity of DNA polymerases, under certain conditions, generates point mutations along the gene of interest during PCR amplification. Certain conditions such as higher magnesium concentrations, increasing the number of cycles of amplification[76] or supplementing the

reaction with manganese can reduce the base-pairing fidelity of the DNA polymerase and increase mutation rates to 10^{-4}~10^{-3} per replicated base[77].

Directed or Focused Mutagenesis

With the many recent advances made in the field of protein crystallography, many proteins have been structurally characterized at a sufficiently high resolution to allow for the identification of residues that are crucial to substrate binding and enzymatic catalysis. Though versatile and simple, all known epPCR methods are significantly limited in their ability to create a large diversity at the gene level. While random mutagenesis can be used to introduce mutations along the entire gene, directed mutagenesis can be utilized to target one or more neighboring codons at the same time. Directed mutagenesis can divided into site-directed mutagenesis (SDM) and site-saturation mutagenesis (SSM). SDM is commonly applied to convert a targeted amino acid to a different chosen amino acid in order to study the function of a single amino acid in relation to the rest of the protein. SSM is a similar method with an additional dimension of complexity, in which a single amino acid can be substituted to become any of the 19 other possible substituents[78]. These experiments are performed by PCR amplification using degenerate synthetic oligonucleotides where the targeted codons are replaced with NNN or NNK where N represents any nucleotide and K represents G or T[79]. The simultaneous saturation mutagenesis of multiple residues can access combinations of mutations that may exhibit epistatic interactions.

Of particular interest to us is the library generation procedure employed by Vogel et. al [80] to modify and study the scope of substrate specificity by a lipase from *Pseudomonas aeruginosa.* In short, they generated focused mutagenesis libraries of mutant enzymes through randomization at several sets of two spatially close amino acids around the active site. The choice of amino acids targeted was aided by the available crystal structure of the enzyme bound by a substrate. The spatially close proximity of the amino acids would potentially allow for occurrence of synergistic conformational effects arising from side-chain orientations that due to its unpredictable nature can not be brought about by single-site saturation mutagenesis. Complete randomization at selected amino acids was then performed in a process, coined as combinatorial active-site saturation test (CAST). This elegantly simple procedure allows for systematic creation of relatively small

libraries of mutants which could be screened by the currently available screening assays leading to positive hits and consequently to the identification of the critical amino acids around the active site that are crucial for substrate selectivity.

Recombination Mutagenesis

While random and directed mutagenesis methods can result in the generation of sequence diversity either along the length of the gene or at specific codons, recombination techniques mimicking the natural evolution path of combining beneficial mutations, are also valuable tools in the generation of a diversified library. Recombination techniques such as DNA shuffling (a procedure that randomly recombines point mutations in vitro)[81,82] and the staggered extension process[83] do not directly create new sequence diversity, but rather take portions of existing sequences and combine them in new arrangements and this results in enhanced diversity. A pioneering paper by Zhang et. al[84] utilized the technique of DNA shuffling to evolve an efficient beta-fucosidase from *Escherichia coli* lacZ beta-galactosidase. The mutated enzyme, selected following seven rounds of DNA shuffling and screening procedures displayed a tenfold increase in k_{cat}/K_m in reactions with the novel para-nitrophenyl-beta-d-fucopyranoside substrate while also showing a 39-fold decrease in reactivity with the natural substrate analogue para-nitrophenyl-beta-d-galactopyranoside. This work was of particular importance as it showed that the library size necessary for obtaining significant enhancements in specificity and activity by reiterative DNA shuffling and screening was within the range of existing high-throughput technology at the time. This work was further enhanced upon by Parikh et. al[85] that showed a significantly more active enzyme could be evolved by a single round of saturation site-directed mutagenesis as opposed to the seven rounds of DNA shuffling. This work demonstrated that saturation site-directed mutagenesis provides a faster and more effective route for this particular evolutionary pathway.

1.6.2. Assaying the Mutagenesis Library

For a successful directed evolution experiment the generation of a large library is non-mandatory, but rather being able to recover the targeted progeny is mandatory. The development of new and improved methods for library analysis is therefore as important as new approaches for creating library diversity. After generating a mutagenesis library,

the major task is to develop a suitable assay for identifying the required protein function. Locating interesting variants from diversified libraries can be accomplished either by an active search of all variants in the library individually (screening) or by applying conditions favoring the exclusive survival of individuals containing the desired variants (selection). Essentially, the difference between screening and selection is that screening is performed on individual genes and as such requires a spatial organization of the screened variants on selected screening media such as agar plates, microtiter plates, arrays, or chips, whereas selections act simultaneously on the entire pool of genes and favour the exclusive growth and survival of clones that contain the intended modified activity.

The basis of all screening and selection methodologies is the link between the gene, the enzyme that it encodes and the activity of the protein. In order to investigate the function of the modified proteins, there must be a link between the mutant DNA and its corresponding protein, in order words, each protein must be linked to the unique DNA sequence that encoded for its production. In addition, there are several criteria that should be satisfied before a screen or selection method is chosen. First, ideally the method should be designed to measure directly the property of interest, such as a modified enzymatic activity. To this end the substrate used in the screen or selection should be identical, or as similar as is possible to the targeted substrate. Also, detection of the reaction product(s) should be sensitive and ideally analyzed under multiple turnover conditions so as to ensure the selection of the best candidates for further property enhancements. Second, the assay should be sensitive over the desired dynamic range. The first rounds of directed evolution experiments require isolation of all variants with even marginal improvements. This would allow for application of more stringent conditions to ensure the isolation of the best variants[86,87].

Screening Methods

Genetic screens have been used by scientist for years to investigate the genes associated with a specific phenotype. The same approach can be utilized in directed evolution where the individual members of the diversified library can be investigated for altered activity or phenotypic changes. There are several well documented screening strategies. Below is a brief description of a selected few of these screens.

Agar/Microtiter Plate Screening

Plate screening is by far the simplest format for screening and involves the incubation of individual colonies from the mutagenic library with the enzyme substrate that has a visualization tag attached to it [88]. Individual members of the library are grown on either agar plates or in liquid media and the activity of the resultant protein is assessed. The crucial factor in such screens is the conversion of substrate by the modified enzymes to create a visual signal, such as chemiluminescence, fluorescence or colour that allows for the identification of colonies with the preferred enzymatic activity being enhanced [89]. The draw back to these methods is the number of colonies that can be assayed as well as possible limitations in the sensitivity of the assay. On agar plates, soluble products diffuse away from the colony and hence only very active variants are detected. In liquid or microtiter plate screening without the availability of sophisticated robotics, only10^3–10^4 variants are typically screened[88].

Screening by Flow Cytometry.

In contrast to the above mentioned screening methods where spatial separation of individual mutants is required to preserves the linkage between phenotype and genotype, a large library population can be interrogated at the level of individual cells using the cell wall or membrane to maintain genotype–phenotype association. Fluorescence-activated cell sorting (FACS)[90] takes advantage of specially designed non-diffusing fluorescent reporters to automate the identification process so as to identify and isolate cells containing the desired genetic variants. Recent advances in flow cytometry has led to improvements in this method so that now it has one of the highest capacities of any screening methods.

Selection Methods

Selection techniques in general rely on a direct a relationship between cell growth/survival and enzymatic activity. Selection methods can be adapted to be used in vivo if a substantial growth advantage is conferred to those clones that harbour a protein variant with the desired improvement. This is generally achieved by genetic complementation of hosts that are deficient in a certain pathway or activity. Cells from the diversified library are generally plated or grown on a selection media that contains a

substrate that after reaction by cells that possess an enhanced enzymatic activity gives those cells a growth advantage. This protocol is often used in conjunction with a “suicide” substrate, which is used as a media supplement, where reaction of such a compound by mutant enzymes possessing an “undesired” activity leads to cell death[89]. An example of this selection has been reported where lipase catalyzed hydrolysis of the ‘‘wrong’’ enantiomer yields a poison (fluoroacetic acid), whereas hydrolysis of the desirable enantiomer yields a carbon source (acetate)[91].

In comparison to screening methods, selective enrichment and selection of only those clones that express the particular enzyme function can be very efficient as many more library members can be analyzed simultaneously. This is because uninteresting variants are not observed. As a consequence, surveying libraries is much faster and can be carried out with higher throughput. Additionally, most selection techniques are compatible with large libraries and do not require special instrumentation[92].

1.7. Research Objectives

The focus of the work described in chapter 2 was to generate mutagenic libraries and through various screening and selection protocols, identify and characterize enzymes capable of capable of hydrolyzing Kdn glycosides more efficiently than Neu5Ac substrates. Kinetic studies on these clones would allow for determination of enzyme efficiencies and specificities.

For Chapters 3, 4, the work is focused on study of covalent inhibition of α-glucosidase from *Saccharomyces cerevisiae* (GH13). The measured pH-rate profiles for inhibition and reactivation as well as the corresponding catalytic and inhibitory proficiencies suggested that inhibition results from the formation of carbenium ions in the active site that are trapped rapidly by an enzymatic residue.

1.8. Contributions of the author to the work presented in this thesis

Chapter 2: Directed Evolution of the Sialidase from Micromonospora viridifaciens into a Kdnase

All the experimental work was done by the candidate except for the following:

- The initial error prone PCR was performed by Dr. Jacqueline Watson
- Dr. Fahimeh Shidmoossavee synthesized 8FMU α-Kdn-(2→6)-β-D-Galp and 8FMU α-Neu5Ac-(2→6)-β-D-Galp.
- Matthew Deen synthesized *O*-β-D-galactopyranosyl chloramphenicol and *O*-β-D-galactopyranosyl-L-tyrosine.

Chapter 3: A New Class of Glycoside Hydrolase Mechanism-Based Covalent Inhibitors: Glycosylation Transition State Conformations

This chapter is a reproduction of a paper published in the Journal of the American Chemical Society.

For this contribution, I performed all protein preparations and kinetics studies. The acquisition of mass spectral data and Mass Spectrometry experiments were done with the help of Ms. Kyung-Mee Moon, Mr. Jason Rogalski, and Dr. Leonard Foster from the UBC proteomics facility. The synthesis of chemical compounds were completed by Michael Tran, Dr. Anuj Yadav and Dr. Pal John Pal Adabala. Dr. Andrew Bennet supervised the project and was involved in the writing and editing of the paper.

Chapter 4: Characterization of Reaction Coordinates for Covalent Labelling of Glycosidase by Cyclohexene Inhibitors

This chapter is a draft form of a paper that is under preparation for submission to the Journal of Angewandte Chemie International.

For this contribution, I performed all site directed mutant generation, expression, protein preparations and kinetics studies. Dr. Weiwu Ren performed all the synthesis. Crystal structures were generated in Dr. Gloster's laboratory.

Dr. Andrew Bennet supervised the project and was involved in the writing and editing of this draft.

Chapter 5: A Mechanistic Study on the a-N-acetylgalactosaminidase from E. meningosepticum: A Family 109 Glycoside Hydrolase

This chapter is a reproduction of a paper published in the journal MedChemComm.

For this contribution, I repeated the cloning, expression and purification of the enzyme (initially performed by Dr. Saswati Chakladar) as well as several of the kinetic studies with compounds made by Dr. Saswati Chakladar. My involvement was necessary to confirm the protein sequence and activity as Dr. Saswati had cloned a different strain to that reported in the literature. Dr. Andrew Bennet supervised the project and was involved in the writing and editing of the paper.

1.9. References

1. Varki A, Cummings RD, Esko JD. *Essentials of glycobiology.* 2nd edition ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009.

2. Klenk E. Neuraminsäure, das spaltprodukt eines neuen gehirnlipoids. *physiologische Chemie*. 1941;268(1-2):50-58.

3. Blix G. Concerning the carbohydrate groups of submaxillary mucin. *Biol. Chem. Hoppe-Seyler*. 1936;240:43-54.

4. Varki A, Schnaar R, Schauer R. Sialic acids and other nonulosonic acids. In: Varki A, Cummings R, Esko J, eds. *Essentials of glycobiology.* 3rd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2017.

5. Schauer, R., & Kamerling, J. P. Chapter 11 chemistry, biochemistry and biology of sialic acids. In: *New comprehensive biochemistry* ; 1997:243-402.

6. Angata T, Varki A. Chemical diversity in the sialic acids and related alpha-keto acids: An evolutionary perspective. *Chem Rev*. 2002;102(2):439-469. doi: cr000407m [pii].

7. Olofsson S, Kumlin U, Dimock K, Arnberg N. Avian influenza and sialic acid receptors: More than meets the eye? *Lancet Infect Dis*. 2005;5(3):184-188. doi: S1473309905013113 [pii].

8. Hinderlich S, Weidemann W, Yardeni T, Horstkorte R, Huizing M. UDP-GlcNAc 2-epimerase/ManNAc kinase (GNE), a master regulator of sialic acid synthesis. *Topics in Current Chemistry*. 2015(366):97–137.

9. Daines DA, Wright LF, Chaffin DO, Rubens CE, Silver RP. NeuD plays a role in the synthesis of sialic acid in escherichia coli K1. *FEMS Microbiol Lett*. 2000;189(2):281-284. doi: S0378-1097(00)00301-3 [pii].

10. Edwards U, Frosch M. Sequence and functional analysis of the cloned *Neisseria meningitidis* CMP-NeuNAc synthetase. *FEMS Microbiol Lett*. 1992;75(2-3):161-166.

11. Eckhardt M, Muhlenhoff M, Bethe A, Gerardy-Schahn R. Expression cloning of the golgi CMP-sialic acid transporter. *Proc Natl Acad Sci U S A*. 1996;93(15):7572-7576.

12. Tsuji S. Molecular cloning and functional analysis of sialyltransferases. *J Biochem*. 1996;120(1):1-13.

13. Wieser RJ, Baumann CE, Oesch F. Cell-contact mediated modulation of the sialylation of contactinhibin. *Glycoconj J*. 1995;12(5):672-679.

14. Yogeeswaran G, Salk PL. Metastatic potential is positively correlated with cell surface sialylation of cultured murine tumor cell lines. *Science*. 1981;212(4502):1514-1516.

15. Kelm S, Schauer R. Sialic acids in molecular and cellular interactions. *Int Rev Cytol*. 1997;175:137-240.

16. Paulson JC, Rogers GN, Carroll SM, et al. Selection of influenza virus variants based on sialyloligosaccharide receptor specificity . *Pure Appl. Chem.* 1984;56(7):797-805.

17. Schauer R. Sialic acids and their role as biological masks. *Trends Biochem. Sci.* 1985;10:357–360.

18. Nadano D, Iwasaki M, Endo S, Kitajima K, Inoue S, Inoue Y. A naturally occurring deaminated neuraminic acid, 3-deoxy-D-glycero-D-galacto-nonulosonic acid (KDN). its unique occurrence at the nonreducing ends of oligosialyl chains in polysialoglycoprotein of rainbow trout eggs. *J Biol Chem*. 1986;261(25):11550-11557.

19. Inoue S, Kitajima K. KDN (deaminated neuraminic acid): Dreamful past and exciting future of the newest member of the sialic acid family. *Glycoconj J*. 2006;23(5-6):277-290. doi: 10.1007/s10719-006-6484-y [doi].

20. Li YT, Yuziuk JA, Li SC, et al. A novel sialidase capable of cleaving 3-deoxy-D-glycero-D-galacto-2-nonulosonic acid (KDN). *Arch Biochem Biophys*. 1994;310(1):243-246. doi: S0003-9861(84)71163-5 [pii].

21. Song Y, Kitajima K, Inoue S, Inoue Y. Isolation and structural elucidation of a novel type of ganglioside, deaminated neuraminic acid (KDN)-containing glycosphingolipid, from rainbow trout sperm. the first example of the natural occurrence of KDN-ganglioside, (KDN)GM3. *J Biol Chem*. 1991;266(32):21929-21935.

22. Inoue S, Lin SL, Chang T, et al. Identification of free deaminated sialic acid (2-keto-3-deoxy-D-glycero-D-galacto-nononic acid) in human red blood cells and its elevated expression in fetal cord red blood cells and ovarian cancer cells. *J Biol Chem*. 1998;273(42):27199-27204.

23. Inoue S, Kitajima K, Inoue Y. Identification of 2-keto-3-deoxy-D-glycero--galactononononic acid (KDN, deaminoneuraminic acid) residues in mammalian tissues and human lung carcinoma cells. chemical evidence of the occurrence of KDN glycoconjugates in mammals. *J Biol Chem*. 1996;271(40):24341-24344.

24. Abbott DW, Macauley MS, Vocadlo DJ, Boraston AB. *Streptococcus pneumoniae* endohexosaminidase D, structural and mechanistic insight into substrate-assisted catalysis in family 85 glycoside hydrolases. *J Biol Chem*. 2009;284(17):11676-11689. doi: 10.1074/jbc.M809663200 [doi].

25. Yuzwa SA, Macauley MS, Heinonen JE, et al. A potent mechanism-inspired O-GlcNAcase inhibitor that blocks phosphorylation of tau in vivo. *Nat Chem Biol*. 2008;4(8):483-490. doi: 10.1038/nchembio.96 [doi].

26. KOSHLAND DE. Stereochemistry and the mechanism of enzymatic reactions. *Biological Reviews*. 1953;28(4):416-436. doi: 10.1111/j.1469-185X.1953.tb01386.x.

27. Zechel DL, Withers SG. Glycosidase mechanisms: Anatomy of a finely tuned catalyst. *Acc Chem Res*. 2000;33(1):11-18. doi: ar970172+ [pii].

28. Mhlongo NN, Skelton AA, Kruger G, Soliman ME, Williams IH. A critical survey of average distances between catalytic carboxyl groups in glycoside hydrolases. *Proteins*. 2014;82(9):1747-1755. doi: 10.1002/prot.24528 [doi].

29. Vocadlo DJ, Davies GJ. Mechanistic insights into glycosidase chemistry. *Curr Opin Chem Biol*. 2008;12(5):539-555. doi: 10.1016/j.cbpa.2008.05.010 [doi].

30. Roggentin P, Schauer R, Hoyer LL, Vimr ER. The sialidase superfamily and its spread by horizontal gene transfer. *Mol Microbiol*. 1993;9(5):915-921.

31. Henrissat B, Bairoch A. Updating the sequence-based classification of glycosyl hydrolases. *Biochem J*. 1996;316 : 695-696.

32. Varghese JN, Laver WG, Colman PM. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 A resolution. *Nature*. 1983;303(5912):35-40.

33. Varghese JN, McKimm-Breschkin JL, Caldwell JB, Kortt AA, Colman PM. The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. *Proteins*. 1992;14(3):327-332. doi: 10.1002/prot.340140302 [doi].

34. Shtyrya YA, Mochalova LV, Bovin NV. Influenza virus neuraminidase: Structure and function. *Acta Naturae*. 2009;1(2):26-32.

35. Taylor G. Sialidases: Structures, biological significance and therapeutic potential. *Curr Opin Struct Biol*. 1996;6(6):830-837. doi: S0959-440X(96)80014-5 [pii].

36. Kim S, Oh DB, Kang HA, Kwon O. Features and applications of bacterial sialidases. *Appl Microbiol Biotechnol*. 2011;91(1):1-15. doi: 10.1007/s00253-011-3307-2 [doi].

37. Buschiazzo A, Alzari PM. Structural insights into sialic acid enzymology. *Curr Opin Chem Biol*. 2008;12(5):565-572. doi: 10.1016/j.cbpa.2008.06.017 [doi].

38. Gaskell A, Crennell S, Taylor G. The three domains of a bacterial sialidase: A beta-propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure*. 1995;3(11):1197-1205.

39. Moustafa I, Connaris H, Taylor M, et al. Sialic acid recognition by vibrio cholerae neuraminidase. *J Biol Chem*. 2004;279(39):40819-40826.

40. Amaya MF, Watts AG, Damager I, et al. Structural insights into the catalytic mechanism of trypanosoma cruzi trans-sialidase. *Structure*. 2004;12(5):775-784.

41. Vimr ER. Microbial sialidases: Does bigger always mean better? *Trends Microbiol*. 1994;2(8):271-277.

42. Chan J, Sandhu G, Bennet AJ. A mechanistic study of sialic acid mutarotation: Implications for mutarotase enzymes. *Org Biomol Chem*. 2011;9(13):4818-4822.

43. Severi E, Muller A, Potts JR, et al. Sialic acid mutarotation is catalyzed by the escherichia coli beta-propeller protein YjhT. *J Biol Chem*. 2008;283(8):4841-4849.

44. Watts AG, Damager I, Amaya ML, et al. *Trypanosoma cruzi* trans-sialidase operates through a covalent sialyl-enzyme intermediate: Tyrosine is the catalytic nucleophile. *J Am Chem Soc*. 2003;125(25):7532-7533. doi: 10.1021/ja0344967 [doi].

45. Wicki J, Rose DR, Withers SG. Trapping covalent intermediates on beta-glycosidases. *Methods Enzymol*. 2002;354:84-105.

46. Damager I, Buchini S, Amaya MF, et al. Kinetic and mechanistic analysis of *Trypanosoma cruzi* trans-sialidase reveals a classical ping-pong mechanism with acid/base catalysis. *Biochemistry*. 2008;47(11):3507-3512. doi: 10.1021/bi7024832 [doi].

47. Watson JN, Dookhun V, Borgford TJ, Bennet AJ. Mutagenesis of the conserved active-site tyrosine changes a retaining sialidase into an inverting sialidase. *Biochemistry*. 2003;42(43):12682-12690. doi: 10.1021/bi035396g [doi].

48. Corfield AP, Higa H, Paulson JC, Schauer R. The specificity of viral and bacterial sialidases for alpha(2-3)- and alpha(2-6)-linked sialic acids in glycoproteins. *Biochim Biophys Acta*. 1983;744(2):121-126.

49. Monti E, Preti A, Venerando B, Borsani G. Recent development in mammalian sialidase molecular biology. *Neurochem Res*. 2002;27(7-8):649-663.

50. Sakurada K, Ohta T, Hasegawa M. Cloning, expression, and characterization of the *Micromonospora viridifaciens* neuraminidase gene in streptomyces lividans. *J Bacteriol*. 1992;174(21):6896-6903.

51. von Itzstein M, Thomson R. Anti-influenza drugs: The development of sialidase inhibitors. *Handb Exp Pharmacol*. 2009;(189):111-54. doi(189):111-154. doi: 10.1007/978-3-540-79086-0_5 [doi].

52. Chan J, Lu A, Bennet AJ. Turnover is rate-limited by deglycosylation for *Micromonospora viridifaciens* sialidase-catalyzed hydrolyses: Conformational implications for the michaelis complex. *J Am Chem Soc*. 2011;133(9):2989-2997. doi: 10.1021/ja109199p [doi].

53. Guo X, Laver WG, Vimr E, Sinnott ML. Catalysis by two sialidases with the same protein fold but different stereochemical courses: A mechanistic comparison of the enzymes from influenza A virus and salmonella typhimurium . *Journal of the American Chemical Society*. 1994;116(13):5572-5578.

54. Toney MD, Kirsch JF. Direct bronsted analysis of the restoration of activity to a mutant enzyme by exogenous amines. *Science*. 1989;243(4897):1485-1488.

55. Copeland RA. Reversible inhibitors. In: *Enzymes: A practical introduction to structure, mechanism, and data analysis.* 2nd ed.; 2002:266-304.

56. Vosseller K, Wells L, Lane MD, Hart GW. Elevated nucleocytoplasmic glycosylation by O-GlcNAc results in insulin resistance associated with defects in akt activation in 3T3-L1 adipocytes. *Proc Natl Acad Sci U S A*. 2002;99(8):5313-5318. doi: 10.1073/pnas.072072399 [doi].

57. Kajimoto T, Node M. Inhibitors against glycosidases as medicines. *Curr Top Med Chem*. 2009;9(1):13-33.

58. von Itzstein M, Wu WY, Kok GB, et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*. 1993;363(6428):418-423. doi: 10.1038/363418a0 [doi].

59. Vasella A, Davies GJ, Bohm M. Glycosidase mechanisms. *Curr Opin Chem Biol*. 2002;6(5):619-629. doi: S1367593102003800 [pii].

60. Siriwardena A, Strachan H, El-Daher S, et al. Potent and selective inhibition of class II α-D-mannosidase activity by a bicyclic sulfonium salt. *ChemBioChem*. 2005;6(5):845-848. doi: 10.1002/cbic.200400397.

61. Wolfenden R. Degrees of difficulty of water-consuming reactions in the absence of enzymes. *Chem Rev*. 2006;106(8):3379-3396. doi: 10.1021/cr050311y [doi].

62. Powell KA, Ramer SW, Del Cardayre SB, et al. Directed evolution and biocatalysis. *Angew Chem Int Ed Engl*. 2001;40(21):3948-3959. doi: 10.1002/1521-3773(20011105)40:213.0.CO;2-N [pii].

63. Murashima K, Kosugi A, Doi RH. Thermostabilization of cellulosomal endoglucanase EngB from *Clostridium cellulovorans* by in vitro DNA recombination with non-cellulosomal endoglucanase EngD. *Mol Microbiol*. 2002;45(3):617-626. doi: 3049 [pii].

64. Hirokawa K, Kajiyama N, Murakami S. Improved practical usefulness of firefly luciferase by gene chimerization and random mutagenesis. *Biochim Biophys Acta*. 2002;1597(2):271-279. doi: S0167483802003023 [pii].

65. Bevis BJ, Glick BS. Rapidly maturing variants of the discosoma red fluorescent protein (DsRed). *Nat Biotechnol*. 2002;20(1):83-87. doi: 10.1038/nbt0102-83 [doi].

66. Campbell RE, Tour O, Palmer AE, et al. A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A*. 2002;99(12):7877-7882. doi: 10.1073/pnas.082243699 [doi].

67. Osanjo G, Dion M, Drone J, et al. Directed evolution of the alpha-L-fucosidase from *Thermotoga maritima* into an alpha-L-transfucosidase. *Biochemistry*. 2007;46(4):1022-1033. doi: 10.1021/bi061444w [doi].

68. Saumonneau A, Champion E, Peltier-Pain P, et al. Design of an alpha-L-transfucosidase for the synthesis of fucosylated HMOs. *Glycobiology*. 2016;26(3):261-269. doi: 10.1093/glycob/cwv099 [doi].

69. Smith JM. Natural selection and the concept of a protein space. *Nature*. 1970;225(5232):563-564.

70. Wright S. Evolution in mendelian populations. *Genetics*. 1931;16(2):97-159.

71. Cedrone F, Menez A, Quemeneur E. Tailoring new enzyme functions by rational redesign. *Curr Opin Struct Biol*. 2000;10(4):405-410. doi: S0959-440X(00)00106-8 [pii].

72. Lai YP, Huang J, Wang LF, Li J, Wu ZR. A new approach to random mutagenesis in vitro. *Biotechnol Bioeng*. 2004;86(6):622-627. doi: 10.1002/bit.20066 [doi].

73. Myers RM, Lerman LS, Maniatis T. A general method for saturation mutagenesis of cloned DNA fragments. *Science*. 1985;229(4710):242-247.

74. Bridges BA, Woodgate R. Mutagenic repair in *Escherichia coli*: Products of the recA gene and of the umuD and umuC genes act at different steps in UV-induced mutagenesis. *Proc Natl Acad Sci U S A*. 1985;82(12):4193-4197.

75. Leung DW, Chen E, Goeddel DV. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. 1, 11–15 (1989). *Technique*. 1989;1:11-15.

76. Cirino PC, Mayer KM, Umeno D. Generating mutant libraries using error-prone PCR. *Methods Mol Biol*. 2003;231:3-9. doi: 1-59259-395-X-3 [pii].

77. Eckert KA, Kunkel TA. High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucleic Acids Res*. 1990;18(13):3739-3744.

78. Siloto RMP, Weselake RJ. Site saturation mutagenesis: Methods and applications in protein engineering. *Biocatalysis and Agricultural Biotechnology*. 2012;1(3):181-189. doi: http://dx.doi.org/10.1016/j.bcab.2012.03.010.

79. Hughes MD, Nagel DA, Santos AF, Sutherland AJ, Hine AV. Removing the redundancy from randomised gene libraries. *Journal of Molecular Biology*. 2003;331:973-979.

80. Reetz MT, Bocola M, Carballeira JD, Zha D, Vogel A. Expanding the range of substrate acceptance of enzymes: Combinatorial active-site saturation test. *Angew Chem Int Ed Engl*. 2005;44(27):4192-4196. doi: 10.1002/anie.200500767 [doi].

81. Stemmer WP. Rapid evolution of a protein in vitro by DNA shuffling. *Nature*. 1994;370(6488):389-391. doi: 10.1038/370389a0 [doi].

82. Stemmer WP. DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution. *Proc Natl Acad Sci U S A*. 1994;91(22):10747-10751.

83. Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotechnol*. 1998;16(3):258-261. doi: 10.1038/nbt0398-258 [doi].

84. Zhang JH, Dawes G, Stemmer WP. Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. *Proc Natl Acad Sci U S A*. 1997;94(9):4504-4509.

85. Parikh MR, Matsumura I. Site-saturation mutagenesis is more efficient than DNA shuffling for the directed evolution of beta-fucosidase from beta-galactosidase. *J Mol Biol*. 2005;352(3):621-628. doi: S0022-2836(05)00808-9 [pii].

86. Schmidt-Dannert C, Arnold FH. Directed evolution of industrial enzymes. *Trends Biotechnol*. 1999;17(4):135-136. doi: S0167-7799(98)01283-9 [pii].

87. Aharoni A, Griffiths AD, Tawfik DS. High-throughput screens and selections of enzyme-encoding genes. *Curr Opin Chem Biol*. 2005;9(2):210-216. doi: S1367-5931(05)00017-7 [pii].

88. Goddard JP, Reymond JL. Recent advances in enzyme assays. *Trends Biotechnol*. 2004;22(7):363-370. doi: 10.1016/j.tibtech.2004.04.005 [doi].

89. Leemhuis H, Kelly RM, Dijkhuizen L. Directed evolution of enzymes: Library screening strategies. *IUBMB Life*. 2009;61(3):222-228. doi: 10.1002/iub.165 [doi].

90. Fulwyler MJ. Electronic separation of biological cells by volume. *Science*. 1965;150(3698):910-911.

91. Reetz MT, Hobenreich H, Soni P, Fernandez L. A genetic selection system for evolving enantioselectivity of enzymes. *Chem Commun (Camb)*. 2008;(43):5502-4. doi(43):5502-5504. doi: 10.1039/b814538e [doi].

92. Brakmann S. Discovery of superior enzymes by directed molecular evolution. *Chembiochem*. 2001;2(12):865-871. doi: 10.1002/1439-7633(20011203)2:123.0.CO;2-6 [pii].

Chapter 2. Directed Evolution of the Sialidase from *Micromonospora viridifaciens* into a Kdnase

2.1. Introduction

Sialic acids are found in a wide range of organisms including higher eukaryotes as well as bacteria[1,2], fungi[3] and protozoans[4]. Both the pathogenic and non-pathogenic bacteria and protozoans, have gained the ability either to synthesize these nine carbon negatively charged sugars or to capture them from their environment for various uses including as a carbon or nitrogen source. Some pathogenic microorganisms have evolved to bind to host sialic acid glycoconjugates and to transfer these sugars to their cell surface in order to evade the host's immune response. Additionally, these cell surface sialic acids are utilized specifically to interact with host cell receptors; for example, *Campylobacter jejuni*, the bacterium responsible for both gastroenteritis and Guillain-Barré syndrome, interacts with host macrophages by binding to sialic-binding lectins (Siglec-1)[5]. Replacement of the aminoacyl groups of sialic acid at C5 with a hydroxyl group generates Kdn. While Kdn is structurally very similar to Neu5Ac, this seemingly minute change leads to near complete abolishment of the hydrolytic activity of bacterial *exo*-sialidases that are commonly used for the identification and structure-function studies of sialoglycoconjugates[6,7].

This intrinsic inhibitory property of Kdn glycosides towards many sialidases may have a potential role in the protection of cells and organs, adorned with Kdn-conjugated glycoproteins, from bacterial and viral attacks[6,8,9]. Of note, Kdn-containing glycoconjugates are widely distributed, in a variety of biological sources including in mammalian tissues[10], human lung and ovarian cancer cells[11], rainbow trout (Salmo gairdneri) eggs[12], as well as bacterial sources.[13]. However, this resistance to cleavage by cellular *exo*-sialidases poses a problem for the elucidation of the exact biological roles of Kdn-conjugates. As a result, the availability of a sialidase, "Kdnase" that could selectively and quantitatively cleave the α-ketosidic linkages in Kdn-glycoconjugates would be an important chemical biology tool for studies that seek to determine the structure, biosynthesis, localization, and function of Kdn-glycoconjugates.

To this end, we constructed a genetic library, using the technique of error prone PCR to introduce random mutations in the gene that encodes the wild-type sialidase from the soil bacterium (*Mv*NA). We identified a number of recurring DNA mutations in the sialidase gene *Micromonospora viridifaciens* that gave rise to a more efficient hydrolysis of synthetic Kdn natural substrate analogues, such as 8FMU α-Kdn-(2→6)-β-D-Galp. In the interest of time, we used the available 3D structure of wild type *Mv*NA bound to the natural inhibitor, DANA (Neu2en5Ac) and identified amino acids potentially involved in recognition and binding to the *N*-acetyl group in sialoside substrates. Next, we performed site directed saturation mutagenesis of these amino acids to generate a mutagenesis library. After an additional round of random mutagenesis, by subjecting the library to multiple screen and selection assays, we were able to isolate modified sialidases with enhanced Kdnase cleavage activity.

2.2. Methods and Materials

2.2.1. Chemicals and Reagents

N-acetylneuraminic acid was purchased from Rose Scientific. Dithiothreitol, bactotryptone, yeast extract and urea were purchased from Bioshop. Glycerol was purchased from VWR. Cytidine 5′-triphosphate disodium salt (CTP) was purchased from 3B Scientific Corporation. Potassium phosphate, dibasic and monobasic, sodium acetate and magnesium chloride were purchased from Caledon Laboratories Ltd. Imidazole was purchased from AK Scientific, Inc. and IPTG was purchased from Invitrogen. All other chemicals and reagents were purchased from Sigma-Aldrich Canada, Ltd.

2.2.2. Enzymes

Micromonospora viridifaciens sialidase was expressed and purified as previously reported [14]. Phusion DNA polymerase and T4 DNA ligase were purchased from New England Biolabs. All restriction endonucleases were purchased from Thermo Scientific. *Escherichia coli N*-acetylneuraminic acid (Neu5Ac) aldolase was purchased from Codexis. *Neisseria meningitidis* CMP-Neu5Ac synthase[15] and *Photobacterium sp*. JT-ISH-224 α-

2,6-sialyltransferase [16] were expressed as reported. GeneMorph II Random Mutagenesis kit was purchased from Agilent Technologies.

2.2.3. Substrate synthesis

8FMU α-Kdn-(2→6)-β-D-Galp and 8FMU α-Neu5Ac-(2→6)-β-D-Galp were synthesized by Dr. Fahimeh Shidmoossavee. *O*-β-D-galactopyranosyl chloramphenicol and *O*-β-D-galactopyranosyl-L-tyrosine were synthesized by Matthew Deen. These compounds were used for the enzymatic synthesis Kdn and Neu5Ac conjugates using the enzymes CMP-sialic acid synthetase and α-2,6-sialyltransferase.

Enzymatic synthesis of α-Kdn-(2→6)-β-D-Gal-Tyrosine, α-Neu5Ac -(2→6)-β-D-Gal-Tyrosine and α-Neu5Ac-(2→6)-β-D-galactopyranosyl chloramphenicol

We used a published protocol for the enzyme-catalyzed synthesis of Kdn and Neu5Ac conjugates with minor modifications[17]. Specifically, we added fresh supernatant (500 µL) from an expression of CMP-sialic acid synthetase[15] to a 50 mL falcon tube containing a reaction medium of *N*-acetylneuraminic acid (0.372 mmol) or Kdn (0.410 mmol), cytidine 5′-triphosphate disodium salt (1.12 mmol), $MgCl_2$ (200 mM) and dithiothreitol (4 mM) in Tris buffer (500 mM; pH 8.01) for a total volume of 10 mL. The reaction was incubated for 3 hr at 37 ^{0}C and centrifuged at 4000 rpm for 30 min. The supernatant was separated from the pellet and transferred to a new 50 mL falcon tube, after which *O*-β-D-galactopyranosyl chloramphenicol (0.95 mmol) or *O*-β-D-galactopyranosyl-L-tyrosine (0.82 mmol) and fresh supernatant (500 µL) from a α-2,6-sialyl transferase expression[16] were added to the solution, and the resultant mixture was incubated at 37 °C for 5 hours. The progress of the enzymatic reaction was monitored by ^{1}H NMR spectroscopy.

Purification of α-Kdn-(2→6)-β-D-Gal-Tyrosine and α-Neu5Ac -(2→6)-β-D-Gal-Tyrosine

The reaction mixture was centrifuged at 4000 rpm for 30 min and the supernatant was loaded onto a column packed with Sephadex G-10 media (GE Healthcare Life Sciences). The column was eluted using water under low pressure. Fractions containing product were identified by Thin-layer chromatography (TLC), pooled and lyophilized to

afford the product as a white solid. The presence of the desired product was further confirmed through ^{1}H NMR spectroscopy.

Purification of α-Neu5Ac-(2→6)-β-D-Gal-chloramphenicol

The reaction mixture was centrifuged at 4000 rpm for 30 min and the supernatant was loaded onto a reversed-phased C18 sep-pak cartridge (20 cc, 5 g), and the cartridge was eluted successively with H_2O (~50 mL) and 1:19 v/v MeCN/H_2O (~150 mL). The fractions containing product were identified by Thin-layer chromatography (TLC), pooled and lyophilized to afford the product as a white solid. The presence of the desired product was further confirmed through ^{1}H NMR spectroscopy.

2.2.4. Random Mutant Library Construction and Expression

The pJWβ-OHS plasmid was used as the template for the sialidase gene [14]. A library of random mutants of the sialidase gene was generated via error prone PCR using the GeneMorph II Random Mutagenesis kit according to manufacturer's instructions. The conditions were set such that an approximate frequency of 1 mutation per 10 base pairs was achieved. The 5'-end of the gene was amplified using the Eco* primer and the 3'-end of the gene was amplified using the ggHindR reverse primer to produce a 1.2 kb fragment (Table 2.1).

Table 2.1 DNA oligonucleotide sequences

Primer	Nucleotide Sequence
Eco*	5'-CCGGAATTCACTGCGAATCCGTACCTCCGC-3'
ggHindR	5'-CCCAAGCTTCAGCCAGGCGAGGTTG-3'

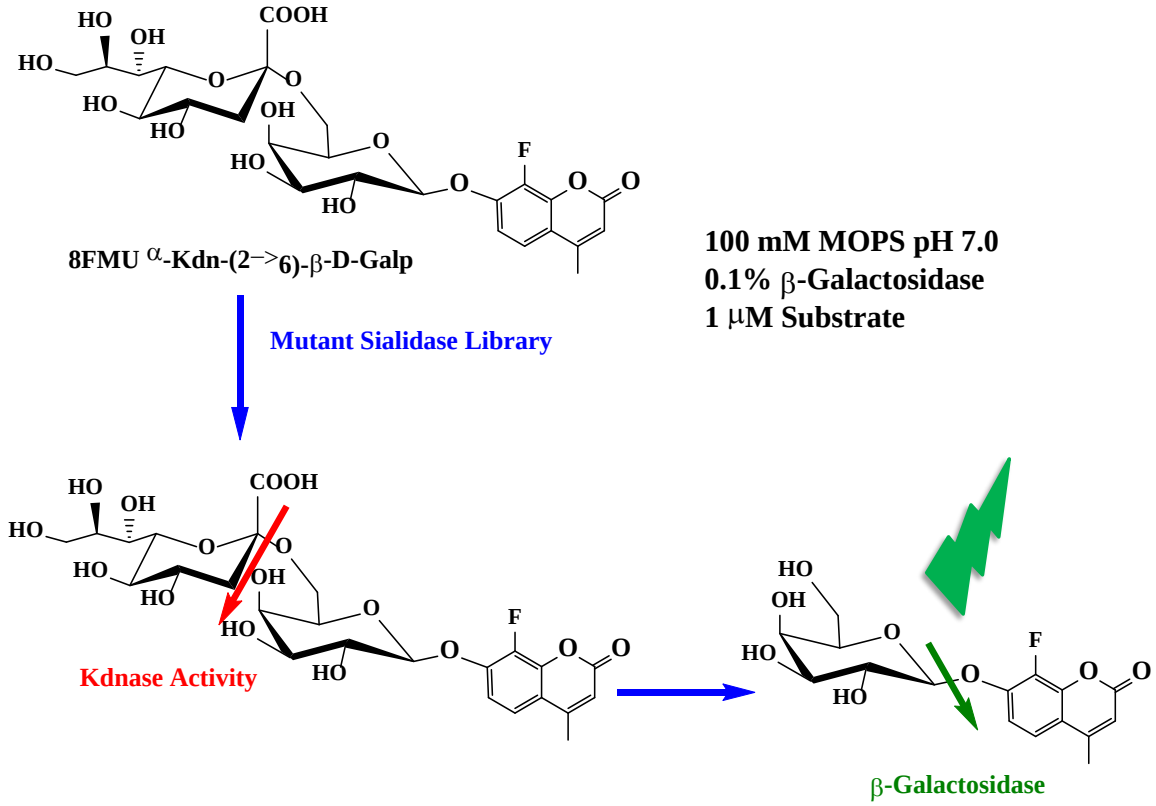
The full 1.24 kb gene and pJWβ plasmid were digested with *Eco* RI and *Hind* III. The digested insert gene and plasmid vector were separated on 1% agarose gel and the insert was excised and purified. The purified restriction enzyme-digested insert was ligated into purified digested vector with 2:1 ratio using T4 DNA ligase. The ligation reaction was incubated at 16 °C overnight.

20-40 ng of DNA from the ligation mixture was transformed into *E. coli* BL21-Gold (DE3) competent cells (Agilent Technologies) and plated on Luria broth (LB) agar media supplemented with 30 μg/mL Kanamycin antibiotic. Five thousand (5000) individual clones were picked into in 96 well plates containing M9 minimal media supplemented with 30 μg/mL Kanamycin, using Molecular Devices Qpix-2-xt colony picker. Following a 4-hour incubation at 37 °C to allow for growth, protein production from the transformed plasmids was induced by adding IPTG to a final 1.5 mM concentration in each well. The plates were incubated for 5 hours at room temperature, 120 RPM to allow for ample protein production. The plates were then centrifuged (3500 rmp, 20 min at 4 °C) to pellet the cells as the protein of interest is exported into the media.

2.2.5. Random Mutagenesis Library Screening

The rationale behind library screening protocol is shown in Scheme 2.1.

A cocktail containing the buffer MOPS (100 mM, pH 7.0), β-galactosidase (final percent solution of 0.1%) and 8FMU α-Kdn-(2→6)-β-D-Galp (final concentration 1 μM) was added to each well and the plates were incubated overnight. These 96 well plates were mounted into a Cary Eclipse fluorescence plate reader and the fluorescence emission was measured (excitation wavelength = 355 nm and emission wavelength = 485 nm). Each plate contained two (2) reference wells; one well with all the components except for the expressed enzyme, which served to identify the background fluorescence, and one well containing wild type sialidase incubated with 8FMU α-Kdn-(2→6)-β-D-Galp (1 μM). The measured data was then used to identify clones with at least 2.5 times the fluorescence intensity of the background and those clones were mini-prepped using QIAprep Spin Miniprep Kit (Qiagen) and sent for sequencing.



Scheme 2.1 The fluorescent reading of individual clones from the random mutant library – protocol 1. Clones expressing modified enzyme capable of cleaving Kdn linkages will hydrolyze the Kdn-gal linkage, releasing the 8 FMU-gal moiety which is then hydrolyzed by β-galactosidase, producing fluorescence at significantly higher levels than the background.

2.2.6. Sequencing Analysis

Sequencing of clones were performed by Eurofins MWG Operon. The sequencing results were analyzed using BioEdit Sequence Alignment Editor and Multiple Sequence Alignment tool, Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/).

2.2.7. Site Directed Mutagenesis Library Construction

A specific mutation that was noted in 3 individual clones, Leu135His, from the random mutagenesis library was incorporated into the wild-type *Mv*NA using the QuikChange™ Site-Directed Mutagenesis System developed by Stratagene. The primer

set used are labelled Leu135For and Leu135Rev with the intended mutation underlined (Table 2.2)

Table 2.2 DNA oligonucleotide sequences.
N refers to any four nucleotides, 25 % of A, T, C, or G. K refers to 50 % of either T or G nucleotide and M refers to 50 % of either A or C nucleotide.

Primer	Nucleotide Sequence
Leu135For	5'-CGACCCCAGCTACCATGTCGACCGGGAAAC-3'
Leu135Rev	5'-GTTTCCCGGTCGACATGGTAGCTGGGGTCG-3'
KDN-Sat(3)-F	5'-TACTCCCAGCGGCAGNNKTTCNNKGGCAGC-3'
KDN-Sat(3)-R	5'-CTGCCGCTGGGAGTAMNNGTGGAAGTTGAA-3'
NcoI.pSF	5'TATGCCATGGGAACTGCGAATCCGTACCTCCGCC-3'
XbaI.pSR	5'-TATCTCTAGACCAGCCAGGCGAGGTTGAAGTT-3'

Additionally, taking advantage of the available structure of wild type *Mv*NA bound to the natural inhibitor, DANA,(Figure 2.2) we identified three amino acids potentially involved in recognition and binding to the *N*-acetyl group in sialosides and we proceeded to include saturation mutations to these selected amino acids.

Saturation site-directed mutations were made to the following three amino acids, Val148, Gly154 and Ala156 using the *Mv*NA containing the Leu135His mutation as the template. Two mutagenic primers as well as two plasmid specific primers were designed to produce a library of mutants using the strand-overlap extension method[18].

Mutagenesis was ensured by making the mutations in each strand in separate PCR experiments. The 5'-end (front) of the gene was amplified using the NcoI.pSF primer with the reverse mutagenic primer KDN-Sat(3)-R to give a 0.46 kb product. The 3'-end (back) of the gene was amplified using the forward mutagenic primer KDN-Sat(3)-F with

the XbaI.pSR reverse primer to produce a 0.78 kb fragment. All PCR reactions used Phusion polymerase (ThermoFisher Scientific) with HF buffer in 10% DMSO and 0.25 mM dNTPs. The front and back fragment portion of PCR reactions were performed at a denaturing temperature of 98 °C for 30 secs (1 cycle) followed by denaturing at 98 °C for 10 sec, annealing at 55 °C for 30 sec, elongation at 72 °C for 1 min (30 cycles), the final elongation at 72 °C was allowed to react for 2 min (1 cycle). The fragments were separated on a 1% agarose gel, sliced and purified using GeneJET Gel Extraction kit (ThermoFisher Scientific). The concentration of the purified samples was determined using NanoDrop Microvolume Spectrophotometer (ThermoFisher Scientific) and stored at -20°C freezer.

The full fragment amplification PCR was set up by using equal amounts of the above purified fragments as the templates. One complete run (98 °C for 40 sec, 50 °C for 1 min, 72 °C for 1:10 min) was allowed to run without addition of primers to allow the fragments to anneal at the overlapping region. Following this step, by addition of primers, NcoI.pSF and XbaI.pSR primers (0.5 μM), the full 1.24 kb fragment was amplified and purified using GeneJET PCR Purification Kit (ThermoFisher Scientific).

The full 1.24 kb gene and pSF-OXB19 (Oxford Genetics) plasmid were digested with *NcoI* and *XbaI*. The digested insert gene and plasmid vector were separated on 1% agarose gel and were then sliced and purified. The purified restriction enzyme-digested insert was ligated into purified digested vector with 2:1 ratio using T4 DNA ligase. The ligation reaction was incubated at 16 °C overnight followed by the addition of 1 μL fresh ligase and the incubation was continued for another 3 hours at room temperature. The ligation mixture was then transformed into *E. coli* BL21-Gold(DE3) competent cells (Agilent Technologies). Twenty clones were picked to perform restriction and sequence diversity check. Following confirmation of the saturation mutagenesis at selected amino acids, the library was subjected to another round of error prone PCR as previously described and cloned back into pSF-OXB19 vector.

2.2.8. Site Directed Mutagenesis Library Screening and Selection Process

Tyrosine Auxotroph Cell Line

For expression and screening purposes, we took advantage of an *E. coli* Tyrosine auxotroph cell line, JW2581-1[19] that contains a deletion of the gene *tyrA*, part of the subpathway for biosynthesis of L-tyrosine in *E. coli*. The cell line was ordered from *E. coli* Genetic Stock Center (CGSC) at Yale University. Competent cells were prepared from this stock according to a published protocol from the Untergasser Laboratory[20].

Initial Parameters of Positive and Negative Evolutionary Forces Directed Screening

To determine the initial amounts of α-Kdn-(2→6)-β-D-Gal-Tyrosine and of α-Neu5Ac-(2→6)-β-D-Gal-chloramphenicol needed for the screen, we inoculated MOPS minimal media[21] in a 48 well plate, supplemented with an increasing concentration of α-Neu5Ac -(2→6)-β-D-Gal-chloramphenicol (0.25 – 2.5 µg/ mL) with an overnight culture of wildtype *Mv*NA cloned into pSF-OXB19 vector and transformed into the Tyr auxotroph line. Following a 7-hour incubation at 37 ^{0}C, the absorbance of each well was measured at 600 nm and used as an indication of relative growth. A parallel plate, supplemented with increasing concentration of α-Neu5Ac-(2→6)-β-D-Gal-Tyrosine (50 – 250 µg/mL) was inoculated with *Mv*NA in a similar manner. Additionally, a third plate of MOPS minimal media supplemented with increasing concentration of α-Kdn-(2→6)-β-D-Gal-Tyrosine (200 – 500 µg/ mL) was inoculated with *Mv*NA.

Screening Process

For the screening purposes, 30-40 ng of ligation mix was transformed into 100 µL of Tyrosine auxotroph line and plated on MOPS minimal media supplemented with 2 µg/mL of α-Neu5Ac-(2→6)-β-D-Gal-chloramphenicol and 350 µg/mL of α-Kdn-(2→6)-β-D-Gal-Tyrosine. 150 colonies growing on this media, were then individually picked using Molecular Devices Qpix-2-xt colony picker and inoculated into 200 µL of MOPS media supplemented with 50 µg/mL L-Tyrosine and 2 µg/mL of chloramphenicol and allowed to grow for 7 hours at 37 ^{0}C to be used as cultures for further selection steps. The growth rate of each well was estimated by measuring the absorbance at 600 nm. The culture

containing these colonies then used to inoculate wells of MOPS minimal media supplemented with increasing concentrations of α-Neu5Ac-(2→6)-β-D-Gal-chloramphenicol until the number of colonies were reduced to 12 colonies. These clones have reduced sialidase activity. The selected 12 colonies were then grown on decreasing concentration of α-Kdn-(2→6)-β-D-Gal-Tyrosine to identify clones with enhanced KDNase activity.

2.2.9. Purification and Kinetics Analysis

Selected colonies from the selection process were then grown in Terrific Broth (TB) media supplemented with 50 μg/mL spectinomycin antibiotic and mini-prepped. The plasmids were used as template to amplify the 1.24 kb gene using Eco* and HindR primers and cloned into pET28a vector. This would ensure the presence of a Histidine tag at the C-terminus and purification of the protein of interest by Ni-NTA column. All purifications were performed using the published protocol[22].

Calibration curves were constructed by incubating 8FMU in the range of 0.2 to 2.0 μM in MES (100 μM) at pH value of 5.5 and MOPS (100 μM) at pH value of 6.2, 6.9 and 7.4. To determine the enzymatic efficiency of selected clones in hydrolysis of Neu5Ac vs. Kdn glycosides, following purification and quantification of each individual enzyme, equal volumes were added to a 200 μL mix of MOPS (100 μM), 8FMU α-Kdn-(2→6)-β-D-Galp or 8FMU α-Neu5Ac-(2→6)-β-D-Galp (10 μM) and β-galactosidase (0.1%). The reaction was monitored at 5-minute intervals for 800 mins at an excitation and emission wavelength of 365 nm and 465 nm respectively.

2.3. Results

Random Mutagenesis Library Screen

A total of fifty (50) 96-well plates were screened using screening protocol 1. The background control reactions were performed in parallel. 155 clones with a fluorescent reading at least 2 times larger than the background control were identified. These clones

were individually grown, the corresponding plasmid was mini-prepped and sent for sequencing. Table 2.3 outlines these findings.

Table 2.3 Fluorescent readings based on protocol 1

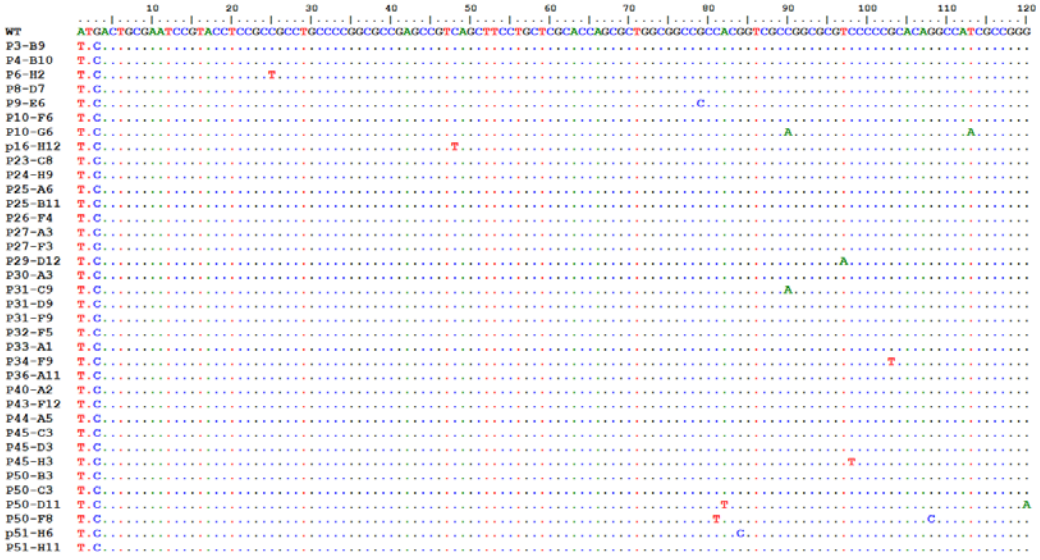
Fluorescent Controls	Fluorescent Intensity (Arbitrary Value)	Number of Clones Identified
Background	55	4874
Background X 2	110	152

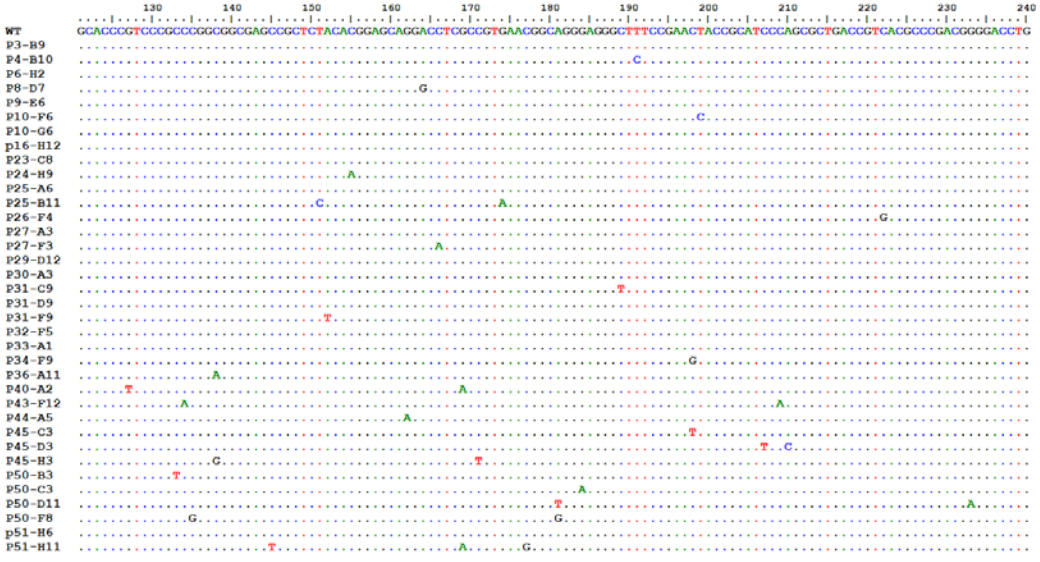
The sequencing results for each clone were aligned with the wildtype sequence to identify potential mutations favoring increased Kdnase activity. Figure 2.1 is a representation of the sequencing results. As is evident from the sequencing results, there are regions within the gene sequences where mutations are not tolerated while certain regions contain multitude of mutations.

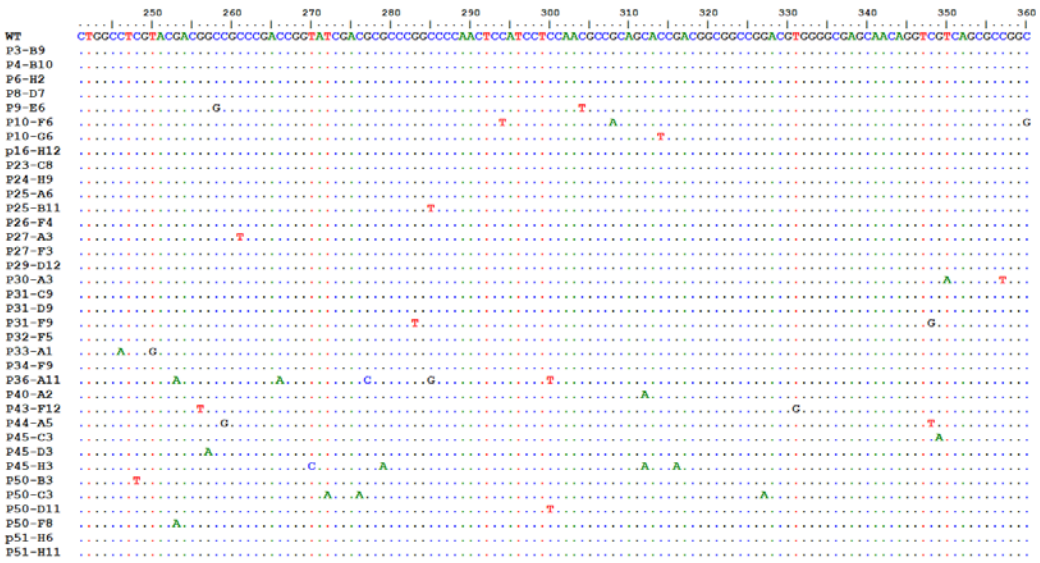
In the interest of time, based on the 3D structure of the DANA bound *Mv*NA (PDB code 1EUS), we looked at the structural loops close to the active site which theoretically would play an important role in the selectivity of the active site. Through this approach, we identified three separate clones where a Leu135His mutation was observed. We deduced this mutation likely increases the Kdnase activity.

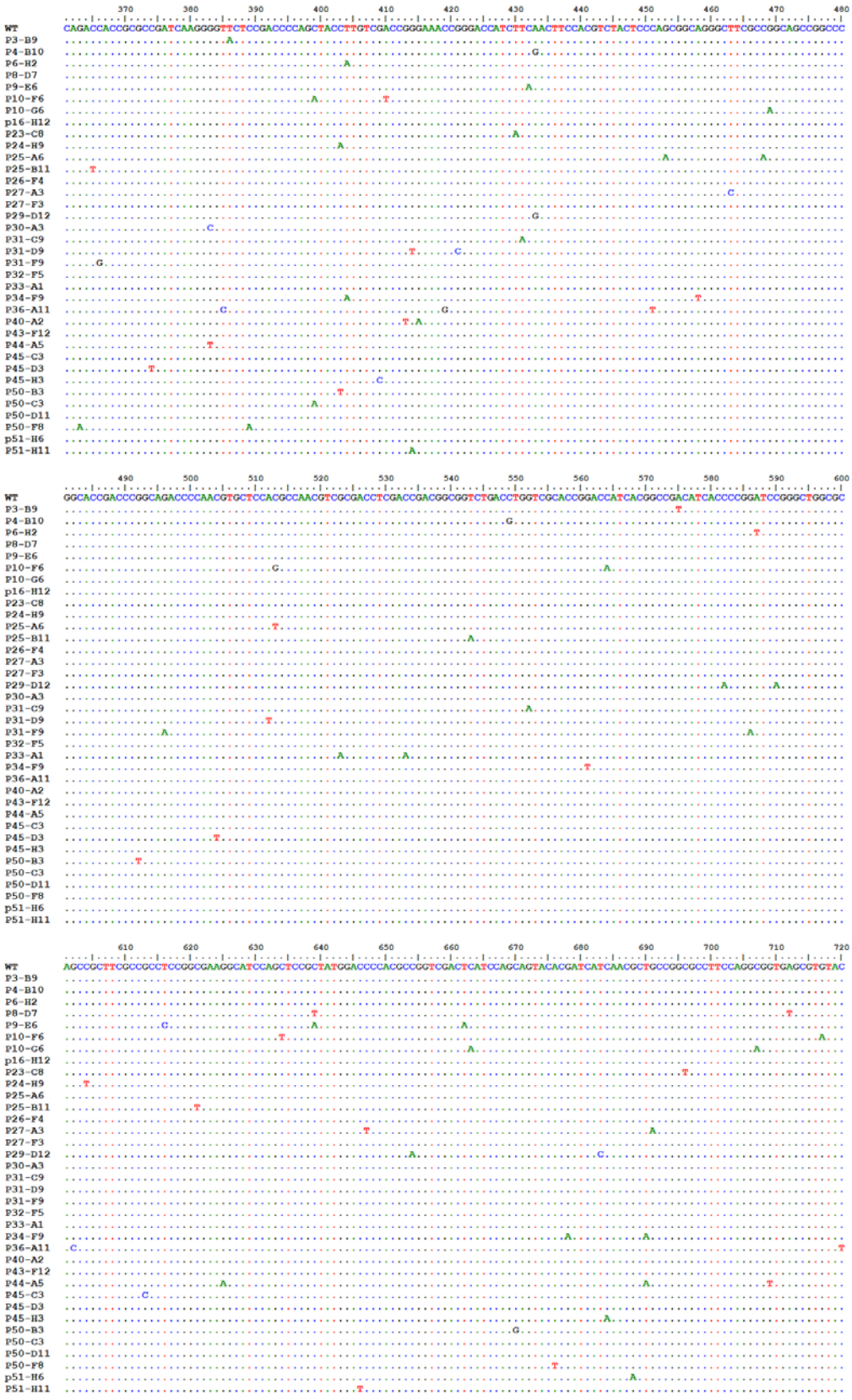
2.3.1. Site Directed and Random Mutagenesis Library Generation

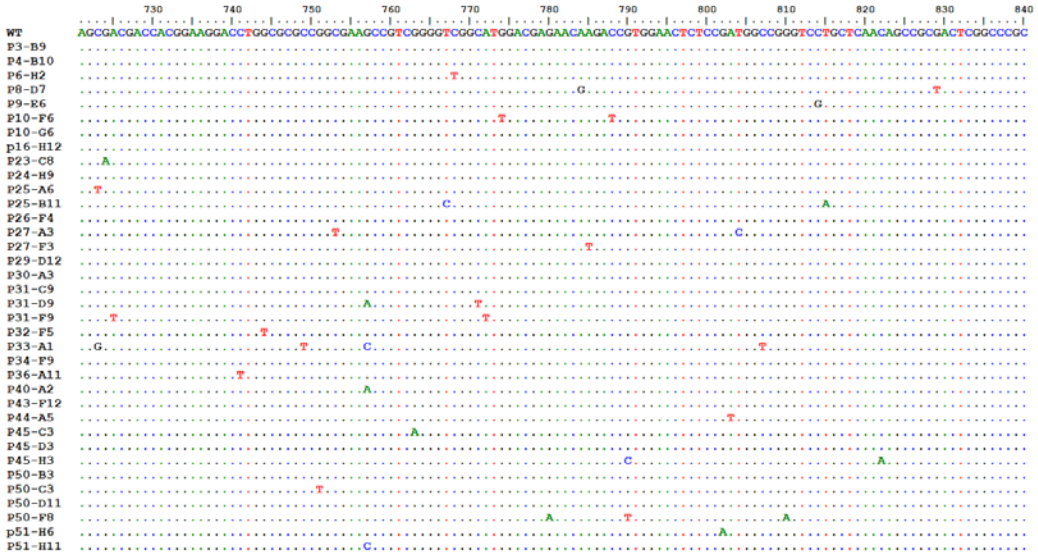
Taking advantage of the 3D structure of the DANA bound *Mv*NA (PDB code 1EUS)[23], we identified the amino acid residues Val148, Gly154 and Ala156 as residues that would potentially be crucial in conferring the ability of recognizing Kdn over Neu5Ac selection and hydrolysis to the modified enzymes (Figure 2.2). A saturation mutagenesis protocol on these amino acids as well as introduction of the Leu135His mutation identified in the random mutagenesis library screen into wildtype *Mv*NA allowed for generation of a library that would be used for further mutagenesis and screening/selection protocols. Figure 2.3 is a depiction of the range of mutations that were introduced in the DNA sequences coding for these three amino acids.


10 20 30 40 50 60 70 80 90 100 110 120
WT ATGACTGCGAATCCGTACCTCCCGCCGCCCTGCCCCGGCGCCGAGCCGTCAGCTTCCTGCTCGCACCAGCGCTGGCGGCCGCCACGGTCGCCGGCGCGTCCCCCGCACAGGCCATCGCCGGG
P3-B9 T.C
P4-B10 T.C
P6-H2 T.C T
P8-D7 T.C
P9-E6 T.C C
P10-F6 T.C
P10-G6 T.C A A
p16-H12 T.C T
P23-C8 T.C
P24-H9 T.C
P25-A6 T.C
P25-B11 T.C
P26-F4 T.C
P27-A3 T.C
P27-F3 T.C
P29-D12 T.C A
P30-A3 T.C
P31-C9 T.C A
P31-D9 T.C
P31-F9 T.C
P32-F5 T.C
P33-A1 T.C
P34-F9 T.C T
P36-A11 T.C
P40-A2 T.C
P43-F12 T.C
P44-A5 T.C
P45-C3 T.C
P45-D3 T.C
P45-H3 T.C T
P50-B3 T.C
P50-C3 T.C
P50-D11 T.C T A
P50-F8 T.C T C
p51-H6 T.C G
P51-H11 T.C

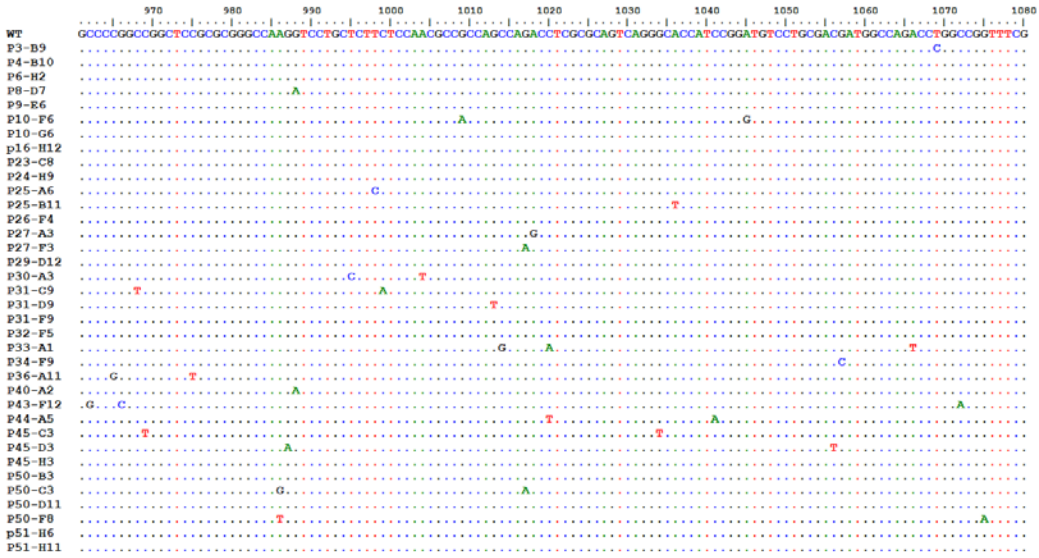

130 140 150 160 170 180 190 200 210 220 230 240
WT GCACCGTCCCGCCCGGCGGCGAGCCGCTCTACACGGAGCAGGACGTCGCCGTGAACGGCAGGGAGGGCTTTCCGAACTACCGCATCCCAGCGCTGACCGTCACGCCCGACGGGGACCTG
P3-B9
P4-B10 C
P6-H2
P8-D7 G
P9-E6
P10-F6 C
P10-G6
p16-H12
P23-C8
P24-H9 A
P25-A6
P25-B11 C A
P26-F4 G
P27-A3
P27-F3 A
P29-D12
P30-A3
P31-C9 T
P31-D9
P31-F9 T
P32-F5
P33-A1
P34-F9 G
P36-A11 A
P40-A2 T A
P43-F12 A A
P44-A5 A
P45-C3 T
P45-D3 T C
P45-H3 G T
P50-B3 T
P50-C3 A
P50-D11 T A
P50-F8 G G
p51-H6
P51-H11 T A G


250 260 270 280 290 300 310 320 330 340 350 360
WT CTGGCCTCGTACGACGGCCGCCCGACCGGTATCGACGCGCCCCGGCCCCAACTCCATCCTCCAACGCCGCAGCACCGACGGCCGCCGGACGTGGGGCCGAGCAACAGCTCGTCAGCGCCGGC
P3-B9
P4-B10
P6-H2
P8-D7
P9-E6 G T
P10-F6 T A G
P10-G6 T
p16-H12
P23-C8
P24-H9
P25-A6
P25-B11 T
P26-F4
P27-A3 T
P27-F3
P29-D12
P30-A3 A T
P31-C9
P31-D9
P31-F9 T G
P32-F5
P33-A1 A G
P34-F9
P36-A11 A A C G T A
P40-A2 A
P43-F12 T G
P44-A5 G T
P45-C3 A
P45-D3 A
P45-H3 C A A A
P50-B3 T
P50-C3 A A A
P50-D11 T
P50-F8 A
p51-H6
P51-H11


WT
P3-B9
P4-B10
P6-H2
P8-D7
P9-E6
P10-F6
P10-G6
p16-H12
P23-C8
P24-H9
P25-A6
P25-B11
P26-F4
P27-A3
P27-F3
P29-D12
P30-A3
P31-C9
P31-D9
P31-F9
P32-F5
P33-A1
P34-F9
P36-A11
P40-A2
P43-F12
P44-A5
P45-C3
P45-D3
P45-H3
P50-B3
P50-C3
P50-D11
P50-F8
p51-H6
P51-H11


730 740 750 760 770 780 790 800 810 820 830 840
WT AGCGACGACCACGGAAGGACCTGGCGCGCGGCGGAAGCCGTCGGGGTCGGCATGGACGAGAACAAGACCGTGGAACTCTCCGATGGCCGGGTCCTGCTCAACAGCGCGACTCGGCCCGC
P3-B9
P4-B10
P6-H2
P8-D7
P9-E6
P10-F6
P10-G6
p16-H12
P23-C8
P24-H9
P25-A6
P25-B11
P26-F4
P27-A3
P27-F3
P29-D12
P30-A3
P31-C9
P31-D9
P31-F9
P32-F5
P33-A1
P34-F9
P36-A11
P40-A2
P43-F12
P44-A5
P45-C3
P45-D3
P45-H3
P50-B3
P50-C3
P50-D11
P50-F8
p51-H6
P51-H11


850 860 870 880 890 900 910 920 930 940 950 960
WT AGCGGATACCGTAAGGTGGCCGTCTCCACTGACGGCGGCCACAGCTACGGCCCGGTGACCATCGACCGCGACCTCCCCGACCCGACGAACAACGCATCGATCATCCGGGCTTCCCTGAC
P3-B9
P4-B10
P6-H2
P8-D7
P9-E6
P10-F6
P10-G6
p16-H12
P23-C8
P24-H9
P25-A6
P25-B11
P26-F4
P27-A3
P27-F3
P29-D12
P30-A3
P31-C9
P31-D9
P31-F9
P32-F5
P33-A1
P34-F9
P36-A11
P40-A2
P43-F12
P44-A5
P45-C3
P45-D3
P45-H3
P50-B3
P50-C3
P50-D11
P50-F8
p51-H6
P51-H11


970 980 990 1000 1010 1020 1030 1040 1050 1060 1070 1080
WT GCCCCGGCCGGCTCCGCGCGGGCCAAGGTCCTGCTCTTCTCCAACGCCGCCAGCCAGACCTCGCGCAGTCAGGGCACCATCCGGATGTCCTGCGACGATGGCCAGACCTGGCCGGTTTCG
P3-B9
P4-B10
P6-H2
P8-D7
P9-E6
P10-F6
P10-G6
p16-H12
P23-C8
P24-H9
P25-A6
P25-B11
P26-F4
P27-A3
P27-F3
P29-D12
P30-A3
P31-C9
P31-D9
P31-F9
P32-F5
P33-A1
P34-F9
P36-A11
P40-A2
P43-F12
P44-A5
P45-C3
P45-D3
P45-H3
P50-B3
P50-C3
P50-D11
P50-F8
p51-H6
P51-H11

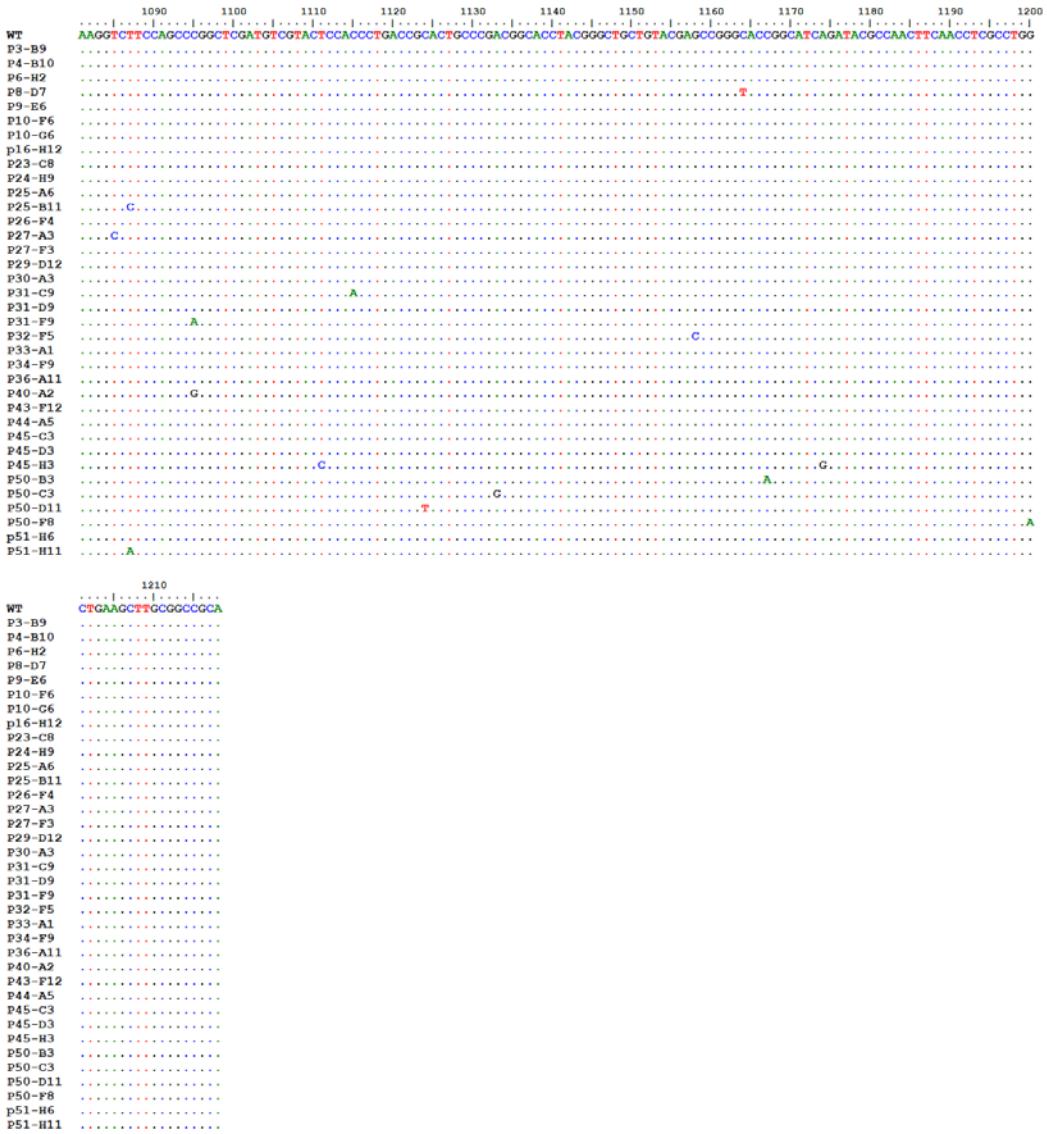


Figure 2.1 Alignment of sequences from colonies selected from the random mutagenesis library.

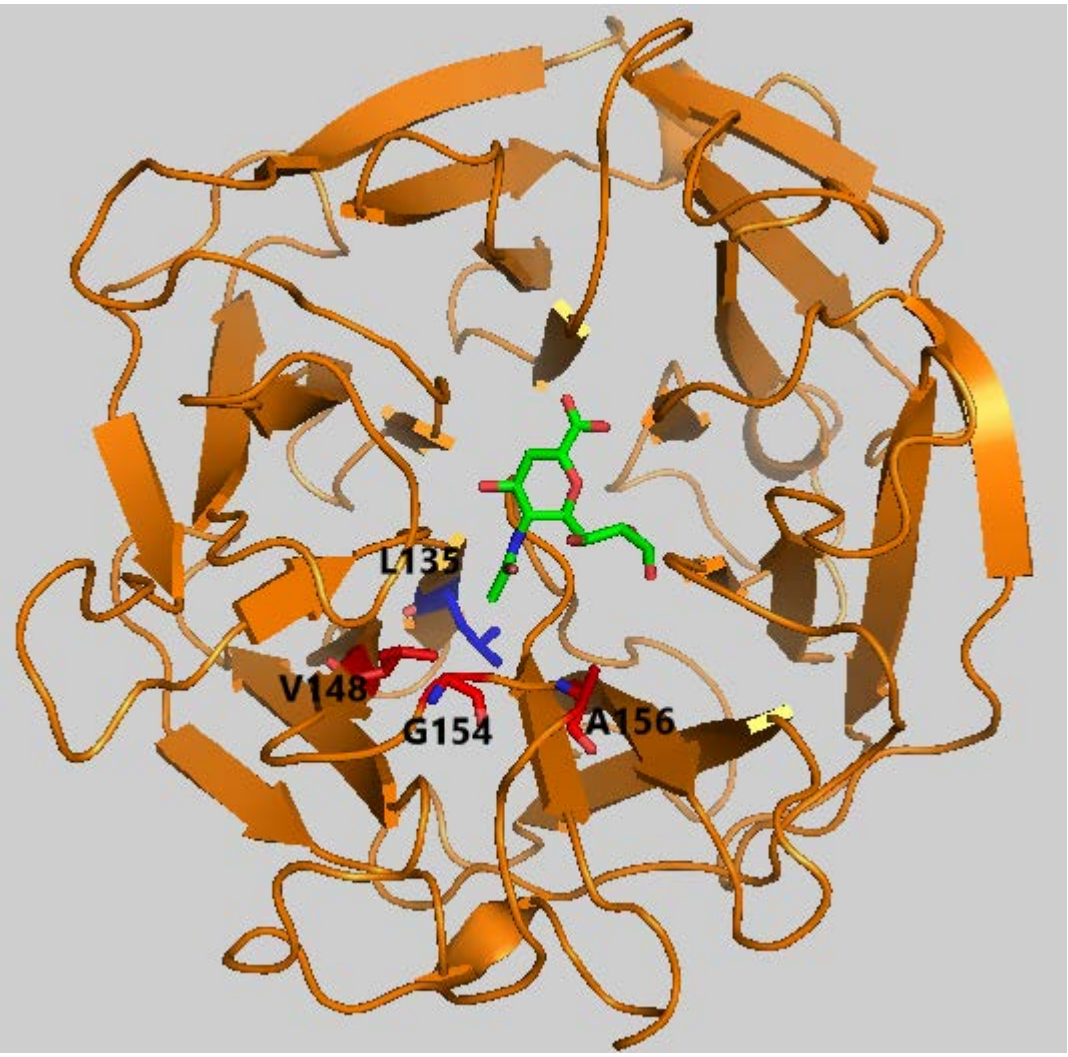


Figure 2.2 3D structure of the DANA bound *Mv*NA (PDB code 1EUS)[23], with amino acids residues Leu135, Val148, Gly154 and Ala156 labelled.

Figure 2.3 Depiction of the range of mutations in the DNA coding for selected amino acids.

Based on analysis of the sequences there is only a 15% retention of the WT sequences in these amino acids, hence we feel confident the range of mutations is sufficient for selections in the next steps. A few clones were sent for sequencing to ensure accurate cloning and presence of various amino acids variations. Figure 2.4 is depiction of the sequencing results and the corresponding amino acid substitutions.

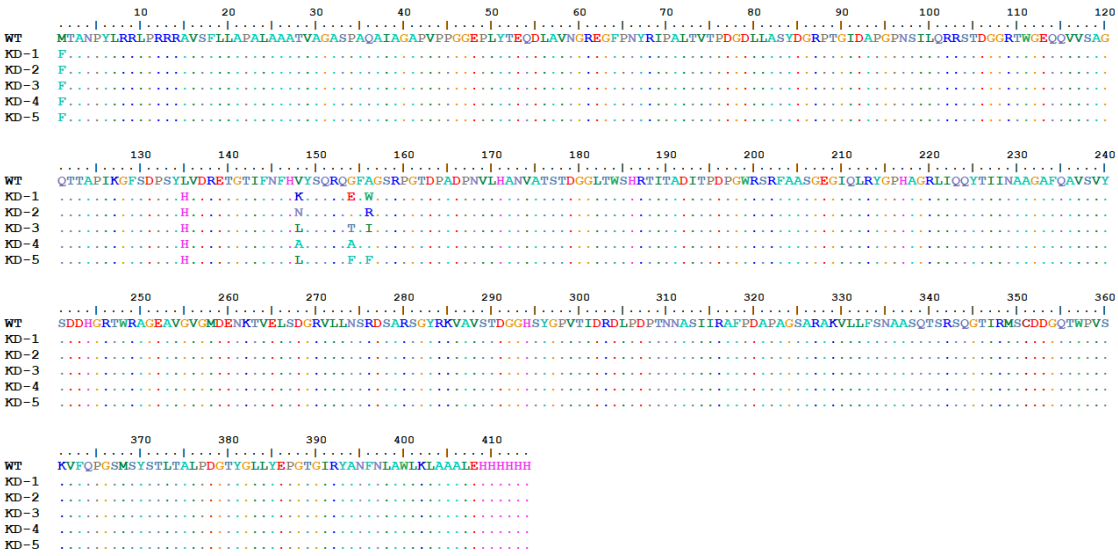


Figure 2.4 Amino acid substitutions at the sites of interest.

2.3.2. Positive and Negative Evolutionary Selection

In order to determine the concentration of chloramphenicol antibiotic that would be lethal to the tyrosine auxotroph line expressing the wildtype *Mv*NA, we inoculated MOPS minimal media[21] in a 48 well plate, supplemented with an increasing concentration of α-Neu5Ac-(2→6)-β-D-Gal-chloramphenicol with an overnight culture of the cell line with gene encoding the wildtype *Mv*NA, *nedA,* cloned into pSF-OXB19 vector and transformed into the Tyr auxotroph line. Following a 5-hour incubation at 37 ^{0}C, the absorbance of each well was measured at 600 nm and used as in indication of relative growth. Figure 2.5 shows the absorbance values corresponding to growth of the above cell line on an increasing α-Neu5Ac-(2→6)-β-D-Gal-chloramphenicol gradient. Based on this data, we concluded that 1.75 – 2.00 µg/mL of antibiotic would be lethal to our cell line carrying wild type copy the gene.

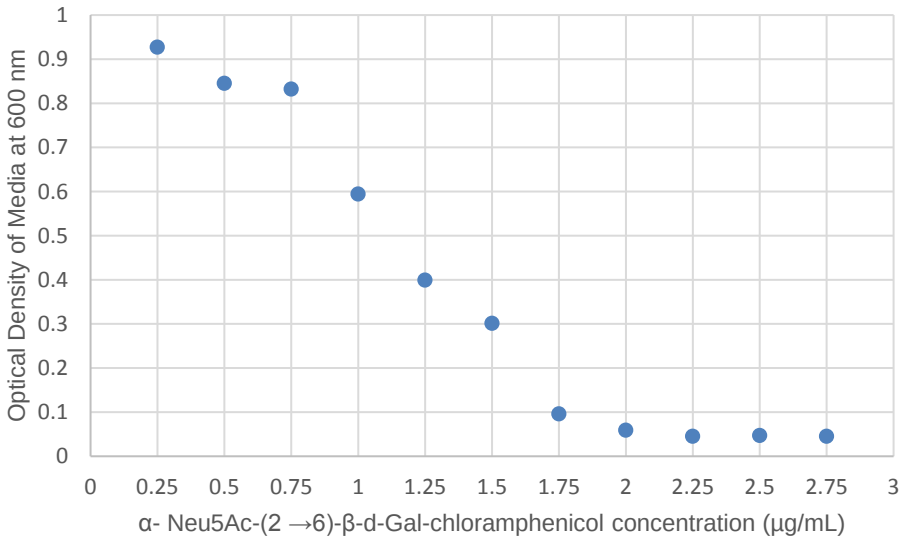


Figure 2.5 Absorbance values corresponding to growth of tyrosine auxotroph line expressing *MvNA* on an α- Neu5Ac -(2→6)-β-D-Gal-chloramphenicol gradient.

The recommended minimal amount for the addition of tyrosine to MOPS minimal media that is essential for growth of a tyrosine auxotroph line is 50 µg/mL[19,21]. However, similar to the above gradient, inoculating MOPS minimal media supplemented with an increasing range of α-Neu5Ac-(2→6)-β-D-Gal-Tyrosine following a 5-hour incubation at 37 ^{0}C showed that the minimal amount needed for significant growth is 150 µg/mL. Figure 2.6 shows the absorbance values corresponding to growth of cell expressing wild type *Mv*NA on an increasing α-Neu5Ac-(2→6)-β-D-Gal-Tyrosine gradient.

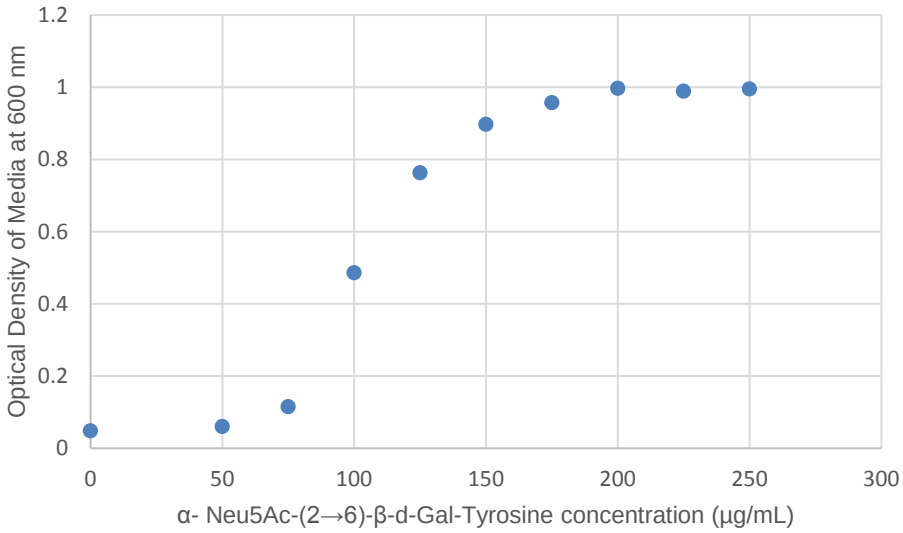


Figure 2.6 Absorbance values corresponding to growth of tyrosine auxotroph line expressing *MvNA* on an α-Neu5Ac -(2→6)-β-D-Gal-Tyrosine gradient.

In a parallel experiment, MOPS minimal media was supplemented with α-Kdn-(2→6)-β-D-Gal-Tyrosine and inoculated with MvNA to determine the level of Kdnase activity of the wildtype enzyme and as evident in data presented Figure 2.7, there was no growth seen in concentrations as high as 500 µg/mL.

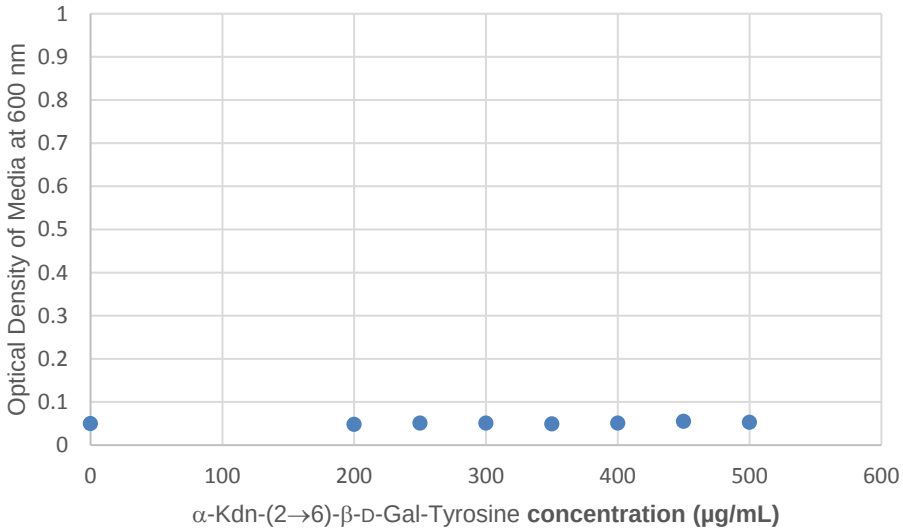


Figure 2.7 Absorbance values corresponding to growth of tyrosine auxotroph line expressing *MvNA* on an α-Kdn-(2→6)-β-D-Gal-Tyrosine gradient.

Based on these data, following the transformation of the mutagenesis library into Tyr auxotroph line, the library was plated on a selection media that contained 500 µg/mL of α-Kdn-(2→6)-β-D-Gal-Tyrosine and 2.0 µg/mL α-Neu5Ac-(2→6)-β-D-Gal-chloramphenicol. Colonies capable of growth on this media were then grown in liquid MOPS media supplemented with increasing concentration of α-Neu5Ac-(2→6)-β-D-Gal-chloramphenicol (up to 4 µg/mL) and reduced the number of colonies expressing an enzyme with reduced sialidase activity to 12.

The corresponding DNA sequence for each clone expressing a modified enzyme capable of hydrolyzing Kdn linkages was amplified and sent for sequencing. Figure 2.8 depicts the amino acid changes observed in the 12 clones compared to the gene coding for the wildtype *Mv*NA.

These 12 selected colonies were then grown on a decreasing concentration of α-Kdn-(2→6)-β-D-Gal-Tyrosine and 3 colonies, namely KDN-A11, KDN-C12 and KDN-H12 capable of growing on 350 µg/mL of α-Kdn-(2→6)-β-D-Gal-Tyrosine which indicates their increased Kdnase activity were selected.

Figure 2.8 Amino acid substitutions seen in the 12 selected clones.

Additionally, the DNA sequence for each clone was cloned into a pET28a vector and subsequently the desired protein was purified and quantified. Figure 2.9 is an SDS-PAGE depiction of the purified proteins, of which KDN-C12, KDN-A11 and KDN-H12 were selected for further kinetics studies.

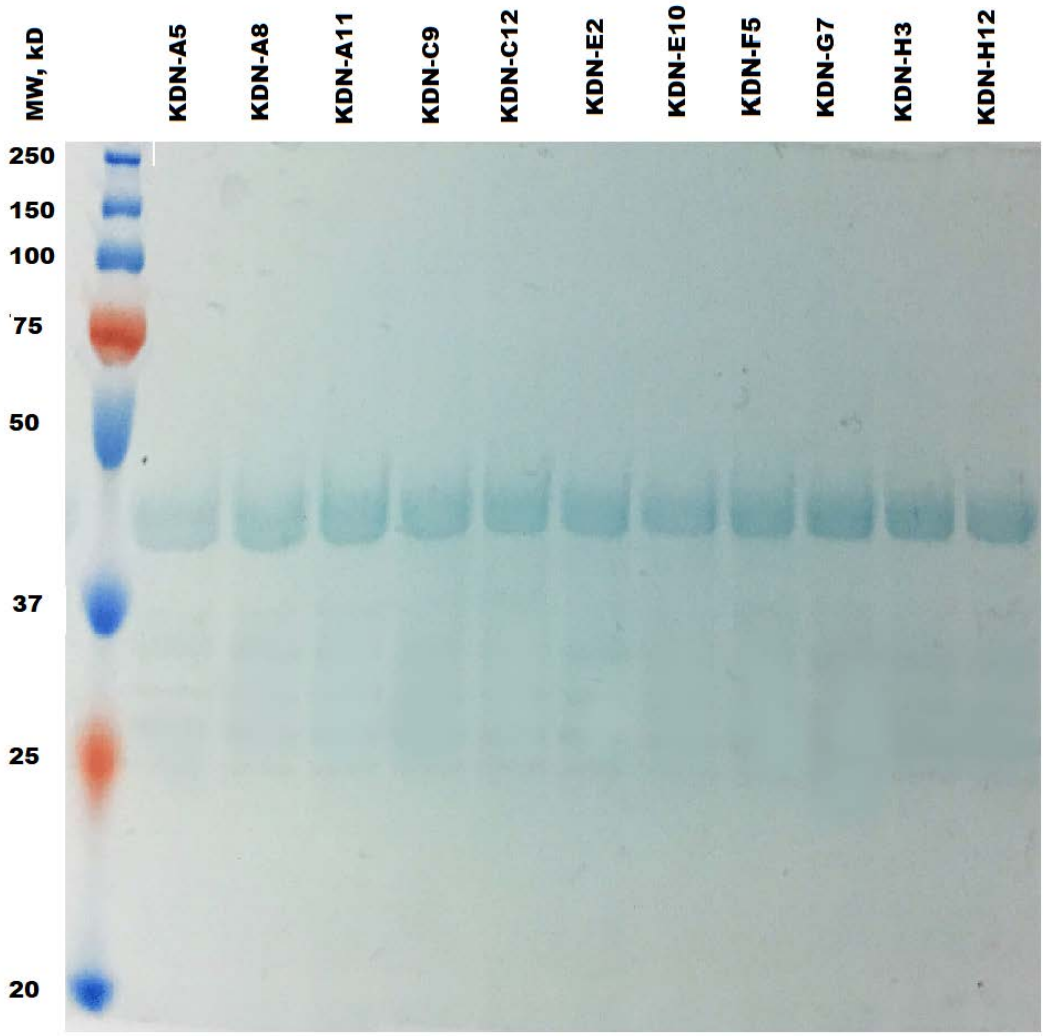


Figure 2.9 SDS-PAGE gel of the purified enzymes from selected clones.

To measure the kinetic parameters of each purified protein, initially a calibration curve of 8-FMU in MOPS buffer was constructed and this is shown in Figure 2.10.

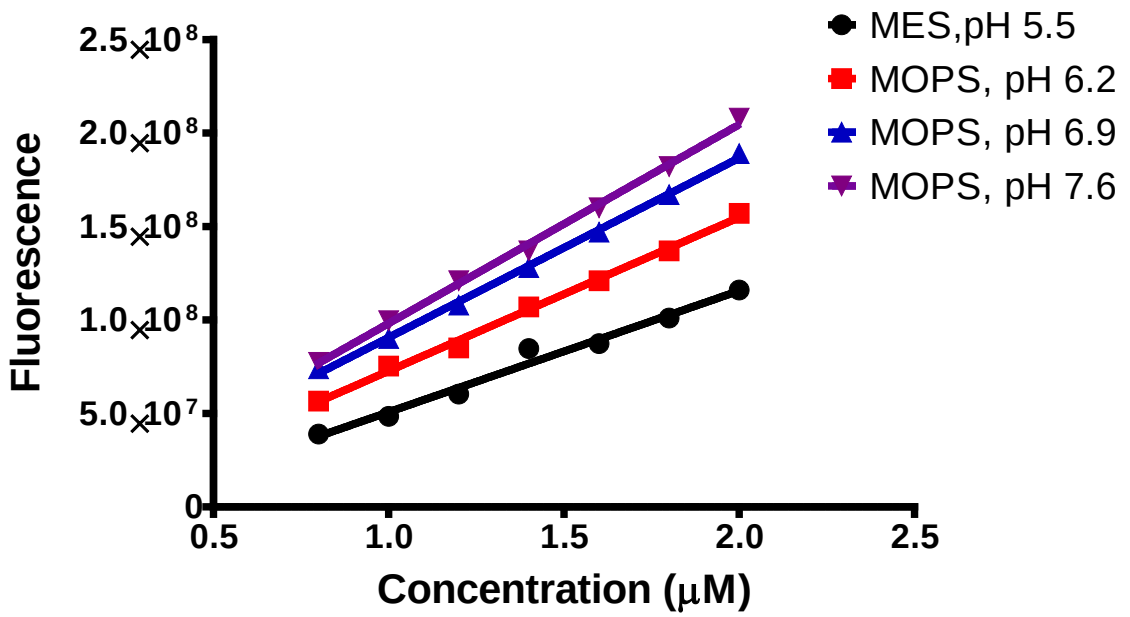


Figure 2.10 Calibration curve of 8-FMU in the range of 0.8 µM to 2.0 µM. The various pH conditions are color coded. The data was collected at λ_{ex} = 360 nm and λ_{em} = 465 nm.

The slope of each calibration curve calculated by fitting the data to a linear regression equation using the computer program prism. The unit of the calculated slopes are Fluorescence/µM. Fitted slopes are shown in Table 2.4.

Table 2.4 Slope values determined from calibration curves at various pH values.

	MES, pH 5.5	MOPS, pH 6.2	MOPS, pH 6.9	MOPS, pH 7.6
10^{-7} × Slope (Fluorescence /µM)	6.49 ± 0.40	8.22 ± 0.25	9.61 ± 0.18	10.64 ± 0.25

For each purified protein of interest, as well as the Kdnase from *Aspergillus fumigatus* (*Af*Kdnase). hydrolysis curves were constructed by plotting the Fluorescence arbitrary units, which gives an indication of the rate hydrolysis of 8FMU α-Kdn-(2→6)-β-D-Galp or 8FMU α-Neu5Ac-(2→6)-β-D-Galp by the mutant enzyme vs. time. The data was fit to a linear regression line and the slope was calculated which has the units of Fluorescence/min. An example of this data for clone KDN-C12 at pH 7.4 is shown in Figure 2.11. Dividing these values for each individual enzyme by the slope of the calibration curve

at that pH followed by division by the product of enzyme and substrate concentration, allowed for generation of k_{cat}/K_m for hydrolysis of 8FMU α-Kdn-(2→6)-β-D-Galp or 8FMU α-Neu5Ac-(2→6)-β-D-Galp by each individual enzyme. The data is summarized in Table 2.5.

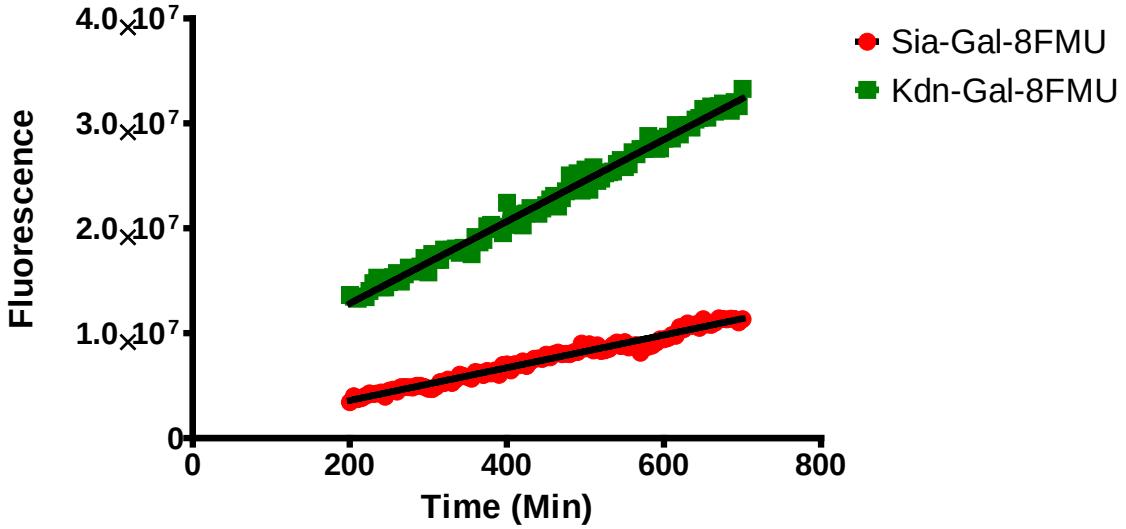


Figure 2.11 Hydrolysis curve of 8FMU α-Kdn-(2→6)-β-D-Galp and 8FMU α-Neu5Ac-(2→6)-β-D-Galp by clone KD-C12.

In order to determine the relative Kdnase activity of an individual enzyme against its sialidase activity, the k_{cat}/K_m (M^{-1} s^{-1}) values for the enzyme's Kdn linkage hydrolysis activity are divided by k_{cat}/K_m (M^{-1} s^{-1}) value for its sialidase activity. These values are shown in Table 2.6.

To determine the specificity of the purified enzymes to catalyze hydrolysis of natural Kdn glycosides linkages vs. Neu5Ac analogues relative to the WT *Mv*NA, we divided the k_{cat}/K_m (Kdn) / k_{cat}/K_m (Neu5Ac) ratio for each enzyme from Table 2.6 by the corresponding k_{cat}/K_m (Kdn) / k_{cat}/K_m (Neu5Ac) of WT enzyme at each pH. The data is summarized in Table 2.7.

Table 2.5 k_{cat}/K_m **(M^{-1} s^{-1}) for each purified enzyme.**

	$10^{-5} \times k_{cat}/K_m$ (M^{-1} s^{-1})									
	WT *Mv*NA		*Af*Kdnase		KD-C12		KD-A11		KD-H12	
	Neu5Ac	Kdn	Neu5Ac	Kdn	Neu5Ac	Kdn	Neu5Ac	Kdn	Neu5Ac	Kdn
pH 5.5	9.14 ± 0.58	0.62 ± 0.05	0.42 ± 0.04	0.64 ± 0.05	0.53 ± 0.04	0.87 ± 0.07	6.07 ± 0.46	0.80 ± 0.06	3.36 ± 0.26	0.58 ± 0.05
pH 6.2	16.0 ± 0.9	0.54 ± 0.03	0.54 ± 0.03	1.94 ± 0.10	7.53 ± 0.04	10.6 ± 0.6	9.39 ± 0.51	1.13 ± 0.06	8.16 ± 0.44	0.95 ± 0.05
pH 6.9	11.2 ± 0.5	0.48 ± 0.02	0.36 ± 0.02	0.99 ± 0.05	2.88 ± 0.14	7.21 ± 0.34	6.88 ± 0.33	1.16 ± 0.06	5.70 ± 0.27	1.24 ± 0.06
pH 7.4	8.86 ± 0.45	0.48 ± 0.02	0.34 ± 0.02	0.94 ± 0.05	4.62 ± 0.24	5.72 ± 0.28	7.18 ± 0.37	1.20 ± 0.06	4.98 ± 0.25	0.95 ± 0.05

To further probe the evolution of enzymatic activity for each enzyme, the k_{cat}/K_m value of respective Neu5c and Kdn of each enzyme is divided by the corresponding WT *Mv*NA k_{cat}/K_m values at each pH. The data is shown in Table 2.8.

Table 2.6 The ratio of the Kdnase vs. sialidase activity of each individual enzyme is calculated by taking the ratio of the k_{cat}/K_m (Kdn) / k_{cat}/K_m (Neu5Ac).

	k_{cat}/K_m (Kdn) / k_{cat}/K_m (Neu5Ac)				
	WT MvNA	AfKdnase	KD-C12	KD-A11	KD-H12
pH 5.5	0.067 ± 0.007	1.53 ± 0.19	0.17 ± 0.02	0.13 ± 0.01	0.17 ± 0.02
pH 6.2	0.034 ± 0.003	3.57 ± 0.29	1.41 ± 0.11	0.12 ± 0.01	0.12 ± 0.01
pH 6.9	0.043 ± 0.003	2.75 ± 0.23	2.51 ± 0.17	0.17 ± 0.01	0.22 ± 0.01
pH 7.4	0.055 ± 0.004	2.78 ± 0.24	1.24 ± 0.09	0.17 ± 0.01	0.19 ± 0.01

Table 2.7 **Specificity Values: The k_{cat}/K_m (Kdn) / k_{cat}/K_m (Neu5Ac) ratio for each enzyme is divided by the the corresponding k_{cat}/K_m (Kdn) / k_{cat}/K_m (Neu5Ac) of WT enzyme at each pH to determine the specificity of each enzyme in hydrolysis of Kdn linkages in comparison to the WT *Mv*NA.**

	Ratio of k_{cat}/K_m (Kdn) / k_{cat}/K_m (Neu5Ac) for each enzyme over k_{cat}/K_m (Kdn) / k_{cat}/K_m (Neu5Ac) of WT *Mv*NA		
	KD-C12	KD-A11	KD-H12
pH 5.5	2.48 ± 0.37	1.97 ± 0.29	2.57 ± 0.39
pH 6.2	41.9 ± 4.5	3.57 ± 0.39	3.44 ± 0.37
pH 6.9	58.5 ± 5.6	3.92 ± 0.38	5.08 ± 0.49
pH 7.4	22.6 ± 2.3	3.06 ± 0.31	3.48 ± 0.35

Table 2.8 **Evolution of enzymatic activity. The k_{cat}/K_m value of respective Neu5c and Kdn of each enzyme is divided by the corresponding WT *Mv*NA k_{cat}/K_m values at each pH.**

	Ratio of k_{cat}/K_m (Individual enzyme) / k_{cat}/K_m (WT *Mv*NA)					
	KD-C12		KD-A11		KD-H12	
	Neu5Ac	Kdn	Neu5Ac	Kdn	Neu5Ac	Kdn
pH 5.5	0.57 ± 0.06	1.42 ± 0.16	0.66 ± 0.07	1.30 ± 0.14	0.37 ± 0.04	0.94 ± 0.11
pH 6.2	0.47 ± 0.04	19.7 ± 1.5	0.59 ± 0.04	2.09 ± 0.16	0.51 ± 0.04	1.75 ± 0.13
pH 6.9	0.26 ± 0.02	15.2 ± 1.0	0.61 ± 0.04	2.41 ± 0.16	0.51 ± 0.03	2.58 ± 0.18
pH 7.4	0.52 ± 0.04	11.8 ± 0.8	0.81 ± 0.06	2.48 ± 0.18	0.56 ± 0.04	1.96 ± 0.14

2.4. Discussion

Sialic acids, with over 50 naturally occurring derivatives, comprise the most chemically and structurally diverse carbohydrate family. Neu5Ac and Kdn, two major members of this family, differ only in the C5 position with Kdn and Neu5Ac having OH and NH groups respectively. Sialidases, or neuraminidases are enzymes that cleave the terminal sialic acid from a variety of glycoconjugates and as such, play an important and often essential role in pathogenesis, bacterial nutrition, and cellular interactions[24].

Given the variety of sialic acids present in nature and their importance, evolution of sialidases, that would recognize and cleave specific types of sialic acid family members would be beneficial in elucidating the role of specific sialic acids and allow for further probe into the effect of various substitutions on their role and function in cellular pathways.

We aimed to evolve a Kdnase, or a sialidase that would recognize and cleave Kdn glycosidic linkages from the *Micromonospora viridifaciens's* sialidase (*Mv*NA) which is a catalytically efficient sialidase. That is, we choose this enzyme because we expected a drastic decrease in activity during early evolution steps and we reasoned that a robust enzyme would be more suitable for our downstream purposes which is reduction of neuraminidase activity and enhancement of Kdnase activity. This enzyme also has very poor affinity in hydrolysing Kdn linkages and more importantly is remarkably tolerant of mutagenesis in the active site residues. To this end, we generated a random mutagenesis library that allowed for identification of a Leu135His mutation that likely plays a role in affording *Mv*NA, Kdnase activity. Taking advantage of the crystal structure of *Mv*NA, we identified three amino acids, namely Val148, Gly154 and Ala156 that would likely play a role in recognition of Kdn vs. Neu5Ac. We generated saturation site directed mutagenesis library of these amino acids and through positive and negative evolutionary selection protocols were able to identify and purify 3 independent modified *Mv*NAs that possess Kdnase activity.

From the sequencing data it is clear that there is a bias retained in the genetic library toward WT retention of the targeted amino acids with Val148 and Gly154 and Ala156 being approximately 48%, 61% and 37% respectively. The sequence analysis is shown in Table 2.9 and 2.10. Despite this observed bias, there is only an 18% retention of the WT amino acids in the sequenced reads. From these 69622 sequence reads, we identified over 3000 different combination of the three amino acids that were targeted. Of note, no read contained either a Cys or Trp.

Deep sequencing data gives an idea of the range of mutations present in the library, however given that one of the active clones has both a cysteine and tryptophan means that the deep sequencing hasn't picked up all low abundance sequences. In the next round of selection procedures, we would like to explore the replacement of the

targeted residues to ensure a better coverage of cysteine and tryptophan as the presence of these amino acids seems to be conducive to reduction of sialidase activity.

Table 2.9 **The percentage of each amino acid from the deep sequencing data is shown. The amino acid replacements observed in 3 selected clones are color coded. WT *Mv*NA(underlined), KDN-C12(purple), KDN-A11(green) and KDN-H12(blue).**

	Amino Acid percentage (%) present in deep sequencing data										
WT	**Ala**	**Cys**	**Asp**	**Glu**	**Phe**	**Gly**	**His**	**Iso**	**Lys**	**Leu**	**Met**
V148	**2.26**	**0.00**	**1.76**	**2.07**	**0.73**	**2.54**	**1.57**	**1.55**	**6.98**	**3.03**	**1.63**
G154	**3.80**	**0.00**	**1.13**	**2.08**	**5.49**	**61.07**	**0.31**	**0.59**	**0.38**	**2.63**	**0.48**
A156	**36.76**	**0.00**	**1.82**	**2.15**	**7.34**	**17.48**	**0.37**	**0.94**	**0.42**	**4.30**	**0.77**

Table 2.10 **The percentage of each amino acid from the deep sequencing data is shown. The amino acid replacements observed in 3 selected clones are color coded. WT *Mv*NA(underlined), KDN-C12(purple), KDN-A11(green) and KDN-H12(blue).**

	Amino Acid percentage (%) present in deep sequencing data										
WT	**Asn**	**Pro**	**Gln**	**Arg**	**Ser**	**Thr**	**Val**	**Trp**	**Tyr**	**STOP**	**Total**
V148	**5.20**	**5.17**	**1.92**	**3.27**	**4.64**	**4.80**	**47.54**	**0.00**	**1.27**	**2.06**	**69622**

G154	0.27	0.61	0.61	4.09	3.04	0.87	5.92	0.00	0.89	5.73	69622
A156	0.31	1.57	0.48	4.35	3.47	1.38	12.11	0.00	0.60	3.37	69622

From the positive and negative evolutionary screens, we were able to identify 12 clones that possessed both a decreased sialidase activity and an increased Kdnase activity. All twelve clones were expressed and purified, except for clone B7 which resisted purification effort, with the expected size of ~44 kD as shown in Figure 2.9.

In order to measure the kinetics parameters of the modified proteins, we set up an assay using natural substrate analogues, namely 8FMU α-Kdn-(2→6)-β-D-Galp and 8FMU α-Neu5Ac-(2→6)-β-D-Galp. Using data from the modified protein assays and the calibration curves shown in Figure 2.10, we calculated the second-order rate constant (k_{cat}/K_m) and the respective Kdn/Neu5Ac ratios, so as to further analyze the specificity and selectivity of the purified proteins. The Km for WT *Mv*NA with a natural sialoside analogue containing a α-2,6 sialyl–galactosyl linkages, as well as an attached 4-methylumbelliferone was reported to be 240 μM[25]. Given the difficulty of synthesizing our analogues, we were unable to measure the Km for each of our selected clones. However, given that we performed the continues assay at a substrate concentration far below the Km, the calculated values are indeed second-order rate constants.

As shown in Table 2.5, the measured k_{cat}/K_m values for WT *Mv*NA are similar to those reported in literature[25] for hydrolysis of MU α-Neu5Ac-(2→6)-β-D-Galp. This in turn lends credibility to our protocol and as a result, we conclude that we have a sensitive and accurate assay to measure the kinetics parameters of the modified enzymes.

As shown in Table 2.6, WT *Mv*NA has a 15 to 30-fold higher selectivity for the acetamido group at C5 position present in Neu5Ac relative to C5-OH of Kdn.

The selectivity of the *Aspergillus fumigatus* (*Af*Kdnase) for hydrolyzing Kdn glycoside linkages over Neu5Ac glycoside linkages that we measured in our assays, is lower than the reported values from the literature[26]. However, it should be noted that the literature values report on activated substrates, MU-Kdn and MU-Neu5Ac were used while in our study we take advantage of compounds containing natural glycoside linkages. Additionally, the natural substrate for *Af*Kdnase, based on the available crystal structure and the presence of two Kdn binding sites, has been proposed to be poly-Kdn which could potentially translate to lower selectivity of this enzyme when cleaving our natural substrate analogues.

Based on the sequencing and kinetic data, the three (3) selected mutants, fall into two categories. First, where in clones KDN-A11 and KDN-H12 there is a significant reduction in the sialidase activity with a small increase in the Kdnase activity. Second, in clone KDN-C12, there is a marked increase in the Kdnase activity especially at a pH of 6.2 and higher as shown in Tables 2.6 and 2.7. While the Kdnase activity of clone KDN-C12 is increased by 20 fold at higher pH, unfortunately, the decrease in sialidase activity is not as great as we had hoped. Thus, the next round of mutagenesis and selection protocols should be with the aim of reducing the sialidase activity. The dependency KDN-C12 enzyme's k_{cat}/K_m on pH is unexpected considering the WT *Mv*NA's k_{cat}/K_m values shows no dependency on pH in the range of 5.2 to 7.4. This observed pH dependency of modified enzymes is likely a consequence of the selection conditions where the clones were grown and selected for in media with a pH of close to 7.

Surprisingly, the highest level of Kdnase activity occurs in the instance where a G154A and A156V replacement along with the WT V148 results in presence of three hydrophobic amino acids which creates a smaller binding pocket more suited for the smaller OH group of Kdn as opposed to bulkier acetamido group of Neu5Ac.

The KDN-A11 mutant, remarkably retains high sialidase activity even though modification of V148P, G154C and A156W would suggest a drastic reduction in the size of the C5 pocket. This retention of sialidase activity likely is due to the side chains of the modified amino acids, namely A156W pointing outward of the C5 pocket, similar to A156 methyl group facing away from the binding pocket as seen in the crystal structure of WT

*Mv*NA bound to DANA shown in Figure 2.12. Confirmation of this awaits the crystal structure of the modified enzyme. Additionally, next round of mutations should be targeting the amino acids between G154 and A156 to see the effect of replacement of these amino acids and their respective sidechains orientation relative to the binding pocket.

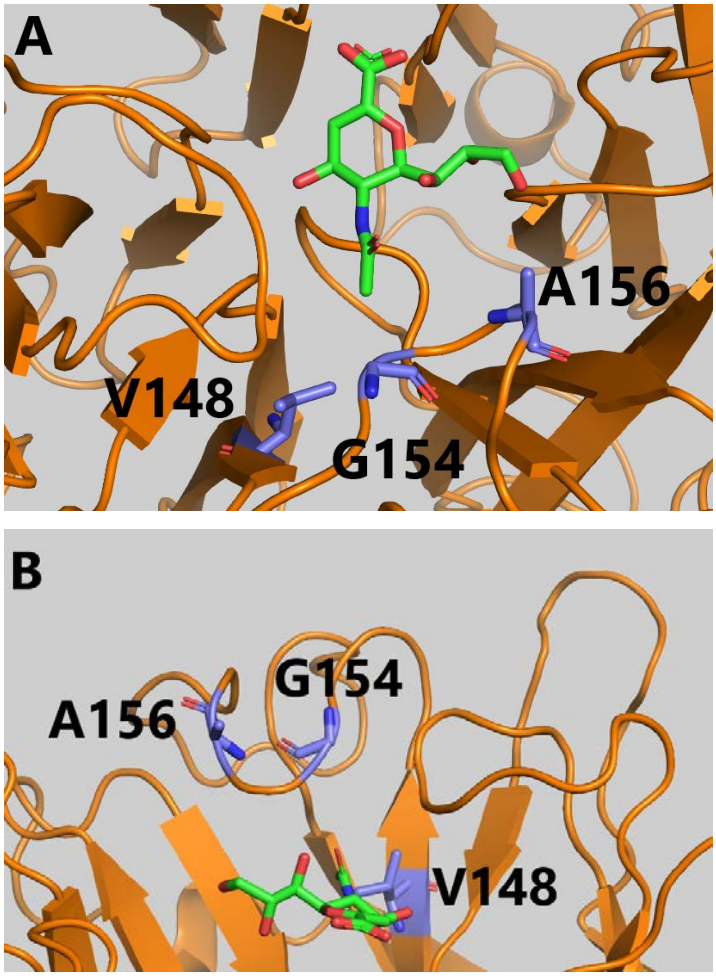


Figure 2.12 Crystal structure of MvNA bound by DANA (PDB code 1EUS)[23]. The amino acids targeted in saturation mutagenesis library are labelled. A) top view b) side view

It should also be noted that all our kinetic studies are done with α2-6 linked sialic acids and we would like to like to also test our mutagenesis library against α2-3 linked sialic acid natural analogues.

2.5. References

1. Severi E, Hood DW, Thomas GH. Sialic acid utilization by bacterial pathogens. *Microbiology*. 2007;153(Pt 9):2817-2822. doi: 153/9/2817 [pii].

2. Vimr ER, Kalivoda KA, Deszo EL, Steenbergen SM. Diversity of microbial sialic acid metabolism. *Microbiol Mol Biol Rev*. 2004;68(1):132-153.

3. Alviano CS, Travassos LR, Schauer R. Sialic acids in fungi: A minireview. *Glycoconj J*. 1999;16(9):545-554.

4. Friedrich N, Matthews S, Soldati-Favre D. Sialic acids: Key determinants for invasion by the *Apicomplexa*. *Int J Parasitol*. 2010;40(10):1145-1154. doi: 10.1016/j.ijpara.2010.04.007 [doi].

5. Heikema AP, Bergman MP, Richards H, et al. Characterization of the specific interaction between sialoadhesin and sialylated *Campylobacter jejuni* lipooligosaccharides. *Infect Immun*. 2010;78(7):3237-3246. doi: 10.1128/IAI.01273-09 [doi].

6. Nadano D, Iwasaki M, Endo S, Kitajima K, Inoue S, Inoue Y. A naturally occurring deaminated neuraminic acid, 3-deoxy-D-glycero-D-galacto-nonulosonic acid (KDN). its unique occurrence at the nonreducing ends of oligosialyl chains in polysialoglycoprotein of rainbow trout eggs. *J Biol Chem*. 1986;261(25):11550-11557.

7. Song Y, Kitajima K, Inoue S, Inoue Y. Isolation and structural elucidation of a novel type of ganglioside, deaminated neuraminic acid (KDN)-containing glycosphingolipid, from rainbow trout sperm. the first example of the natural occurrence of KDN-ganglioside, (KDN)GM3. *J Biol Chem*. 1991;266(32):21929-21935.

8. Angata T, Kitajima K, Inoue S, et al. Identification, developmental expression and tissue distribution of deaminoneuraminate hydrolase (KDNase) activity in rainbow trout. *Glycobiology*. 1994;4(4):517-523.

9. Li YT, Yuziuk JA, Li SC, et al. A novel sialidase capable of cleaving 3-deoxy-D-glycero-D-galacto-2-nonulosonic acid (KDN). *Arch Biochem Biophys*. 1994;310(1):243-246. doi: S0003-9861(84)71163-5 [pii].

10. Kanamori A, Inoue S, Xulei Z, et al. Monoclonal antibody specific for alpha 2-->8-linked oligo deaminated neuraminic acid (KDN) sequences in glycoproteins. preparation and characterization of a monoclonal antibody and its application in immunohistochemistry. *Histochemistry*. 1994;101(5):333-340.

11. Inoue S, Kitajima K. KDN (deaminated neuraminic acid): Dreamful past and exciting future of the newest member of the sialic acid family. *Glycoconj J*. 2006;23(5-6):277-290. doi: 10.1007/s10719-006-6484-y [doi].

12. Inoue S, Kanamori A, Kitajima K, Inoue Y. KDN-glycoprotein: A novel deaminated neuraminic acid-rich glycoprotein isolated from vitelline envelope of rainbow trout eggs. *Biochem Biophys Res Commun*. 1988;153(1):172-176. doi: S0006-291X(88)81204-X [pii].

13. Knirel YA, Kocharova NA, Shashkov AS, Kochetkov NK, Mamontova VA, Soloveva TF. Structure of the capsular polysaccharide of *Klebsiella ozaenae* serotype K4 containing 3-deoxy-D-glycero-D-galacto-nonulosonic acid. *Carbohydr Res*. 1989;188:145-155.

14. Watts AG, Damager I, Amaya ML, et al. *Trypanosoma cruzi* trans-sialidase operates through a covalent sialyl-enzyme intermediate: Tyrosine is the catalytic nucleophile. *J Am Chem Soc*. 2003;125(25):7532-7533. doi: 10.1021/ja0344967 [doi].

15. Karwaski MF, Wakarchuk WW, Gilbert M. High-level expression of recombinant *Neisseria* CMP-sialic acid synthetase in *Escherichia coli*. *Protein Expr Purif*. 2002;25(2):237-240. doi: S1046592802000049 [pii].

16. Kakuta Y, Okino N, Kajiwara H, et al. Crystal structure of *Vibrionaceae photobacterium* sp. JT-ISH-224 alpha2,6-sialyltransferase in a ternary complex with donor product CMP and acceptor substrate lactose: Catalytic mechanism and substrate recognition. *Glycobiology*. 2008;18(1):66-73. doi: cwm119 [pii].

17. Chan J, Lu A, Bennet AJ. Turnover is rate-limited by deglycosylation for *Micromonospora viridifaciens* sialidase-catalyzed hydrolyses: Conformational implications for the michaelis complex. *J Am Chem Soc*. 2011;133(9):2989-2997. doi: 10.1021/ja109199p [doi].

18. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*. 1989;77(1):51-59. doi: 0378-1119(89)90358-2 [pii].

19. Baba T, Ara T, Hasegawa M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The keio collection. *Mol Syst Biol*. 2006;2:2006.0008. doi: msb4100050 [pii].

20. Untergasser A. Preparation of chemical competent cells. *Untergasser's Lab.* 2008.

21. Neidhardt FC, Bloch PL, Smith DF. Culture medium for enterobacteria. *J Bacteriol*. 1974;119(3):736-747.

22. Watson JN, Indurugalla D, Cheng LL, Narine AA, Bennet AJ. The hydrolase and transferase activity of an inverting mutant sialidase using non-natural beta-sialoside substrates. *Biochemistry*. 2006;45(44):13264-13275. doi: 10.1021/bi061489x [doi].

23. Gaskell A, Crennell S, Taylor G. The three domains of a bacterial sialidase: A beta-propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure*. 1995;3(11):1197-1205.

24. Varki A, Cummings RD, Esko JD. *Essentials of glycobiology.* 2nd edition ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009.

25. Indurugalla D, Watson JN, Bennet AJ. Natural sialoside analogues for the determination of enzymatic rate constants. *Org Biomol Chem*. 2006;4(24):4453-4459. doi: 10.1039/b613909d [doi].

26. Telford JC, Yeung JH, Xu G, et al. The *Aspergillus fumigatus* sialidase is a 3-deoxy-D-glycero-D-galacto-2-nonulosonic acid hydrolase (KDNase): Structural and mechanistic insights. *J Biol Chem*. 2011;286(12):10783-10792. doi: 10.1074/jbc.M110.207043 [doi].

Chapter 3. A New Class of Glycoside Hydrolase Mechanism-Based Covalent Inhibitors: Glycosylation Transition State Conformations

Saeideh Shamsi Kazem Abadi,[1] Michael Tran,[2] Anuj Yadav,[2] Pal John Pal Adabala,[2] Saswati Chakladar,[2] and Andrew J. Bennet[2,*]

Departments of Biochemistry and Molecular Biology,[1] and Chemistry[2] Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, V5A 1S6, Canada

E-mail:bennet@sfu.ca

3.1. Abstract

The design of covalent inhibitors in glycoscience research remains an important challenge for the development of chemical tools. Here we report the synthesis of a new carbocyclic mechanism-based covalent inhibitor of an α-glucosidase (glycoside hydrolase family 13). The enzyme efficiently catalyzes its alkylation via either an allylic cation or a cationic TS. We show that this allylic covalent inhibitor has very different catalytic proficiencies for pseudo-glycosylation and deglycosylations. Our carbasugar analogues have the potential to be useful chemical biology tools.

3.2. Body

The catalytic transfer of carbohydrate moieties frequently involves anomeric positive charge delocalization by the endocyclic oxygen atom at the enzymatic transition state(s); we show for the first time that an alkene can perform the same task. Of note, such enzymes, which catalyze either addition or removal of carbohydrate residues, are often critical components of cellular regulation involving complex interactions between glycoconjugates and various biological receptors.[1-3] The enzymes that hydrolytically remove sugar residues by cleavage of glycosidic bonds are called glycoside hydrolases (GHs).[4] Nature has evolved several catalytic strategies for the modes of action for GHs. That being said, most GHs catalyze hydrolysis by using a pair of active site aspartic and/or glutamic acid (Asp/Glu) residues,[5,6] with the distance between the two carboxylic acids being a determining factor in whether the reaction occurs with retention or inversion of anomeric configuration.[7,8]

Retaining glycoside hydrolases that have two catalytic Asp/Glu residues operate via two sequential inversions of configuration at the anomeric center. The first results in the formation of a covalent glycosyl-enzyme intermediate (Figure 1) and the second, not shown, involves hydrolysis of this intermediate. Both the glycosylation and deglycosylation transition states (TSs) for pyranoside hydrolysis possess pyranosylium ion-like character,

and the six-membered ring adopts one of several allowed conformations.[6,9] In the current example, retaining α-glucoside hydrolases (α-glucosidases from GH13)[10] react via pyranosylium ion-like TSs that are traversed during the catalytic cycle. Moreover, the pyran ring conformation at the TSs is postulated to be close to a $^4H_3$ half-chair.[11] Also, for GH13 enzymes the structure of the enzyme bound intermediate is a $^4C_1$ chair (Figure 1).[12] Kinetic isotope effect data suggests that the productive Michaelis complex is not a ground state chair conformation, but is likely a $^1S_3$ skew boat (Figure 1).[13,14]

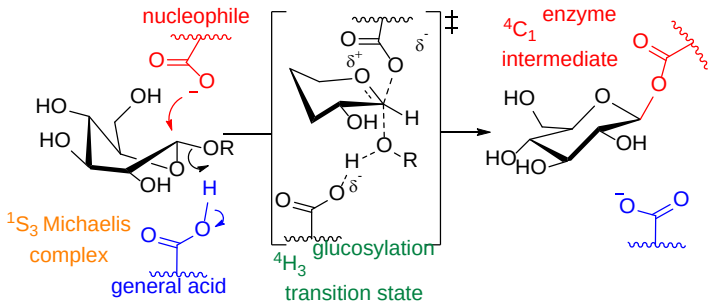


Figure 1. Proposed mechanism of glucosylation for a GH13 family retaining α-glucosidase; for clarity, some hydroxyl groups are not shown for the transition state. Conformations are shown for the Michaelis complex ($^1S_3$), the glucopyranosylium ion-like transition state ($^4H_3$) and the glucosylated enzyme-bound intermediate ($^4C_1$).

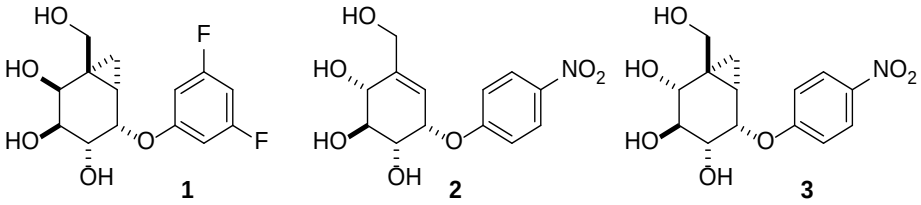
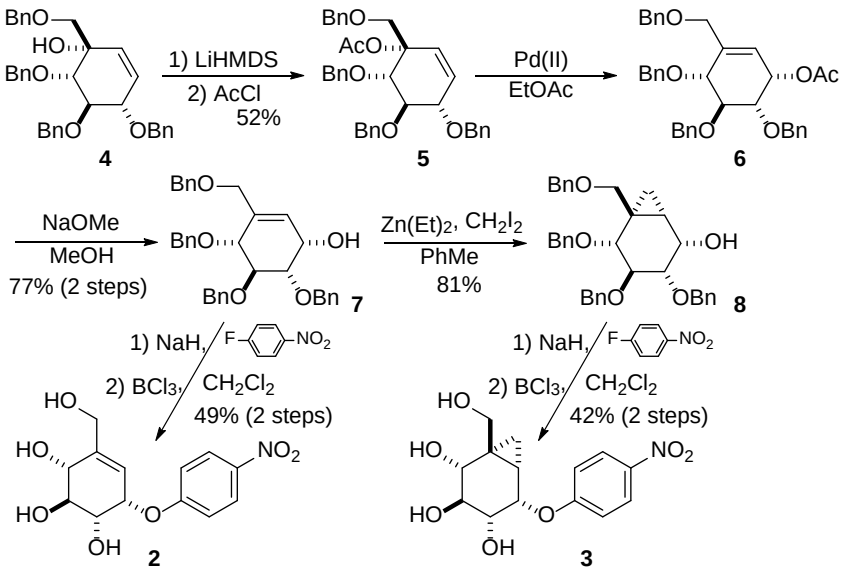


Figure 2. Structures of the mechanism-based covalent inhibitor (1) for a GH36 family α-D-galactosidase from *Thermotoga maritima*, and the GH13 inhibitors (2 and 3) that were evaluated as covalent inhibitors in the current study.

Previously, we have utilized the requirement for transition state positive charge delocalization that is a fundamental factor in catalysis by most GHs in our design of a cyclopropyl-containing mechanism-based covalent inhibitor (**1**) of an α-D-galactosidase from *Thermotoga maritima*, a GH36 retaining enzyme.[15,16] Here we describe the synthesis of two carbocyclic analogues of D-glucose (**2** and **3**) that are covalent inhibitors of a GH13

retaining α-glucosidase. We show that these compounds lead to a single covalent labeling of the enzyme, and importantly the rate constants for 'pseudo'-glycosylation and deglycosylation for these two covalent inhibitors are distinct and provide insight into the conformational itinerary for this family of GH enzymes.

We synthesized **2** and **3** (Scheme 1) from **4**, which we made in four steps (42% yield) from commercially available 2,3,4,6-tetra-*O*-benzyl-D-glucopyranose according to the method of Kapferer et al.[17] First, we acetylated the tertiary alcohol in **4** to give **5**, which smoothly underwent a palladium-catalyzed [3,3]-sigmatropic rearrangement to give the pseudo anomeric acetate **6**.[18] Standard deacetylation conditions gave allylic alcohol **7**, which was subjected to a Furukawa modified Simmons-Smith cyclopropanation to give **8**. Both **7** and **8** underwent facile S_NAr reactions followed by global debenzylation with BCl_3 to give covalent inhibitors **2** and **3**, respectively.



Scheme 1. Synthesis of carbocyclic inhibitors 2 and 3.

We then tested carbasugar analogues **2** and **3** for their activity against commercially available yeast α-glucosidase. Shown in Figure 3 are the measured pseudo-first-order rate constants for the loss of enzyme activity as a function of the concentration of the carbasugar analogue. Notably, the allylic

inhibitor **2** (red circles) is less active that the bicyclo[4.1.0]heptyl inactivator **3** (blue circles).

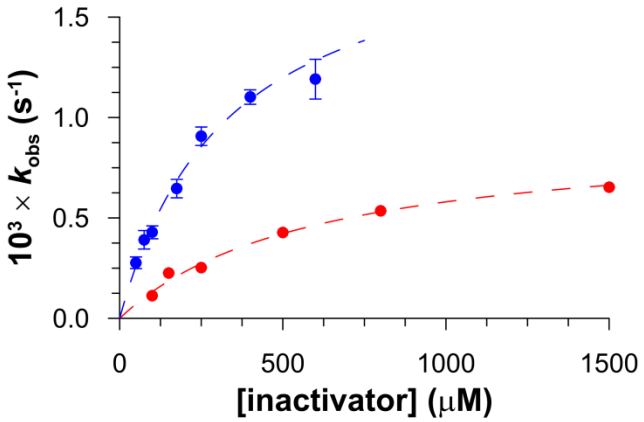


Figure 3. Inactivation kinetics for covalent inhibition of yeast GH13 α-glucosidase. Shown in red circles are the data for covalent labeling by the allylic inhibitor 2, while the blue circles represent labeling by the bicyclic inhibitor 3. Error bars that are not visible are encompassed within the symbol. Dashed lines represent the best non-linear fits to a standard Michaelis–Menten equation. Conditions for all experiments were T=25 °C in sodium phosphate buffer (50 mM, pH 6.84, [BSA] = 1 mg/ mL).

We then measured the rate constants for the reactivation of the covalently-modified α-glucosidase after removal of excess inactivator (Figure 4). Remarkably, the measured first-order rate constant for regaining of enzyme activity following inhibition by allylic inhibitor **2** (red circles) is more rapid than that for bicyclo[4.1.0]heptyl inactivated enzyme (blue circles). The kinetic data were fit to the standard kinetic scheme for covalent inhibition (Scheme 2) and the derived rate and equilibrium constants are tabulated in Table 1.

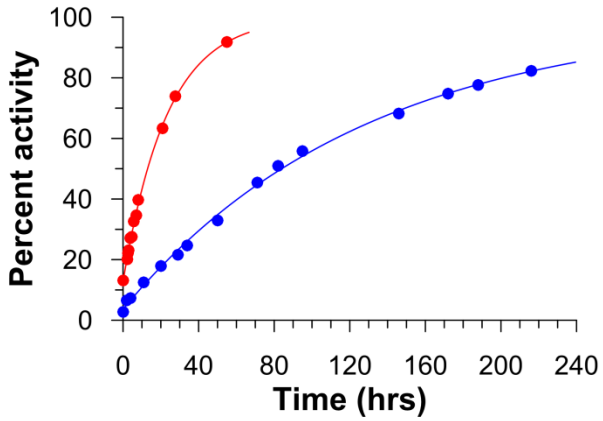


Figure 4. Reactivation kinetics for covalently inhibited yeast GH13 α-glucosidase. Shown in red circles are the data for the dealkylation of enzyme that had been labeled by allylic inhibitor 2, while the blue circles represent α-glucosidase reactivation previously treated with the bicyclic inhibitor 3. The solid lines represent the best non-linear fits to a standard first-order rate equation. Conditions for all experiments were T=25 °C in sodium phosphate buffer (50 mM, pH 6.84, [BSA] = 1 mg/ mL).

$$E + I \underset{K_i}{\rightleftharpoons} E{:}I \xrightarrow{k_{inact}} E\text{-}I \xrightarrow{k_{react}} E + P$$

Scheme 2. Kinetic scheme for the covalent inhibition of GH13 yeast α-glucosidase by carbasugar analogues 2 and 3.

We next tried to identify the sites of labeling. Specifically, we incubated the yeast enzyme with excess inactivators and tried to analyze, by ESI tandem mass spectrometry (MS/MS), the tryptic (and peptic) peptides obtained by digestion of both the inactivated and the untreated enzymes. Unfortunately, we were unable to obtain satisfactory peptide fragmentation that remained covalently modified after tryptic digestion. However, we successfully demonstrated that yeast GH13 α-glucosidase is singly labeled by the expected mass addition of the carbon skeleton portions of **2** and **3** to the molecular weight of the enzyme (Figure S1 Supporting Information). That is, the mass spectrum of intact enzyme shows a single peak for the native enzyme at 67275.7, while that after reaction with **2** shows the intact enzyme and a mono-alkylated species ($C_7H_{10}O_4$ = 158.1) at 67433.9, and the corresponding mass spectrum for the enzyme covalently-modified by **3**

displays a single peak at 67448.8, which corresponds to addition of the carbocyclic skeleton of **3** ($C_8H_{12}O_4$ = 172.1 + H).

Table 1. Kinetic parameters for the covalent inhibition and reactivation of yeast α-glucosidase by the allylic and bicyclo[4.1.0]heptyl compounds 2 and 3. Conditions for all experiments were T=25 °C in sodium phosphate buffer (50 mM, pH 6.84, [BSA] = 1 mg/ml).

Inactivator	K_I (μM)	k_{inact} (s^{-1})	$t_{½}$ (mins)	k_{inact}/K_I (M^{-1} s^{-1})	k_{react} (s^{-1})	$t_{½}$ (hrs)
2	570 ± 90	$(9.05 \pm 0.63) \times 10^{-4}$	12.8	1.59 ± 0.27	$(1.19 \pm 0.07) \times 10^{-5}$	16.3
3	285 ± 45	$(1.82 \pm 0.14) \times 10^{-3}$	6.3	6.4 ± 1.1	$(2.17 \pm 0.14) \times 10^{-6}$	88.7

The kinetic data listed in Table 1 show two remarkable features: (i) both the first- (k_{inact}) and second-order (k_{inact}/K_i) rate constants for inactivation of yeast α-glucosidase are larger for the bicyclic inhibitor; and (ii) the reactivation rate constant of labeled enzyme (k_{react}) is larger for the allylic covalent adduct.

Both allylic[19] and cyclopropylcarbinyl compounds[20-22] undergo S_N1-like reactions at accelerated rates via allylic and non-classical bicyclobutonium cationic intermediates,[23,24] respectively. Of note, distinct conformations are required for formation of delocalized carbocations for carbasugar covalent inhibitors **2** and **3**. Specifically, enzyme-catalyzed covalent labeling within the enzymatic active site requires a conformation in which a π-type molecular orbital can participate in glycosidic C–O bond cleavage, a process that occurs from an oxygen n-type lone pair for natural glycoside substrates. In the current study, the cyclopropyl-containing inhibitor **3** requires a pseudo-equatorial aglycone for effective σ-bond participation (Fig. 5, Panel A),[25-27] while allylic participation from the double bond in carbocyclic inhibitor **2** entails a pseudo-axial aglycone (Fig. 5, Panel B).

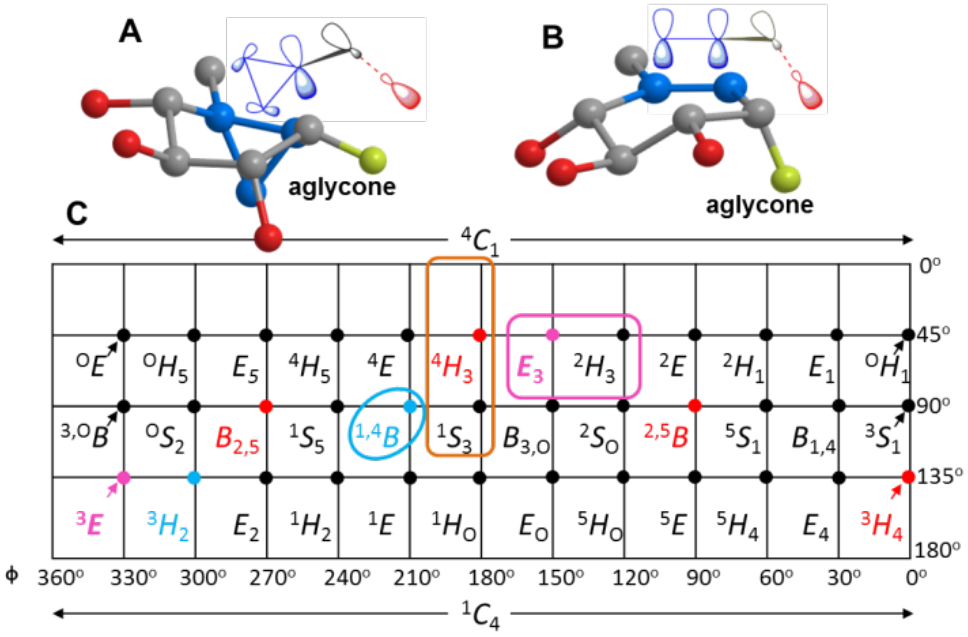
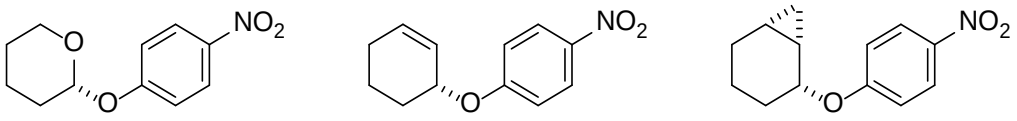


Figure 5. Conformations for π-type orbital participation (blue atoms at back) into the σ* of the glycosidic C–O bond; the C6 hydroxyl group is omitted for clarity. Panel A) α-cyclopropyl inhibitor 3; B) α-allylic compound 2; C) Mercator projection of six-membered ring conformations. The currently accepted reaction coordinate for a GH13 enzyme is indicated by the orange box. Possible conformations for a pyranosylium ion-like transition state are shown in red. The bisected conformations for cyclopropyl assisted ionization of 3 are shown in teal, and the closest to the proposed enzymatic reaction coordinate is circled. The two lowest energy conformations for an allylic cation (between C5–C6–C1) are labeled in rose (bold font) and the proposed reaction coordinate for ionization of 2 is indicated by the rose box.

Based on current theories,[5,6] we assume that the GH13 yeast α-glucosidase has evolved to stabilize pyranosylium ion-like $^4H_3$ TSs from a bound $^1S_3$ Michaelis complex and we propose that our bicyclo[4.1.0]carbasugar **3** reacts with the bisected geometry required for bicyclobutonium ion formation that is closest on the six-membered ring conformational itinerary (Figure 5, panel A, a ^{1,4}B boat) to that for the catalyzed-hydrolysis reactions of GH13 enzymes. That is, the evolved reaction coordinate for α-glucopyranoside hydrolysis, which involves a rate-determining non-chemical step that is likely a conformational change,[13] exhibits a second-order rate constant (k_{cat}/K_m = 6.7 × 10^4 M^{-1} s^{-1}) for 4-nitrophenyl α-glucopyranoside.[13] However, in order to assess the efficiency of covalent labeling it is important to calculate the relative enzymatic proficiencies for

formation of the covalent glycosyl-enzyme intermediate ($k_{cat}/K_m \times 1/k_{uncat}$ or $k_{inact}/K_i \times 1/k_{uncat}$).[15,28,29]



9 **10** **11**

Figure 6. Structures of the model compounds (9–11) used to estimate relative catalytic proficiencies.

Due to the extremely slow hydrolysis rates for the spontaneous reaction of glycosides[30] we used our previous kinetic data for the unsubstituted model compounds **9** and **11**[15] and we made **10** (Figure 6), by following standard procedures (Supporting Information). To evaluate the spontaneous rate constant for the pH-independent hydrolysis of **10** at 25 °C we used the Eyring equation to extrapolate kinetic data acquired at higher temperatures (Supporting Information Table S1). We then calculated the relative rate constants for covalent labeling of the enzyme (pseudo-glycosylation) and for cleavage of the glycosidic bond in the enzyme intermediate (pseudo-deglycosylation). Listed in Table 2 are the second-order rate constants (k_{cat}/K_m and k_{inact}/K_i) for the reactions of yeast α-glucosidase with 4-nitrophenyl α-D-glucopyranoside[13] and our two covalent inhibitors **2** and **3**, the first-order rate constants for spontaneous hydrolyses of the model compounds, and the relative catalytic proficiencies (CP), with a value of 1.0 being set for the 4-nitrophenyl glucoside $(k_{cat}/K_m \times 1/k_{uncat})_{rel}$.

Table 2. Relative enzymatic proficiencies for glycoside hydrolysis and covalent-labeling by 2 and 3.

core structure	k_{cat}/K_m or k_{inact}/K_i (M^{-1} s^{-1})	k_{spont} (s^{-1})[a]	CP_{rel}
pyranosyl	67,000[b]	4.61×10^{-5}	1.0
cyclohexenyl	1.59	4.20×10^{-8}	0.028
bicyclo[4.1.0]	6.4	1.49×10^{-6}	0.0033

[a]Rate constant extrapolated to 25 °C, Supporting Information (Table S2). bData taken from reference 13.

Notably, the catalytic proficiency for covalent labeling by **2** is higher than the corresponding value for reaction with **3**, despite the cyclopropyl inhibitor exhibiting a larger second-order constant (k_{inact}/K_i) for enzyme labeling.

Even though the ground state conformations that permit π-bond participation in **2** ($^2H_3$ half-chair or a $B_{1,4}$ boat) are removed from the GH13 α-glucosidase reaction coordinate (orange box, Figure 5, panel C) it is clear that the enzyme stabilizes formation of an allylic cation-like transition state, which should have five coplanar ring carbon atoms. Thus, we reason that the yeast α-glucosidases binds **2** in a $^2H_3$ half-chair and this results in the catalyzed formation of an E_3 allylic cation (rose box, Figure 5), a species that is structurally close to the evolved TS for glycosylation ($^4H_3$). In the case of our bicyclo[4.1.0]heptyl covalent inhibitor **3**, σ-bond participation requires a bisected geometry; however, in this case the resultant cation likely remains in the original bisected geometry due to the high rotational barrier to conformational changes in bicyclobutonium ions.[31] We conclude that covalent labeling by **2**, relative to **3**, involves a reaction coordinate that more closely matches that of the natural substrates.

Table 3. Relative proficiencies for dealkylation of the yeast α-glucosidase covalent intermediates.

core structure	k_{deglyc} *or* k_{react} *(s^{-1})*	k_{spont} *(s^{-1})*[a]	CP_{rel}
pyranosyl	*>29 s^{-1}* [b]	*4.61×10^{-5}*	*1.0*
cyclohexenyl	*1.19×10^{-5}*	*4.20×10^{-8}*	*<4.6×10^{-4}*
bicyclo[4.1.0]	*2.17×10^{-6}*	*1.49×10^{-6}*	*<2.4×10^{-6}*

[a]Supporting Information (Table S2) [b]Data for most reactive pyridinium glycoside for which k_{cat} reports on the glycosylation step.[13]

Interestingly, the relative catalytic proficiencies for the pseudo-deglycosylation reactions are markedly different (Table 3) than those for the initial covalent labeling event (Table 2). That is, the natural β-glucopyranosyl enzyme intermediate is hydrolyzed much more efficiently relative to the allylic and bicyclic covalent intermediates. We reason that the enzymatic motions that evolved to promote the distortion of the $^4C_1$ glycosyl unit in the covalent intermediate so that it undergoes hydrolysis to form the α-glucopyranose product

in a $^{1}S_{3}$ skew boat conformation are much less effective at promoting the formation of allylic and bicyclobutonium ion-like TSs from the corresponding covalent enzyme-intermediates.

Finally, we envision that these two classes of carbasugar analogue covalent inhibitors will be useful research tools for biological studies. Our covalent inhibitors, unlike other covalent inactivators, such as cyclophellitol and cyclophellitol aziridine[32-34] that irreversibly inactivate glycoside hydrolases, show a time dependent loss and return of enzymatic activity. Moreover, we should be able to customize the rates of covalent-labeling (by changing the leaving group) and reactivation (by choosing either the cyclohexenyl or the bicyclo[4.1.0]heptyl carbon skeletons). That is, our two classes of reversible covalent inhibitors could be used to monitor cellular responses to time-dependent changes in glycoside hydrolase activity. Also, assuming that the rates for each process (pseudo-glycosylation and deglycosylation) depend on the evolutionary derived pyranosylium ion-like TS structure ($^{4}H_{3}$, $^{3}H_{4}$, $B_{2,5}$, or $^{2,5}B$, Figure 5)[5,6,11] and the geometric requirements for π-type orbital participation we suggest that a simple analysis using Figure 5 will allow researchers to target the optimal carbasugar analogue for their particular glycoside hydrolase.

3.3. Associated Content

Supporting Information

Full experimental procedures, spectroscopic data, and rate constants for hydrolysis of **10**. Supporting Information is available free of charge on the ACS Publications website at DOI:

Material and Methods

General. All reactions described were performed under an atmosphere of dry nitrogen using oven dried glassware unless otherwise specified. Normal-phase flash chromatography was carried out with 230–400 mesh silica gel and reverse-phase chromatography was carried out using a C18-cartridge on an automated system. All reagents, solvents and starting materials were purchased from Sigma Aldrich, Alfa Aesar, TCI America, Strem, EMD, Anachemia, Caledon, Fisher or ACP and were used without

further purification unless otherwise specified. Benzene was freshly distilled from Na metal. Cold temperatures were maintained by use of the following conditions: 0 °C, ice-water bath; −78 °C, acetone-dry ice bath; temperatures between −78 °C and 20 °C that were required for longer reaction times were maintained with a Neslab Cryocool Immersion Cooler (CC-100 II) in a EtOH/propan-2-ol bath. Optical rotations were measured using a Perkin-Elmer 341 polarimeter and are reported in units of deg cm^2 g^{-1} (concentrations reported in units of g/100 cm^3). NMR spectra were recorded on either a Bruker Avance 600 equipped with a QNP or TCI cryoprobe (600 MHz) or a Bruker 400 (400 MHz) instrument. Chemical shifts (δ) are listed in ppm downfield from TMS using either the residual solvent peak as internal reference or 4,4- dimethyl-4-silapentane-1-sulfonic acid as external reference in D_2O. ^1H and ^{13}C NMR peak assignments are made based on ^1H-^1H COSY, ^1H-^{13}C HMQC and ^1H-^{13}C HMBC experiments. Coupling constants are reported in Hertz (Hz) and are reported to the nearest 0.1 Hz. Infrared (IR) spectra were recorded on a Perkin Elmer Spectrum Two Fourier transform spectrometer with neat samples. Only selected, characteristic absorption data are provided for each compound. High resolution mass spectra were performed on an Agilent 6210 TOF LC/MS using ESI-MS or were carried out by the Notre Dame University Mass Spectrometry Department using 3 EI techniques. (1S,4S,5R,6S)-4,5,6-tris(benzyloxy)-1-((benzyloxy)methyl)cyclohex-2-en-1-ol (4) was prepared from 2,3,4,6-tetra-O-benzyl-D-glucopyranose in four steps (42% overall yield) according to the method of Kapferer et al.[1]

(1S,4S,5R,6S)-4,5,6-tris(benzyloxy)-1-((benzyloxy)methyl)cyclohex-2-en-1-yl acetate (5): To a solution of 4 (1.11 g, 2 mmol) in dry THF (30 mL) was added a solution of LiHMDS (1.0 M, 4 mL) in THF at 0 °C. This mixture was stirred at this temperature for 30 min, followed by the addition of acetyl chloride (0.19 mL, 2.7 mmol). The mixture was stirred at rt for 24 h under a nitrogen atmosphere. The mixture was then diluted with EtOAc (100 mL), and the solution was washed with sat. aq. NaHCO3 (60 mL), followed by brine (60 mL). The organic layer was dried with Na2SO4 and concentrated under reduced pressure. The resulting residue was purified by flash chromatography (Hexane:EtOAc, 5:1) to afford 5 (623 mg, 52%) as a colorless oil.

(1S,4R,5S,6S)-4,5,6-tris(benzyloxy)-3-((benzyloxy)methyl)cyclohex-2-en-1-yl acetate (6): To 5 (32 mg, 0.055 mmol) in EtOAc (2 mL) was added

bis(benzonitrile)dichloropalladium (2.2 mg). This mixture was refluxed for 26 h under a nitrogen atmosphere. The mixture was then filtered through a silica pad, and washed with EtOAc and the filtrate was concentrated to afford 32 mg of compound 6, which was used without further purification.

(1S,4R,5S,6S)-4,5,6-tris(benzyloxy)-3-((benzyloxy)methyl)cyclohex-2-enol (7): To a solution of 6 (150 mg, 0.26 mmol) in dry MeOH (3 mL) was added a solution of 1.0 M NaOMe in MeOH (3 mL). The mixture was then refluxed for 45 min under a nitrogen atmosphere. The volatiles were evaporated, and the mixture was diluted with EtOAc (50 mL). The mixture was quenched with 10% HCl, and the organic layer was washed with sat. aq. NaHCO3 (20 mL), brine (20 mL), and dried with Na2SO4. The organic layer was concentrated, and the resulting residue was purified by flash chromatography (Hexanes: EtOAc, 5:1- 2:1) to afford compound 7 (111 mg, 4 77%) as a colorless oil. This material exhibited identical ^{1}H and ^{13}C NMR data to those reported in the literature.[2]

(1S,2S,3S,4S,5R,6R)-3,4,5-tris(benzyloxy)-6-((benzyloxy)methyl)bicyclo[4.1.0]heptan-2-ol (8): Under a nitrogen atmosphere, a diethyl zinc solution in hexane (1 M, 3.5 mL) was added to cooled dry toluene (10 mL; –15°C). This mixture was stirred at –15°C for 10 min and then CH_2I_2 (0.38 mL, 4.7 mmol) was added dropwise to the reaction mixture. After 10 min, trifluoroacetic acid (30 µL, 0.4 mmol) was added dropwise to the cooled solution following which the cooling bath was removed and the reaction mixture was stirred at room temperature for 5 min. To the resultant mixture, a solution of 7 (124 mg, 0.23 mmol) in dry toluene (5 mL) was added and the reaction mixture was stirred for 15 h at room temperature. The reaction was quenched by the addition of aqueous HCl (10%), and then diluted by adding EtOAc (30 mL). After separation, the organic layer was washed with saturated aqueous $NaHCO_3$ (25 mL), and brine (25 mL), dried over Na_2SO_4, and the resulting solution was concentrated. The residue was purified by flash chromatography (Hexane:EtOAc, 5:1–2:1) to afford 8 (103 mg, 81%) as a colorless oil. This material exhibited identical ^{1}H and ^{13}C NMR data to those reported in the literature.[2]

(3S,4S,5S,6R)-4,5,6-tris(benzyloxy)-1-((benzyloxy)methyl)-3-(4-nitrophenoxy)cyclohexene (12): A suspension of NaH in mineral oil (60%, 62 mg, 1.6

mmol) was washed with hexane (2 × 5 mL) before being transferred in dry DMSO (25 mL) to a 100 mL flask at rt. To this mixture a solution of 7 (124 mg, 0.23 mmol) in dry DMSO (15 mL) was added dropwise under nitrogen atmosphere. The mixture was left for 30 min at rt before potassium benzoate (63 mg, 0.39 mmol) was added. After a further 30 min, 1-fluoro-4-nitrobenzene (110 µL, 1.5 mmol) was added. The reaction was stirred overnight at room temperature under nitrogen atmosphere, and it 5 was then quenched by the addition of a saturated NH4Cl solution (20 mL). After the addition of brine (75 mL) the resulting mixture was extracted with ether (3 × 100 mL) and the combined organic layers were washed with saturated aqueous $NaHCO_3$ (6 × 100 mL). The organic layer was dried (Na2SO4) and concentrated under reduced pressure. The residue was purified by flash column chromatography (Hexane:EtOAc, 16:1–2:1) to give 12 as a colorless syrup (128 mg, 84%).

(1R,2R,3S,4S,5S,6S)-2,3,4-tris(benzyloxy)-1-((benzyloxy)methyl)-5-(4-nitrophenoxy)bicyclo[4.1.0]heptane (13): A suspension of NaH in mineral oil (60%, 90 mg, 2.25 mmol) was washed with hexane (2 × 5 mL) before being transferred in dry DMSO (25 mL) to a 100 mL flask at rt. To this mixture a solution of 8 (186 mg, 0.338 mmol) in dry DMSO (18 mL) was added dropwise under nitrogen atmosphere. The mixture was left for 30 min at rt before potassium benzoate (92 mg, 0.57 mmol) was added. After a further 30 min, 1-fluoro-4- nitrobenzene (160 µL, 1.52 mmol) was added and the reaction was stirred overnight at rt under a nitrogen atmosphere. The resulting mixture was quenched by the addition of saturated NH4Cl solution (20 mL). Addition of brine (75 mL) was followed by extraction with ether (3 × 100 mL) and the combined organic layers were washed with saturated aqueous $NaHCO_3$ (6 × 100 mL). The organic layer was dried (Na2SO4) and concentrated under reduced pressure. The residue was purified by flash column chromatography (Hexane:EtOAc, 16:1–2:1) to give 13 as a colorless syrup (144 mg, 64%).

(1R,2S,3R,6S)-4-(hydroxymethyl)-6-(4-nitrophenoxy)cyclohex-4-ene-1,2,3-triol (2): To a solution of 12 (4 mg) in anhydrous CH2Cl2 (0.4 mL) that had been cooled to –78 °C was added a solution of BCl_3 in CH2Cl2 (1 M; 40 µL) dropwise, and the resulting mixture was stirred for 1 h at –78 °C. After which the reaction was allowed to warm slowly to room temperature over 2 h, 6 when it was quenched by the addition of 1:1 v/v MeOH/DCM and the volatiles were removed under reduced pressure. The residue was purified by silica gel

column chromatography to give a viscous oil (1 mg, 58%). ^{1}H NMR (600 MHz, CD_3OD) δ 3.75 (dd, J = 10.2, 3.9 Hz, 1H), 3.98 (dd, J = 10.1, 7.3 Hz, 1H), 4.04 (d, J = 7.3 Hz, 1H), 4.23–4.17 (m, 2H), 5.11 (t, J = 4.5 Hz, 1H), 6.04–6.00 (m, 1H), 7.18 (d, J = 9.3 Hz, 2H), 8.19 (d, J = 9.3 Hz, 2H); ^{13}C NMR (151 MHz, CD_3OD) δ 165.35, 147.73, 142.69, 126.71, 118.04, 116.84, 75.22, 74.50, 74.09, 72.00, 62.74, 49.00. HRMS (ESI)+ m/z 320.0741 [C13H15NO7 (M + Na)+ requires 320.0746].

(1R,2R,3S,4R,5S,6S)-1-(hydroxymethyl)-5-(4-nitrophenoxy)bicyclo[4.1.0]heptane-2,3,4-triol (3): A solution of 13 (30 mg) in anhydrous CH2Cl2 (0.7 mL) was cooled to –78 °C. To this mixture a solution of BCl_3 in CH2Cl2 (1 M; 0.29 mL) was added dropwise and the reaction mixture was stirred for 1 h. Then it was slowly warmed to room temperature over 2 h, when it was quenched by the addition of 1:1 v/v MeOH/DCM and the volatiles were removed under reduced pressure. The resulting solid was purified by silica gel column chromatography to obtain pure crystalline 3 (9 mg, 65%); mpt. = 160–162 °C; ^{1}H NMR (600 MHz, CD_3OD) δ 0.42 (dd, J = 9.9, 5.4 Hz, 1H), 0.71 (t, J = 5.5 Hz, 1H), 1.64 (ddd, J = 9.4, 7.8, 5.6 Hz, 1H), 2.89 (d, J = 11.4 Hz, 1H), 3.51 (dd, J = 5.0, 1.6 Hz, 2H), 3.98 (d, J = 11.4 Hz, 1H), 4.10 – 4.06 (m, 1H), 5.06 – 5.02 (m, 1H), 7.17 (d, J = 9.3 Hz, 2H), 8.21 (d, J = 9.3 Hz, 2H); ^{13}C NMR (151 MHz, CD_3OD) δ 164.79, 142.75, 126.71, 116.93, 75.63, 73.92, 72.96, 71.41, 67.18, 32.77, 22.31, 9.86. HRMS (ESI)+ m/z 334.0897 [C14H17NO7 (M + Na)+ requires 334.0903].

3-(4-Nitrophenoxy)cyclohex-1-ene (10): In a round-bottom flask charged with cyclohex-2-en- 1-one (2 g, 20 mmol) and CeCl3.7H2O (2 g, 5 mmol) were dissolved in methanol (20 mL). To this stirred solution was added NaBH4 (700 mg, 20 mmol) in small portions. After the vigorous evolution of gas had subsided the reaction mixture was carefully neutralized with dilute aqueous 7 HCl (15 mL). The resultant solution was extracted with ether (3 × 30 mL) and the combined organic layers were washed with brine (2 × 20 mL) and dried (Na2SO4). The volatiles were removed under reduced pressure to afford the crude allylic alcohol (1.6 g, 82% yield) in >98% purity as determined by ^{1}H NMR spectroscopy. This material was carried forward to the next step without further purification. The resultant allylic alcohol was dissolved in dimethoxyethane (100 mL) and 4-nitrofluorobenzene (2.3 g, 16 mmol) was added with cooling to 0 °C. At this point, NaH (700 mg, 29 mmol, 60% dispersion in oil) was added portion wise with continuous stirring.

Following which, the reaction mixture was warmed to 60 °C and stirring was continued at this temperature for another 12 h. Upon quenching the reaction by addition of water (40 mL), the product was extracted from the aqueous layer with CHCl3 (3 × 50 mL). The organic layers were combined and washed with brine (2 × 20 mL) and dried (Na_2SO_4). This was followed by removal of the volatiles under reduced pressure to afford a yellow syrup. The crude residue was purified via flash column chromatography (5% EtOAC-Hexane, 50% Hexane-Toluene) to obtain pure product as a yellow syrup (2.2 g, 62% yield); ^{1}H NMR (400 MHz, CDCl3) δ 8.27–8.17 (m, 2H), 7.04–6.92 (m, 2H), 6.06–6.01 (m, 1H), 5.85–5.81 (m, 1H), 4.93 (dt, J = 7.2, 3.4, 1H), 2.20– 2.01 (m, 2H), 2.00–1.93 (m, 1H), 1.92–1.79 (m, 2H), 1.72–1.62 (m, 1H); ^{13}C NMR (101 MHz, CDC13) δ 163.14, 141.09, 133.44, 125.92, 124.72, 115.22, 71.65, 28.05, 24.94, 18.68; HRMS (ESI)+ m/z 242.0780 [C12H13NO3 (M + Na)+ requires 242.0793].

Intact Native and Inactivated Protein Mass Spectrometry: 1mg/mL protein samples were diluted 250x in 95:5 water:acetonitrile in 0.1% formic acid. An aliquot of this sample (5 μL) was then injected onto a 5 mm C4 column connected to a Waters Xevo GS-2 QTof mass spectrometer via a NanoAquity UPLC system, and through a Waters Z-spray electrospray ion 8 source. Samples were eluted in a 2 minute gradient from 5–100% acetonitrile at 20 μL/min. Mass spectra were summed and deconvoluted using Waters' MaxEnt algorithm.

Enzyme kinetics: Yeast α-glucosidase (G5003) was purchased from Sigma Aldrich. A stock was made by dissolving the lyophilized powder in sodium phosphate buffer (50 mM, pH 6.84) containing a BSA concentration of 1 mg/ml. To determine the kinetic parameters for inactivation, we used a classical dilution assay method that involved incubating enzyme with various concentrations of covalent inhibitor (2 or 3) in 50 mM sodium phosphate buffer, pH 6.84 and periodically measuring the activity of the enzyme by removing an aliquot (20 μL) of the inactivation stock solution and adding it to a pre-equilibrated solution (25 °C) containing 4-nitrophenyl α-D-glucopyranoside (50 μM) in sodium phosphate buffer (50 mM, pH = 6.84, [BSA] = 1 mg/ml). We determined the pseudo first-order rate constants for inhibition (k_{obs}) at each inactivator concentration by fitting the absorbance versus time data to a standard first-order rate equation. The first- and second-order rate constants (k_{inact} and k_{inact}/K_I) for the inactivation process were determined by fitting the pseudo-first

order rate constants for inactivation (k_{obs}) versus inactivator concentration data to a standard Michaelis–Menten equation.

Reactivation kinetics: In a typical enzyme reactivation assay, we incubated yeast α-glucosidase (10 μL, 44.03 μM) in sodium phosphate buffer (50 mM, pH = 6.84, [BSA] = 1 mg/ml) with each inactivator (5 μL, 10 mM) for 1 hour at 25 °C. Following the incubation, the excess inactivator was removed and the inactivation buffer was exchanged by filtering the solutions using a 5-K molecular weight cutoff centrifugal filter and washing three times with reactivation buffer (3 × 500 μL) at 4 °C to give a final sample volume of 200 μL. Each sample was incubated at 25 °C and the activity of the 9 enzyme was measured by removing an aliquot (20 μL) of the reactivation stock solution at various time points and adding it to a pre-equilibrated solution of 50 μM of 4-nitrophenyl α-Dglucopyranoside in sodium phosphate buffer (50 mM, pH = 6.84, [BSA] = 1 mg/ml). The measured initial rates were fit to a standard first-order rate equation to give estimates of the rate constant of reactivation (k_{react}).

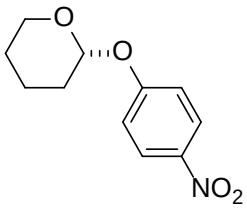
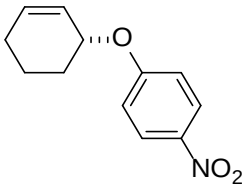
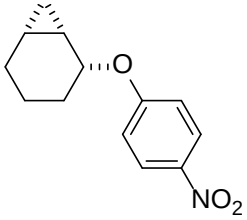
Measurement of Spontaneous Hydrolysis Kinetics: A portion of a stock solution of allylic pNP 10 in dioxane (final concentration 50 μM, dioxane content <5%) was added to a preequilibrated, at various temperatures (60–80 °C), HEPES buffer solution (40 mM, pH 7.5). The rate of production of 4-nitrophenolate was monitored at 400 nm. First-order rate constants were obtained by fitting the time–absorbance data (for at least three half-lives) to a standard first-order rate equation. These spontaneous hydrolysis reactions were monitored at pH values of 7.5 and 8.3 in order to show that the rate constants for hydrolysis were independent of pH under these conditions.

Supplementary Table 1: **Rate constants for the spontaneous hydrolysis of 3-(4-nitrophenyl)cyclohex-1-ene (10) in HEPES buffer (40 mM, pH 7.5).**

T (°C)	k_{obs} s^{-1}
25.0	4.20×10^{-8} [a]
60.0	$(4.29 \pm 0.14) \times 10^{-6}$
70.0	$(1.39 \pm 0.04) \times 10^{-5}$
80.0	$(4.03 \pm 0.15) \times 10^{-5}$

[a] Value calculated using the Eyring equation.

Supplementary Table 2: **Extrapolated first-order rate constants (k_{obs}) for the hydrolyses of 2-(4-nitrophenoxy)tetrahydro-2H-pyran, cis-2-(4-nitrophenoxy)bicyclo[4.1.0]heptane, and 3-(4-nitrophenyl)cyclohex-1-ene in HEPES buffer (40 mM, pH 7.5) at 25 °C, and the relative rates of hydrolyses.**

Compound	**9**	**10**	**11**
Structure			
k_{obs}	4.61×10^{-5} s^{-1} [a]	4.20×10^{-8} s^{-1}	1.49×10^{-6} s^{-1} [b]
k_{rel}	1100	1.0	35

[a] Value interpolated from kinetic data in Chakladar et al.[3] [b] Value extrapolated from kinetic data in Chakladar et al.[3]

References:

(1) Kapferer, P.; Sarabia, F.; Vasella, A. *Helv. Chim. Acta* **1999**, *82*, 645-656.

(2) Dookhun, V.; Bennet, A. J. *Can. J. Chem.* **2004**, *82*, 1361-1364.

(3) Chakladar, S.; Wang, Y.; Clark, T.; Cheng, L.; Ko, S.; Vocadlo, D. J.; Bennet, A. J. *Nat. Commun.* **2014**, *5*, 5590.

Figure S1: Reconstructed MS of intact yeast α-glucosidase; Panel A) native enzyme; B) enzyme after inhibition with **2**; and C) enzyme after inactivation with **3**.

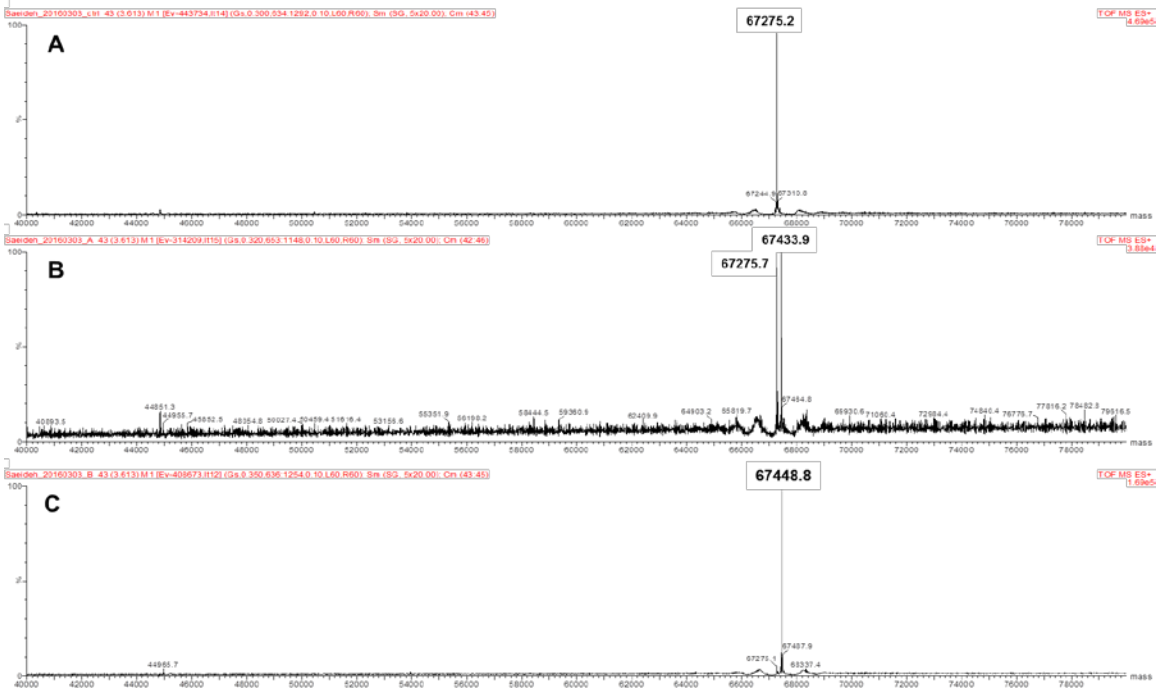


Figure S2: ^{1}H NMR spectrum for **2** in CD_3OD.

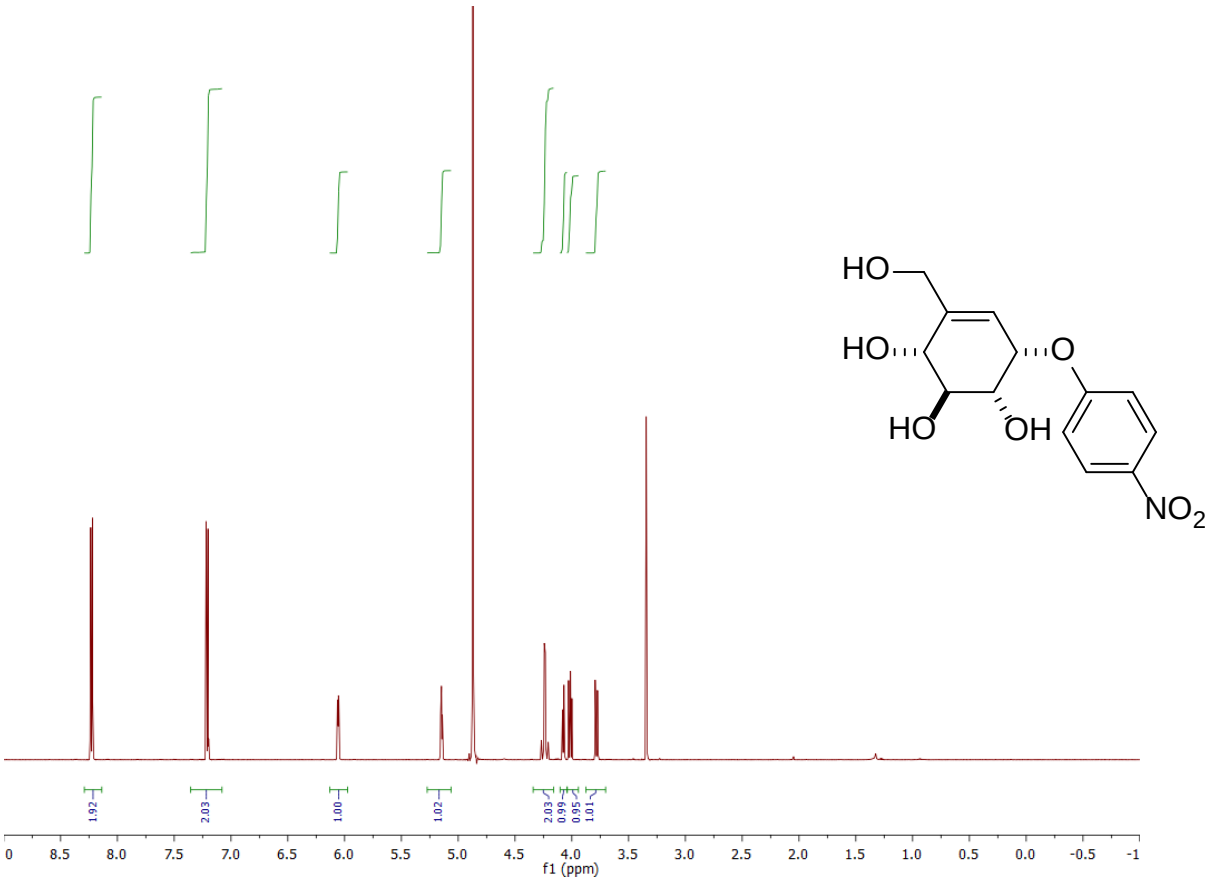


Figure S3: ^{13}C NMR spectrum for **2** in CD_3OD.

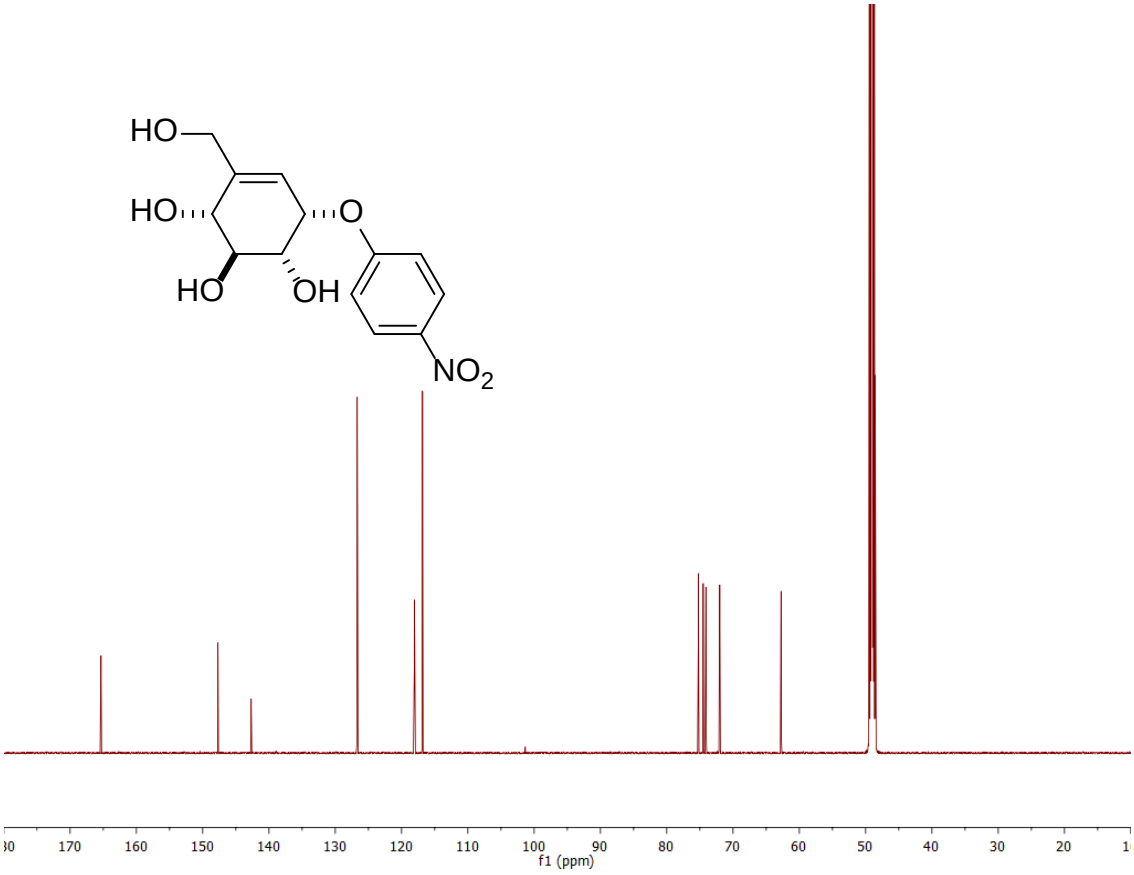


Figure S4: ^{1}H NMR spectrum for **3** in CD_3OD.

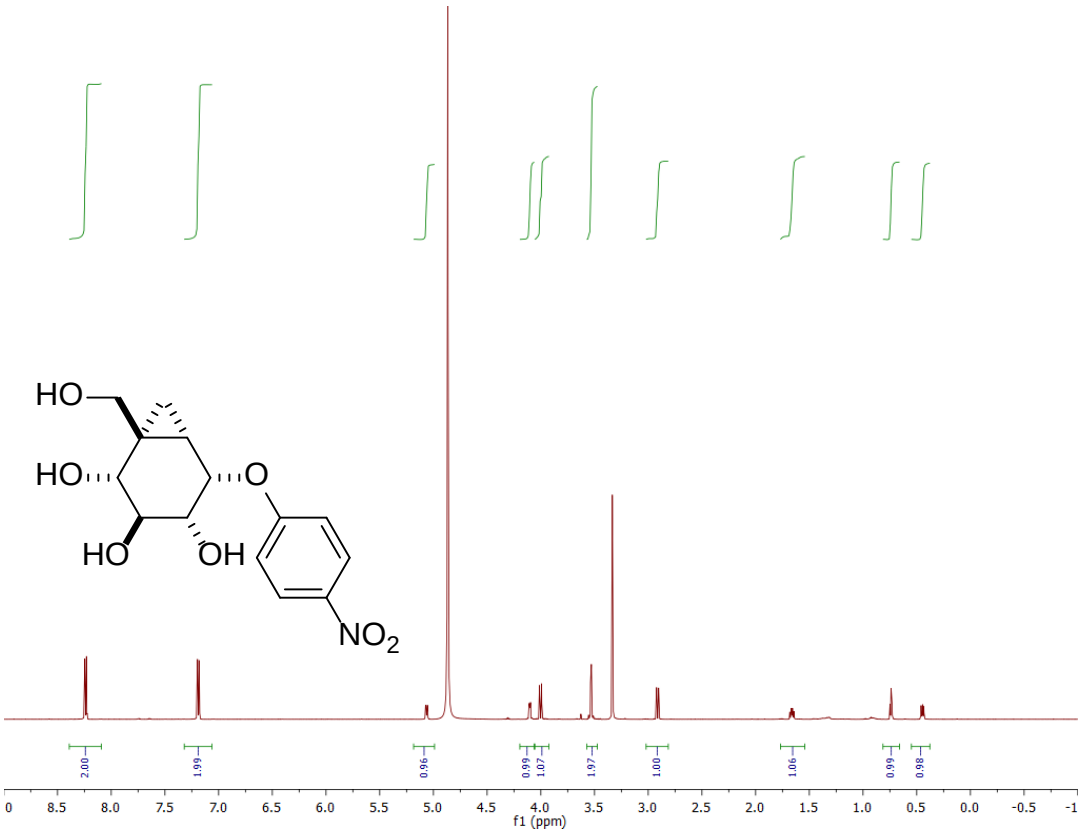
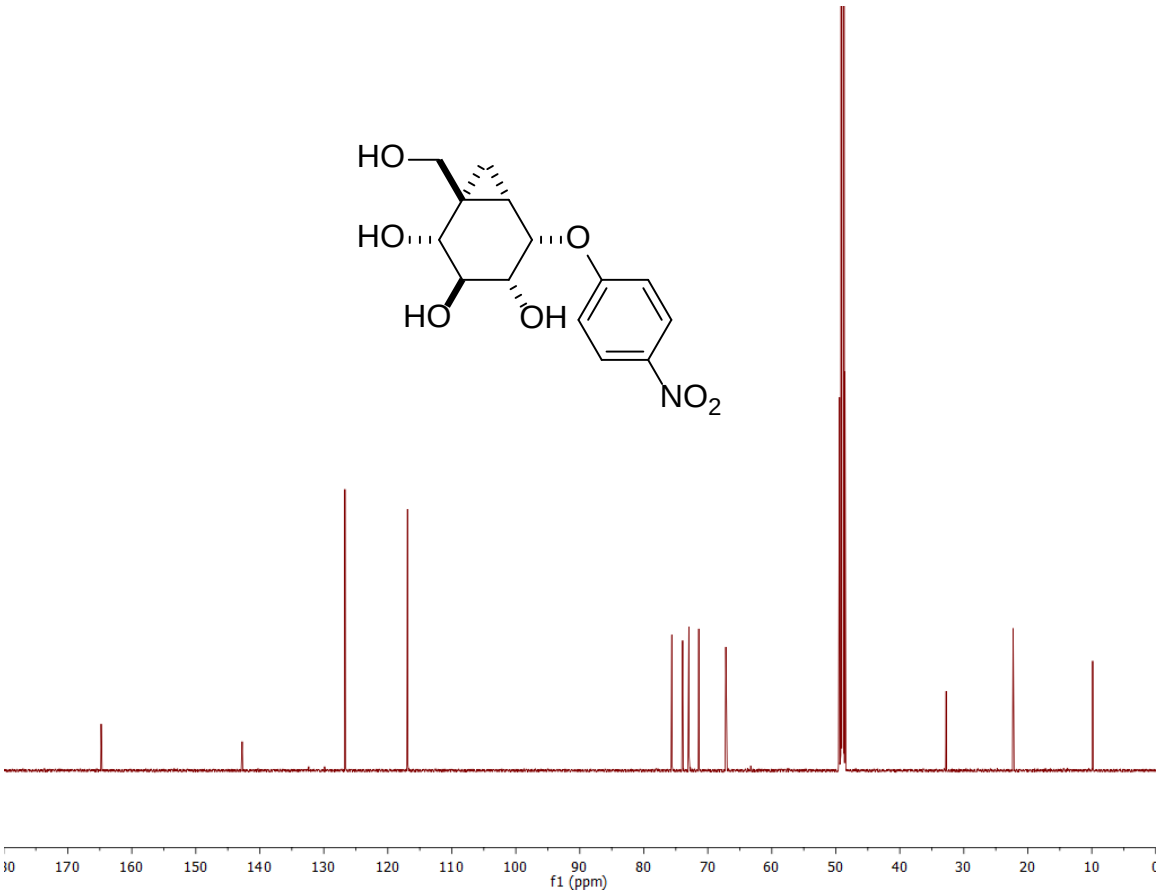


Figure S5: ^{13}C NMR spectrum for **3** in CD_3OD.



Author Information

Corresponding Author

*bennet@sfu.ca

Notes

The authors declare no competing financial interests.

Acknowledgment

The authors thank Ms. Kyung-Mee Moon, Mr. Jason Rogalski, and Dr. Leonard Foster for acquisition of mass spectral data. This work was supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant (AJB: #121348-2012).

3.4. References

1. Witte, M. D.; van der Marel, G. A.; Aerts, J. M.; Overkleeft, H. S. Org Biomol Chem 2011, 9, 5908-5926.
2. Moremen, K. W.; Tiemeyer, M.; Nairn, A. V. Nat. Rev. Mol. Cell. Biol. 2012, 13, 448-462.
3. Schnaar, R. L.; Gerardy-Schahn, R.; Hildebrandt, H. Physiol Rev 2014, 94, 461-518.
4. Lombard, V.; Ramulu, H. G.; Drula, E.; Coutinho, P. M.; Henrissat, B. Nucleic Acids Res. 2014, 42, D490-D495.
5. Vocadlo, D. J.; Davies, G. J. Curr. Opin. Chem. Biol. 2008, 12, 539-555.
6. Davies, G. J.; Planas, A.; Rovira, C. Acc. Chem. Res. 2012, 45, 308-316.
7. Mhlongo, N. N.; Skelton, A. A.; Kruger, G.; Soliman, M. E. S.; Williams, I. H. Proteins 2014, 82, 1747-1755.
8. McCarter, J. D.; Withers, S. G. Curr. Opin. Struct. Biol. 1994, 4, 885-892.
9. Sinnott, M. L. Chem. Rev. 1990, 90, 1171-1202.
10. Sinnott, M. Carbohydrate Chemistry and Biochemistry: Structure and Mechanism; 2nd ed.; RSC Publishing: Cambridge, 2013.
11. Speciale, G.; Thompson, A. J.; Davies, G. J.; Williams, S. J. Curr. Opin. Struct. Biol. 2014, 28, 1-13.
12. Uitdehaag, J. C. M.; Mosi, R.; Kalk, K. H.; van der Veen, B. A.; Dijkhuizen, L.; Withers, S. G.; Dijkstra, B. W. Nat. Struct. Biol. 1999, 6, 432-436.
13. Hosie, L.; Sinnott, M. L. Biochem. J. 1985, 226, 437-446.
14. Huang, X. C.; Tanaka, K. S. E.; Bennet, A. J. J. Am. Chem. Soc. 1997, 119, 11147-11154.
15. Chakladar, S.; Wang, Y.; Clark, T.; Cheng, L.; Ko, S.; Vocadlo, D. J.; Bennet, A. J. Nat. Commun. 2014, 5, 5590.
16. Adamson, C.; Pengelly, R.; Shamsi Kazem Abadi, S.; Chakladar, S.; Draper, J.; Britton, R.; Gloster, T.; Bennet, A. J. Angew. Chem., Int. Ed. Engl. 2016, 14978–14982.
17. Kapferer, P.; Sarabia, F.; Vasella, A. Helv. Chim. Acta 1999, 82, 645-656.
18. Lim, C.; Baek, D. J.; Kim, D.; Youn, S. W.; Kim, S. Org. Lett. 2009, 11, 2583-2586.

19. Lowry, T. H.; Richardson, K. S. Mechanism and theory in organic chemistry; 3rd ed.; Harper & Row: New York, 1987.
20. Roberts, J. D.; Mazur, R. H. J. Am. Chem. Soc. 1951, 73, 2509-2520.
21. Friedrich, E. C.; Jassawalla, J. D. C. J. Org. Chem. 1979, 44, 4224-4229.
22. Moss, R. A.; Zheng, F.; Johnson, L. A.; Sauers, R. R. J. Phys. Org. Chem. 2001, 14, 400-406.
23. Harris, J. M.; Moffatt, J. R.; Case, M. G.; Clarke, F. W.; Polley, J. S.; Morgan, T. K., Jr.; Ford, T. M.; Murray, R. K., Jr. J. Org. Chem. 1982, 47, 2740-2744.
24. Huang, X.; Tanaka, K. S. E.; Bennet, A. J. J. Am. Chem. Soc. 1998, 120, 1405-1409.
25. de Meijere, A. Angew. Chem., Int. Ed. Engl. 1979, 18, 809-826.
26. Olah, G. A.; Prakash, G. K. S.; Rasul, G. J. Am. Chem. Soc. 2008, 130, 9168-9172.
27. Werstiuk, N. H.; Poulsen, D. A. Arkivoc 2009, 38-50.
28. Wolfenden, R.; Snider, M. J. Acc. Chem. Res. 2001, 34, 938-945.
29. Bartlett, P. A.; Marlowe, C. K. Biochemistry 1983, 22, 4618-4624.
30. Wolfenden, R.; Lu, X.; Young, G. J. Am. Chem. Soc. 1998, 120, 6814-6815.
31. Kabakoff, D. S.; Namanworth, E. J. Am. Chem. Soc. 1970, 92, 3234-3235.
32. Li, K. Y.; Jiang, J. B.; Witte, M. D.; Kallemeijn, W. W.; van den Elst, H.; Wong, C. S.; Chander, S. D.; Hoogendoorn, S.; Beenakker, T. J. M.; Codee, J. D. C.; Aerts, J. M. F. G.; van der Marel, G. A.; Overkleeft, H. S. Eur. J. Org. Chem. 2014, 6030-6043.
33. Arcelli, A.; Cerè, V.; Peri, F.; Pollicino, S.; Ricci, A. Tetrahedron 2001, 57, 3439-3444.
34. Kallemeijn, W. W.; Li, K. Y.; Witte, M. D.; Marques, A. R.; Aten, J.; Scheij, S.; Jiang, J.; Willems, L. I.; Voorn-Brouwer, T. M.; van Roomen, C. P.; Ottenhoff, R.; Boot, R. G.; van den Elst, H.; Walvoort, M. T.; Florea, B. I.; Codee, J. D.; van der Marel, G. A.; Aerts, J. M.; Overkleeft, H. S. Angew. Chem., Int. Ed. Engl. 2012, 51, 12529-12533.

Chapter 4. Characterization of Reaction Coordinates for Covalent Labeling of Glycosidase by Cyclohexene Inhibitors.

Weiwu Ren,[1] Verena Oehler,[3] Saeideh Shamsi Kazem Abadi,[2] Jason Draper,[1] Michael Meanwell,[1] Robert Britton,[1,*] Tracey M. Gloster,[3,*] and Andrew J. Bennet[1,*]

Departments of Chemistry[1] and Molecular Biology and Biochemistry,[2] Simon Fraser University, 8888 University Drive, Burnaby, B.C. V5A 1S6, Canada. [3]Biomedical Sciences Research Complex, University of St Andrews, North Haugh, St Andrews, Fife (UK)

Corresponding author e-mail address: rbritton@sfu.ca; tmg@st-andrews.ac.uk; bennet@sfu.ca

4.1. Introduction

Of the three main biological polymeric building blocks carbohydrates are the most structurally diverse. Enzymes that catalyze their transfer from a biomolecule to water (glycoside hydrolases; GHs)[1] are often critical for biological processes including digestion of carbohydrates, degradation of plant biomass, and pathogen infection mechanisms.

We show for the first time the conformational itinerary of an allylic carbohydrate mimic as it undergoes pseudo-glycosylation and deglycosylation with a GH. Insights into the structure of such mimics throughout the catalytic cycle enables a greater understanding of the subtleties of GH mechanism.

It is critical for the design of small molecule transition state analogues (TSAs), which have potential as therapeutics,,[2-4] that we understand how GHs stabilize the positively charged transition states (TSs) of natural substrates (pyranosides). Also, an important component of reaction coordinate stabilization for GH enzymes involves the changes in the pyran ring conformation[12] from that in the Michaelis complexes to that of the glycosyl-enzyme intermediate.. Indeed, GHs have evolved to hydrolyze unactivated glycosides efficiently and these enzymes exhibit catalytically proficiencies of up to 10^{17} M^{-1}.[9-11] Recently several carbasugar mimics that react with GH enzymes via cationic TSs have been reported.[5-8]

There are several mechanisms used by GHs to catalyze hydrolysis of the glycosidic bond. Most GHs that catalyze hydrolysis with overall retention of stereochemistry rely on a pair of active site aspartic and/or glutamic acid (Asp/Glu) residues, in which one functions as a general acid/base while the second acts a nucleophile/leaving group.[13-15] That is, these enzymes employ two sequential S_N2-like reactions each of which involves an inversion of configuration. The first S_N2-like reaction generates a covalent glycosyl-enzyme intermediate while the second results in hydrolysis of this intermediate. The TSs for both chemical steps incorporate positive charge delocalization from the anomeric carbon onto the ring oxygen (pyranosylium ion-like character) with the ring adopting one of several low energy conformations[12] that include half-chairs ($^4H_3/^3H_4$), boats ($^{2,5}B/B_{2,5}$) or envelopes ($^4E/^3E$).[13,16-18]

For enzyme catalyzed reactions non-chemical steps are often kinetically significant, which for GHs can be probed by measuring the effect of leaving group ability on the catalytic rate constants. Of note, the retaining α-galactosidases from *Thermotoga maritima* (*Tm*GalA) was shown to exhibit both k_{cat} and k_{cat}/K_m values for the hydrolysis of aryl α-D-galactopyranosides that were independent of leaving group ability,[19] thus, suggesting that, a kinetically significant non-chemical step precedes glycosidic bond cleavage.[19,20]

Previously, we exploited the ability of most retaining GHs to stabilize positive charge at both the glycosylation and the deglycosylation transition states in the design of cyclopropylmethyl[6,7] and allylic mechanism-based covalent inhibitors (e.g., **1**) of retaining GHs. It is likely that both of these inhibitors undergo a pseudo-glucosylation reaction as shown in Figure 1.[7] Recently, Danby and Withers reported that the allylic carbaglucose analogue **2** was a substrate for several β-glucosidases.[8] have also designed and synthesized inhibitors in which the C2 hydroxyl group is replaced by a fluorine atom, which decreases the rate of hydrolysis of the corresponding enzyme-bound intermediate[21-23] during our study of the conformational itinerary for a cyclopropyl-containing covalent inhibitor.[5]

Herein, we describe the first conformational itinerary for an allylic mechanism-based covalent inhibitor of a glycoside hydrolase. That is, we present the de novo synthesis of allylic carbasugar mimics of galactose (**3**) and 2-deoxy-2-fluorogalactose (**4**) and the kinetic characterization of the reaction of these two covalent inhibitors with *Tm*GalA. We also present the structural characterization of **4** with *Tm*GalA in the form of a Michaelis complex (with an active site mutant), a covalent adduct after loss of the 2,4-dinitrophenyl leaving group, and as the resulting complex with the hydrolyzed product **6**.

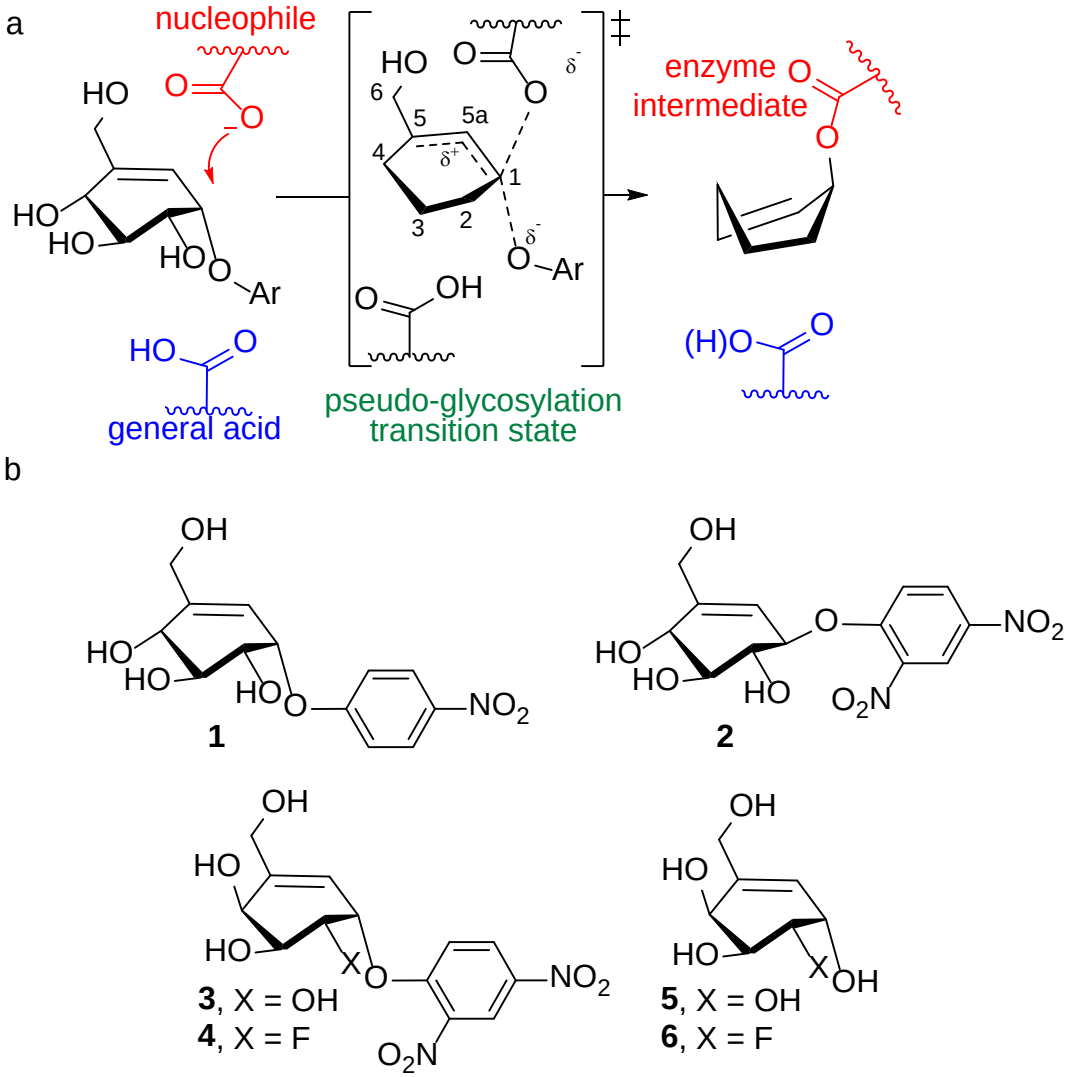


Figure 1. (a) Proposed mechanism of pseudo-glucosylation for the allylic α-glucoside mimic 1; for clarity, most hydroxyl groups are not shown for the transition state or intermediate. (b) Structures of mechanism-based covalent inhibitor 1, β-glucoside analogue 2, the α-galactose analogue inhibitor 3, the 2-deoxy-2-fluoro covalent inhibitor 4, and 5 and 6, the hydrolyzed products for the α-galactose analogue inhibitors 3 and 4, respectively.

4.2. Experimental Section

4.2.1. Materials and Methods

All anhydrous reactions described were performed under an atmosphere of nitrogen using flamedried glassware. Normal phase column chromatography was carried out with 230-400 mesh silica gel (Silicycle, SiliaFlash® P60). Concentration and removal of trace solvents was done with a Büchi rotary evaporator using a dry ice/acetone condenser and vacuum applied from a Büchi V-500 pump. All reagents and starting materials were purchased from Sigma Aldrich, Alfa Aesar, TCI America or Arcos and were used without further purification. All solvents were purchased from Sigma Aldrich, EMD, Anachemia, Caledon, Fisher or ACP and used without further purification unless otherwise specified. CH_2Cl_2 was freshly distilled over CaH_2; Tetrahydrofuran (THF) was freshly distilled over Na metal/benzophenone. Cold temperatures were maintained by use of the following conditions: 0 °C, ice-water bath; −78 °C, acetone-dry ice bath; temperatures between −78 °C and 0 °C required for longer reaction times were maintained with a Neslab Cryocool Immersion Cooler (CC-100 II) in a 2-propanol bath.

Nuclear magnetic resonance (NMR) spectra were recorded on a Bruker Avance 600 equipped with a QNP or TCI cryoprobe (600 MHz), Bruker 500 (500 MHz), or Bruker 400 (400 MHz) using $CDCl_3$ or CD_3OD as solvent. Signal positions (δ) are given in parts per million from tetramethylsilane (δ 0) and were measured relative to the signal of the solvent (^{1}H NMR: $CDCl_3$: δ 7.26, CD_3OD: δ 3.31; ^{13}C NMR: $CDCl_3$: δ 77.16, CD_3OD: δ 49.00). Coupling constants (*J* values) are given in Hertz (Hz) and are reported to the nearest 0.1 Hz. ^{1}H NMR spectral data are tabulated in the order: multiplicity (s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet; br., broad), coupling constants, number of protons. NMR spectra were recorded. Infrared (IR) spectra were recorded on a Perkin Elmer Spectrum Two™ Fourier transform spectrometer with neat samples. Only selected, characteristic absorption data are provided for each compound. High resolution mass spectra were performed on an Agilent 6210 TOF LC/MS using ESI-MS or were carried out by the Notre Dame University Mass Spectrometry Department using EI technique. Optical rotation was measured on a Perkin Elmer 341 Polarimeter at 589 nm.

(*R*)-3-((triisopropylsilyl)oxy)pent-4-enal (11): To a solution of (2*R*,3*R*)-2-iodomethyl-5-methoxytetrahydrofuran-3-ol[24] (5.00 g, 19.4 mmol) in DMF (32 mL) was added imidazole (2.90 g, 42.6 mmol), 4-dimethylaminopyridine (23 mg, 0.194 mmol), and TIPS-Cl (4.56 mL, 21.3 mmol). The mixture was stirred for 16 h and then was quenched with H_2O and the mixture was extracted with Et_2O. The combined organic layers were washed with brine and then were dried over Na_2SO_4. The solvents were removed *in vacuo* and the residue was then purified by flash column chromatography (CH_2Cl_2:pentane, 3:2) to yield acetal as a colorless oil (6.42 g, 80%). To the above acetal in THF/H_2O (4/1, 70 mL) was added Zn dust (10.14 g, 15.5 mmol). The resulting cloudy suspension was refluxed for 2 h, cooled to room temperature, and filtered through a Celite® pad (diethyl ether rinse). The solution was further diluted with diethyl ether and was washed with brine and then were dried over Na_2SO_4. The solvents were removed *in vacuo* to yield **11** as a colorless oil without further purification (3.97 g, 100%).

IR (neat): 2945, 2868, 1727, 1466, 1099 cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$): δ 9.81 (t, *J* = 2.5 Hz, 1H), 5.92 (ddd, *J* = 17.1, 10.4, 6.1 Hz, 1H), 5.28 (apparent dt, *J* = 17.1, 1.2 Hz, 1H), 5.14 (apparent dt, *J* = 10.2, 1.2 Hz, 1H), 4.78-4.74 (m, 1H), 2.62 (dd, *J* = 5.6, 2.5 Hz, 2H), 1.08-1.04 (m, 21H); ^{13}C NMR (151 MHz, $CDCl_3$) δ 201.9, 140.3, 115.2, 70.0, 51.7, 18.17, 18.15, 12.4; HRMS (ESI): m/z $[M + H]^+$ calcd for $C_{14}H_{29}O_2Si$: 257.1931; found: 257.1928; $[\alpha]_D^{20}$ ($CHCl_3$, c = 0.22): –13.9.

(*R*)-4-((1*S*,2*R*,3*S*)-2-chloro-1-hydroxy-3-((triisopropylsilyl)oxy)pent-4-en-1-yl)-2,2-dimethyl-1,3-dioxan-5-one (9): To a solution of **11** (4.66 g, 18.2 mmol) in CH_2Cl_2 (90 mL) were added (*R*)-proline (1.62 g, 14.0 mmol), *N*-chlorosuccinimide (2.12 g, 15.9 mmol), and 2,2-dimethyl-1,3-dioxan-5-one (2.20 mL, 18.7 mmol). The mixture was stirred at ambient temperature for 24 h and then was quenched with H_2O. The mixture was extracted with Et_2O and the combined organic layers were washed with brine and then were dried over Na_2SO_4. The solvents were removed *in vacuo* and the residue was then purified by flash column chromatography (pentane:diethyl ether, 8:1) to yield **9** as a colorless oil (4.59 g, 60%).

IR (neat): 3538, 2944, 2868, 1738, 1223, 1086 cm^{-1}; ^1H NMR (600 MHz, $CDCl_3$) δ: 5.87 (ddd, *J* = 17.2, 10.3, 7.8 Hz, 1H), 5.34 (apparent dt, *J* = 17.2, 1.0 Hz, 1H), 5.28

(apparent dt, *J* = 10.3, 0.9 Hz, 1H), 4.57 (apparent t, *J* = 7.7 Hz, 1H), 4.46 (ddd, *J* = 8.9, 2.5, 1.5 Hz, 1H), 4.37 (dd, *J* = 8.9, 1.3 Hz, 1H), 4.29 (dd *J* = 17.6, 1.4 Hz, 1H), 4.08 (d, *J* = 17.6 Hz, 1H), 3.99 (dd, *J* = 7.5, 1.0 Hz, 1H), 3.48 (dd, *J* = 2.5, 0.9 Hz, 1H), 1.50 (s, 3H), 1.42 (s, 3H), 1.08-1.05 (m, 21H); ^{13}C NMR (151 MHz, $CDCl_3$) δ: 211.4, 138.4, 118.6, 101.7, 75.6, 72.8, 67.9, 66.6, 63.5, 24.0, 23.5, 18.15, 18.13, 12.6; HRMS (ESI): m/z $[M + H]^+$ calcd for $C_{20}H_{38}ClO_5Si$: 421.2172; found: 421.2188; $[\alpha]_D^{20}$ ($CHCl_3$, c = 0.89): +88.5.

(1*S*,2*R*,3*S*)-2-chloro-1-((*S*)-2,2-dimethyl-5-methylene-1,3-dioxan-4-yl)-3-((triisopropylsilyl)oxy)pent-4-en-1-ol (7): To a cooled (–78 °C) solution of 5-(methanesulfonyl)-1-phenyl-1*H*-tetrazole (4.62 g, 20. 8 mmol) in THF (60 mL) was added dropwise LiHMDS (20.8 mL, 1.0 M in THF, 20.8 mmol) and stirred at –78 °C for 30 min. To this yellow solution **9** (4.40 g, 10.4 mmol) in THF (20 mL) was added dropwise at –78 °C and stirred for another 1 h before quenching with H_2O. The mixture was extracted with Et_2O and the combined organic layers were washed with brine and then were dried over Na_2SO_4. The solvents were removed *in vacuo* and the residue was then purified by flash column chromatography (pentane:diethyl ether, 12:1) to yield **7** as a colorless oil (3.24 g, 74%).

IR (neat): 3485, 2968, 1380, 1228, 1067 cm^{-1}; ^{1}H NMR (600 MHz, $CDCl_3$) δ: 5.95 (ddd, *J* = 17.3, 10.4, 7.1 Hz, 1H), 5.36 (m, 1H), 5.35 (apparent dt, *J* = 17.3, 1.0 Hz, 1H), 5.29 (apparent dt, *J* = 10.4, 1.0 Hz, 1H), 5.02 (brs, 1H), 4.70 (ddt, 7.1, 4.1, 1.0 Hz, 1H), 4.36–4.30 (m, 4H), 4.25 (d, *J* = 13.5 Hz, 1H) 3.54 (d, 2.9 Hz, 1H), 1.48 (s, 3H), 1.34 (s, 3H), 1.14–1.08 (m, 21H); ^{13}C NMR (151 MHz, $CDCl_3$) δ: 142.0, 138.1, 118.3, 109.9, 99.7, 79.1, 71.3, 70.8, 65.1, 64.1, 28.3, 21.8, 18.2(0), 18.1(8), 12.6; HRMS (ESI): m/z $[M + H]^+$ calcd for $C_{21}H_{40}ClO_4Si$: 419.2367; found: 419.2379; $[\alpha]_D^{20}$ ($CHCl_3$, c = 0.45): +22.2.

(*S*)-1-((2*R*,3*S*)-3-((*S*)-2,2-dimethyl-5-methylene-1,3-dioxan-4-yl)oxiran-2-yl)prop-2-en-1-ol (15): To a solution of **7** (9.11 g, 21.7 mmol) in EtOH/H_2O (5/1, 150 mL) was added CsOH (50% w/w in H_2O, 21.0 mL, 109 mmol). The resulting mixture was heated to 80 °C and was stirred for 3 h, then cooled to room temperature. The mixture was extracted with Et_2O and the combined organic layers were washed with $NaHCO_3$ (aq.) and brine, dried over Na_2SO_4. The solvents were removed *in vacuo* and the residue was then purified by flash column chromatography (pentane:diethyl ether, 10:1) to yield the

TIPS protected epoxide as a colorless oil (5.91 g, 71%). To a solution of this epoxide in THF (30 mL) was added tetrabutylammonium fluoride (19.1 mL, 1.0 M in THF, 19.1 mmol). The reaction was stirred at ambient temperature for 1 h and was then purified by flash column chromatography (pentane:ethyl acetate, 2:1) to yield **15** as a white solid (3.43 g, 98%).

IR (neat): 3445, 2991, 1372, 1222, 1199, 1158, 1084, 1002 cm^{-1}; ^{1}H NMR (400 MHz, $CDCl_3$) δ: 6.00 (ddd, *J* = 17.4, 10.7, 4.5 Hz, 1H), 5.48 (dt, *J* = 17.4, 1.5 Hz, 1H), 5.28 (dt, *J* = 10.7, 1.5 Hz, 1H), 5.21–5.19 (m, 1H), 5.04–5.02 (m, 1H), 4.39 (d, *J* = 14.0 Hz, 1H), 4.29 (d, *J* = 14.0 Hz, 1H), 4.26 (d, *J* = 7.8 Hz, 1H), 4.18–4.16 (m, 1H), 3.26 (dd, *J* = 7.8, 4.2 Hz, 1H), 3.03 (dd, *J* = 7.8, 4.2 Hz, 1H), 2.24 (d, *J* = 3.8 Hz, 1H), 1.44 (s, 3H), 1.40 (s, 3H); ^{13}C NMR (101 MHz, $CDCl_3$) δ: 142.4, 135.9, 116.6, 109.1, 99.4, 70.9, 69.8, 64.0, 58.8, 57.3, 27.6, 21.4; HRMS (ESI): m/z $[M + Na]^+$ calcd for $C_{12}H_{18}O_4Na$: 249.1097; found: 249.1111; m.p. = 60–61 °C; $[\alpha]_D^{20}$ ($CHCl_3$, c = 0.68): –15.0.

(6*S*,6a*R*,7a*S*,7b*S*)-2,2-dimethyl-6,6a,7a,7b-tetrahydro-4*H*-oxireno[2',3':5,6]benzo[1,2-*d*][1,3]dioxin-6-ol (16): To a solution of **15** (300 mg, 1.33 mmol) in CH_2Cl_2 (40 mL) was added Stewart-Grubbs' catalyst (30 mg, 0.053 mmol). The mixture was refluxed under argon for 72 h. The reaction was cooled to room temperature and concentrated *in vacuo*. The residue was then purified by flash column chromatography (pentane:ethyl acetate, 1:1) to yield **16** as a white solid (240 mg, 91%).

IR (neat): 3424, 2989, 1382, 1223, 1198, 1072, 1013 cm^{-1}; ^{1}H NMR (600 MHz, $CDCl_3$) δ: 5.47–5.46 (m, 1H), 4.82 (brs, 1H), 4.53 (brs, 1H), 4.37 (d, *J* = 14.4 Hz, 1H), 4.17 (d, *J* = 14.4 Hz, 1H), 3.44-3.43 (m, 1H), 3.39–3.37 (m, 1H), 2.28 (br. d, *J* = 4.4 Hz, 1H), 1.51 (s, 3H), 1.43 (s, 3H); ^{13}C NMR (151 MHz, $CDCl_3$) δ: 134.1, 118.0, 100.4, 65.4, 63.5, 62.5, 53.6, 51.9, 27.0, 21.3; HRMS (ESI): m/z $[M + H]^+$ calcd for $C_{10}H_{15}O_4$: 199.0965; found: 199.0973; m.p. = 81–82 °C; $[\alpha]_D^{20}$ ($CHCl_3$, c = 0.78): +95.0.

(3a*R*,4*S*,4a*S*,9a*S*)-4-hydroxy-6,6-dimethyl-3a,4,4a,9a-tetrahydro-8*H*-[1,3]dioxolo[4',5':4,5]benzo[1,2-*d*][1,3]dioxin-2-one (18): A mixture of Cs_2CO_3 (326 mg, 1.0 mmol) and 3Å molecular sieves powder (160 mg) was heated under vaccum for 5 min, then blanketed with CO_2 (g) and cooled to room temperature. A solution of **16** (198 mg, 1.0 mmol) in DMF (2 mL) was then added. The resulting light brown solution was heated

to 45 °C and was stirred for 18 h and was then quenched with NH_4Cl (aq.). The mixture was extracted with Et_2O and the combined organic layers were washed with brine and then were dried over Na_2SO_4. The solvents were removed *in vacuo* and the residue was then purified by flash column chromatography (pentane:ethyl acetate, 3:1) to yield **18** as a white solid (230 mg, 95%).

IR (neat): 3479, 2942, 1803, 1383, 1163, 1043 cm^{-1}; ^1H NMR (400 MHz, $CDCl_3$) δ: 5.60-5.57 (m, 1H), 5.20-5.18 (m, 1H), 4.98 (dd, *J* = 6.8, 3.8 Hz, 1H), 4.66–4.65 (m, 1H), 4.52–4.79 (m, 1H), 4.43 (t, *J* = 3.8 Hz, 1H), 4.21 (dd, *J* = 14.8, 0.8 Hz, 1H), 2.69 (brs, 1H), 1.55 (s, 3H), 1.44 (s, 3H); ^{13}C NMR (101 MHz, $CDCl_3$) δ: 153.9, 136.8, 114.7, 100.2, 74.4, 72.1, 65.9, 64.7, 62.5, 27.6, 21.0; HRMS (ESI): m/z $[M + H]^+$ calcd for $C_{11}H_{15}O_6$: 243.0863; found: 243.0863; m.p. = 156–157 °C; $[\alpha]_D^{20}$ ($CHCl_3$, c = 0.19): +35.0.

(1*R*,2*S*,3*S*,6*S*)-6-(2,4-dinitrophenoxy)-4-(hydroxymethyl)cyclohex-4-ene-1,2,3-triol (3): To a solution of **18** (48.5 mg, 0.2 mmol) in THF/methanol (1/1, 2 mL) at 0 °C was added K_2CO_3 (27.6 mg, 0.2 mmol). The resulting mixture was stirred at 0 °C for 1 h and then filtered through a pad of silica gel. The solvents were removed *in vacuo* and the residue was dissolved in DMF (0.8 mL). Quinuclidine (111 mg, 1.0 mmol) and 4Å molecular sieves (10 beads) were added and the resulting solution was stirred at ambient temperature for 30 min. Then a solution of 2,4-dinitrofluorobenzene (37.2 mg, 0.2 mmol) in DMF (0.2 mL) was added dropwise. The reaction was stirred at ambient temperature for 12 h and then cooled to 0 °C. Methanol (2 mL) was added, followed by aqueous HCl (1.0 M) to pH~3. The reaction was stirred at 0 °C for 20 min and was quickly neutralized by adding trimethylamine, and then purified by flash column chromatography (CH_2Cl_2: methanol, 12:1) to yield **3** as a white foam (12.3 mg, 18%).

IR (neat): 3361, 2930, 1611, 1520, 1348, 1076 cm^{-1}; ^1H NMR (600 MHz, CD_3OD) δ: 8.70 (d, *J* = 2.8 Hz, 1H), 8.45 (dd, *J* = 9.4, 2.8 Hz, 1H), 7.69 (d, *J* = 9.4 Hz, 1H), 5.99–5.98 (m, 1H), 5.42 (t, *J* = 4.2 Hz, 1H), 4.29 (d, *J* = 4.2 Hz, 1H), 4.22 (d, *J* = 15.0 Hz, 1H), 4.19 (dd, *J* = 9.7, 3.8 Hz, 1H), 4.16 (d, *J* = 15.0 Hz, 1H), 4.05 (dd, *J* = 9.7, 4.2 Hz, 1H); ^{13}C NMR (151 MHz, CD_3OD) δ: 158.1, 147.0, 141.3, 140.9, 129.7, 122.5, 118.8, 117.9, 77.8, 70.4, 69.6, 68.1, 63.5; HRMS (ESI): m/z $[M + Na]^+$ calcd for $C_{13}H_{14}N_2NaO_9$: 365.0592; found: 365.0583; $[\alpha]_D^{20}$ (CH_3OH, c = 0.31): +121.7.

(1*S*,2*R*,3*S*)-1-((*S*)-2,2-dimethyl-5-methylene-1,3-dioxan-4-yl)-2-fluoro-3-((triisopropylsilyl)oxy)pent-4-en-1-ol (8): To a solution of **11** (260 mg, 1.0 mmol) in DMF (10 mL) at 5 °C were added Selectfluor® (350 mg, 1.0 mmol) and (*R*)-proline (115 mg, 1.0 mmol). The mixture was stirred at 5 °C for 1 h and then was quenched with H_2O and the mixture was extracted with Et_2O. The combined organic layers were washed with brine and then were dried over Na_2SO_4. The solvents were removed *in vacuo* and the residue was redissolved in CH_2Cl_2 (5 mL). (*R*)-proline (92 mg, 0.8 mmol) and 2,2-dimethyl-1,3-dioxan-5-one (156 mg, 1.2 mmol) were added at 0 °C. The mixture was warmed up to room temperature and stirred for 48 h. The reaction was then quenched with H_2O and was extracted with Et_2O. The combined organic layers were washed with brine and then were dried over Na_2SO_4. The solvents were removed *in vacuo* and the residue was dissolved in THF (3 mL). In another flask LiHMDS (2.0 mL, 1.0 M in THF, 2.0 mmol) was added dropwise to a cooled (–78 °C) solution of 5-(methanesulfonyl)-1-phenyl-1*H*-tetrazole (444 mg, 2.0 mmol) in THF (7 mL) and stirred at –78 °C for 30 min. Then the above solution (3 mL) was added dropwise at –78 °C and stirred for another 1 h before quenching with H_2O. The mixture was extracted with Et_2O and the combined organic layers were washed with brine and then were dried over Na_2SO_4. The solvents were removed *in vacuo* and the residue was then purified by flash column chromatography (pentane:diethyl ether, 15:1) to yield **8** as a colorless oil (161 mg, 40%).

IR (neat): 3478, 2944, 2868, 1464, 1381, 1096, 1071 cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$): δ 5.88 (ddd, *J* = 17.2, 10.5, 6.3 Hz, 1H), 5.42 (apparent d, *J* = 17.2 Hz, 1H), 5.33–5.31 (m, 2H), 5.04 (brs, 1H), 4.86–4.83 (m, 1H), 4.67 (dd, *J* = 44.2, 3.7 Hz, 1H), 4.43 (d, *J* = 8.5 Hz, 1H), 4.36 (d, *J* = 13.0 Hz, 1H), 4.26 (d, *J* = 13.2 Hz, 1H), 4.14 (ddd, *J* = 29.0, 8.7, 2.0 Hz, 1H), 4.11-4.09 (m, 1H), 1.49 (s, 3H), 1.34 (s, 3H), 1.11–1.05 (m, 21H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 142.1, 136.4 (d, $J_{C\text{-}F}$ = 7.8 Hz), 118.2 (d, $J_{C\text{-}F}$ = 1.5 Hz), 109.9, 99.6, 90.1 (d, $J_{C\text{-}F}$ = 185.1 Hz), 76.6 (d, $J_{C\text{-}F}$ = 22.8 Hz), 70.8 (d, $J_{C\text{-}F}$ = 18.3 Hz), 70.5(d, $J_{C\text{-}F}$ = 3.8 Hz), 65.1, 28.3, 22.0, 18.03, 18.02, 12.4; HRMS (ESI): m/z $[M + Na]^+$ calcd for $C_{21}H_{39}FNaO_4Si$: 425.2494; found: 425.2495; $[\alpha]_D^{20}$ ($CHCl_3$, c = 0.45): +19.8.

(1*S*,2*R*,3*S*)-1-((*S*)-2,2-dimethyl-5-methylene-1,3-dioxan-4-yl)-2-fluoro-3-((triisopropylsilyl)oxy)pent-4-en-1-yl acetate (19): To a solution of **8** (201 mg, 0.5 mmol) in CH_2Cl_2 (5 mL) at ambient temperature were added triethylamine (139 µL, 1.0

mmol), acetic anhydride (71 µL, 0.75 mmol), and 4-dimethylaminopyridine (6.1 mg, 0.05 mmol). The reaction was stirred at ambient temperature for 48 h and then was quenched with NH_4Cl (aq.). The mixture was extracted with Et_2O and the combined organic layers were washed with brine and then were dried over Na_2SO_4. The solvents were removed *in vacuo* and the residue was then purified by flash column chromatography (pentane:diethyl ether, 8:1) to yield **19** as a colorless oil (186 mg, 84%).

IR (neat): 2944, 2868, 1749, 1464, 1371, 1232, 1094, 1040 cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$): δ 5.88 (ddd, *J* = 17.2, 10.1, 7.9 Hz, 1H), 5.44 (ddd, *J* = 27.1, 9.0, 1.0 Hz, 1H), 5.28 (d, *J* = 17.2 Hz, 1H), 5.26 (d, *J* = 10.1 Hz, 1H), 4.99 (brs, 1H), 4.97 (brs, 1H), 4.81 (ddd, *J* = 44.9, 6.0, 1.0 Hz, 1H), 4.55-4.52 (m, 1H), 4.45 (d, *J* = 8.9 Hz, 1H), 4.37-4.32 (m, 1H), 4.27 (d, *J* = 13.8 Hz, 1H), 2.02 (s, 3H), 1.45 (s, 3H), 1.38 (s, 3H), 1.05 (brs, 21H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.8, 141.1, 137.2 (d, $J_{C\text{-}F}$ = 4.0 Hz), 118.5 (d, $J_{C\text{-}F}$ = 1.6 Hz), 111.5, 99.6, 91.5 (d, $J_{C\text{-}F}$ = 184.5 Hz), 74.0 (d, $J_{C\text{-}F}$ = 25.4 Hz), 72.1(d, $J_{C\text{-}F}$ = 4.0 Hz), 70.4 (d, $J_{C\text{-}F}$ = 16.3 Hz), 64.0, 28.8, 24.9, 21.1, 18.2, 12.7; HRMS (ESI): m/z $[M + Na]^+$ calcd for $C_{23}H_{41}FNaO_5Si$: 467.2600; found: 467.2595; $[\alpha]_D^{20}$ ($CHCl_3$, c = 0.9): +43.2.

(6*S*,7*S*,8*S*,8a*S*)-7-fluoro-6-hydroxy-2,2-dimethyl-6,7,8,8a-tetrahydro-4H-benzo[d][1,3]dioxin-8-yl acetate (20): To a solution of **19** (160 mg, 0.36 mmol) in THF (3.6 mL) at 0 °C was added a solution of tetrabutylammonium fluoride (0.72 mL, 1.0 M in THF, 0.72 mmol) and acetic acid (43 µL, 0.72 mmol). The reaction was stirred at ambient temperature for 48 h and was then purified by flash column chromatography (pentane:ethyl acetate, 3:1) to yield a 10:1 mixture of the desired deprotection product and acyl migration compound. The mixture was redissolved in DCM (18 mL) and Grubbs' II catalyst (31 mg, 0.036 mmol) was then added. The mixture was heated to 40 °C under argon for 1 h. The reaction was cooled to room temperature and concentrated *in vacuo*. The residue was then purified by flash column chromatography (pentane:ethyl acetate, 1:1.5) to yield **20** as a yellow oil (82 mg, 88%).

IR (neat): 3449, 2992, 2926, 1749, 1377, 1234, 1097, 1064 cm^{-1}; ^1H NMR (400 MHz, $CDCl_3$): δ 5.59 (ddd, *J* = 6.0, 4.2, 1.8 Hz, 1H), 5.53 (brs, 1H), 4.89 (dddd, *J* = 48.9, 5.8, 3.4, 1.1 Hz, 1H), 4.71 (brs, 1H), 4.51–4.47 (m, 1H), 4.45-4.37 (m, 1H), 4.15 (d, *J* = 14.4 Hz, 1H), 2.09 (brs, 1H), 2.09 (s, 3H), 1.51 (s, 3H), 1.37 (s, 3H); ^{13}C NMR (101 MHz,

$CDCl_3$) δ 170.1, 133.1, 120.3 (d, J_{C-F} = 2.1 Hz), 99.7, 88.1 (d, J_{C-F} = 174.9 Hz), 68.0 (d, J_{C-F} = 26.3 Hz), 65.3 (d, J_{C-F} = 18.5 Hz), 64.4 (d, J_{C-F} = 3.5 Hz), 63.3, 28.3, 20.9, 20.1; HRMS (ESI): m/z $[M + Na]^+$ calcd for $C_{12}H_{17}FNaO_5$: 283.0952; found: 283.0954; $[\alpha]_D^{20}$ ($CHCl_3$, c = 0.78): +104.3.

(1*S*,2*S*,5*S*,6*R*)-5-(2,4-dinitrophenoxy)-6-fluoro-3-(hydroxymethyl)cyclohex-3-ene-1,2-diol (4): To a solution of **20** (9.4 mg, 0.036 mmol) in DMF (0.18 mL) was added quinuclidine (20 mg, 0.18 mmol) and 4Å molecular sieves (2 beads). The resulting solution was stirred at ambient temperature for 30 min. Then a solution of 2,4-dinitrofluorobenzene (7.4 mg, 0.040 mmol) in DMF (0.1 mL) was added dropwise. The reaction was stirred at ambient temperature for 12 h and then methanol (0.4 mL) was added, followed by K_2CO_3 (7.5 mg, 0.054 mmol). The reaction was stirred at ambient temperature for another 1 h and was cooled to 0 °C, acidified with aqueous HCl (1.0 M) to pH~3. The reaction was stirred at 0 °C for 15 min and was quickly neutralized by adding trimethylamine, and then was purified by flash column chromatography (CH_2Cl_2: methanol, 20:1) to yield **4** as a white solid (8.2 mg, 66%).

IR (neat): 3363, 2926, 1605, 1532, 1347, 1279, 1068 cm^{-1}; ^1H NMR (600 MHz, CD_3OD): δ 8.69 (d, *J* = 1.5 Hz, 1H), 8.46 (dd, *J* = 9.4, 1.5 Hz, 1H), 7.69 (d, *J* = 9.6 Hz, 1H), 5.98 (brs, 1H), 5.64 (dd, *J* = 4.1, 3.6 Hz, 1H), 5.01 (ddd, *J* = 49.0, 10.2, 3.6 Hz, 1H), 4.31 (apparent t, *J* = 4.0 Hz, 1H), 4.28-4.24 (m, 1H), 4.24 (d, *J* = 15.3 Hz, 1H), 4.15 (d, *J* = 15.3 Hz, 1H); ^{13}C NMR (151 MHz, CD_3OD) δ 157.4, 148.2, 141.7, 141.0, 129.7, 122.3, 117.7, 117.5 (d, J_{C-F} = 4.4 Hz), 90.4 (d, J_{C-F} = 185.3 Hz), 76.3 (d, J_{C-F} = 16.5 Hz), 68.5 (d, J_{C-F} = 10.3 Hz), 68.4, 63.4; HRMS (ESI): m/z $[M + Na]^+$ calcd for $C_{13}H_{13}FN_2NaO_2$: 367.0548; found: 367.0550; m.p. = 158–159 °C; $[\alpha]_D^{20}$ (CH_3OH, c = 0.17): +190.0.

4.2.2. Protein expression and purification

A mutation was introduced into the plasmid that encoded *Tm*GalA harbouring a D387A mutation, using the QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA) with primers, ForD387A, 5′-GATGAGGATAGGACCTGCTACTGCGCCGTTCTGGG-3′ and RevD387A, 5′-CCCAGAACGGCGCAGTAGCAGGTCCTATCCTCATC-3′). The plasmid encoding either

α-galactosidase *Tm*GalA or *Tm*GalA with the D387A mutation were transformed into *E. coli* BL-21(DE3) cells. *Tm*GalA and the D387A mutant were recombinantly expressed and purified as described previously.[5]

4.2.3. Enzyme kinetics

T. maritima α-galactosidase was recombinantly expressed and purified as described previously.[5] Michaelis–Menten kinetic parameter for the hydrolysis of the allylic carbasugar mimics of galactose (**3**) and 2-deoxy-2-fluorogalactose (**4**) were determined from a minimum of six initial rate measurements using a concentration range of at least $K_m/4$ to $4 \times K_m$. The progress of each reaction was continuously monitored for 5 min at 400 nm using a Cary Eclipse Fluorescence Spectrometer equipped with a temperature controller. Each 500 µL reaction mixture was prepared by addition of the appropriate volume of buffer (50 mM HEPES buffer, pH 7.4, T = 37 °C), substrate and enzyme. The rate versus substrate concentration data were fit to a standard Michaelis-Menten equation using a standard nonlinear least-squares computer program (Prizm 7.0).

All inactivation experiments were performed at 37 °C in 50 mM HEPES buffer, pH 7.4 using an Applied Photophysics SX20 stopped-flow spectrophotometer, equipped with an external temperature controller. The stopped-flow spectrometer was used in the sequential double mixing mode, in which rapidly mixed enzyme and inhibitor **4** were incubated for variable time intervals prior to rapid mixing of the enzyme/inhibitor solution with a buffered solution of 4-nitrophenyl α-D-galactopyranoside. The remaining enzyme activity was monitored at a wavelength of 400 nm. Pseudo first-order rate constants for inactivation (k_{obs}) at each inhibitor concentration were calculated by fitting the absorbance versus time data to a standard first-order rate equation using a nonlinear least squares routine in a computer program (Prizm 7.0).

4.2.4. Crystallization

*Tm*GalA (10 mg/ml) was crystallised from 0.2 M $MgSO_4$ and 20% (w/v) poly(ethylene glycol) (PEG) 3350, and crystals were soaked in 0.2 M $MgSO_4$ and 30% (w/v) PEG 3350 containing a minute amount of **4** added as powder for either 2 minutes or

1 hour. The *Tm*GalA mutant D387A (10 mg/ml) was incubated with ~1 mM **4** and co-crystallised from 0.2 M $MgSO_4$ and 20% (w/v) PEG 3350. All crystals were cryo-protected in 0.2 M $MgSO_4$ and 30% (w/v) PEG 3350 prior to vitrification in liquid nitrogen.

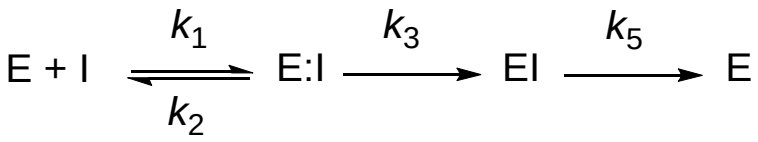
4.2.5. Data collection and processing

X-ray diffraction data were collected at Diamond Light Source (DLS) or the European Synchrotron Radiation Facility (ESRF) beamlines and the experimental details are listed in Table S1 (Supporting Information). Diffraction data were processed using the Xia2[25] pipeline to run DIALS or XDS[26] with Aimless[27] from the CCP4 suite[28]. Molecular replacement was performed using MOLREP[29] or Phaser[30] with Protein Data Bank (PDB) entry 5M0X as the search model. Refinement was performed using REFMAC5[31] and manual model building was done using Coot.[32] Structures were validated using PDB_REDO.[33] Models for the inhibitors were built in JSME[34] and the library generated with PRODRG.[35]

4.3. Results and Discussion

4.3.1. Synthesis of Dinitrophenyl Allylic Galactose Analogues 3 and 4.

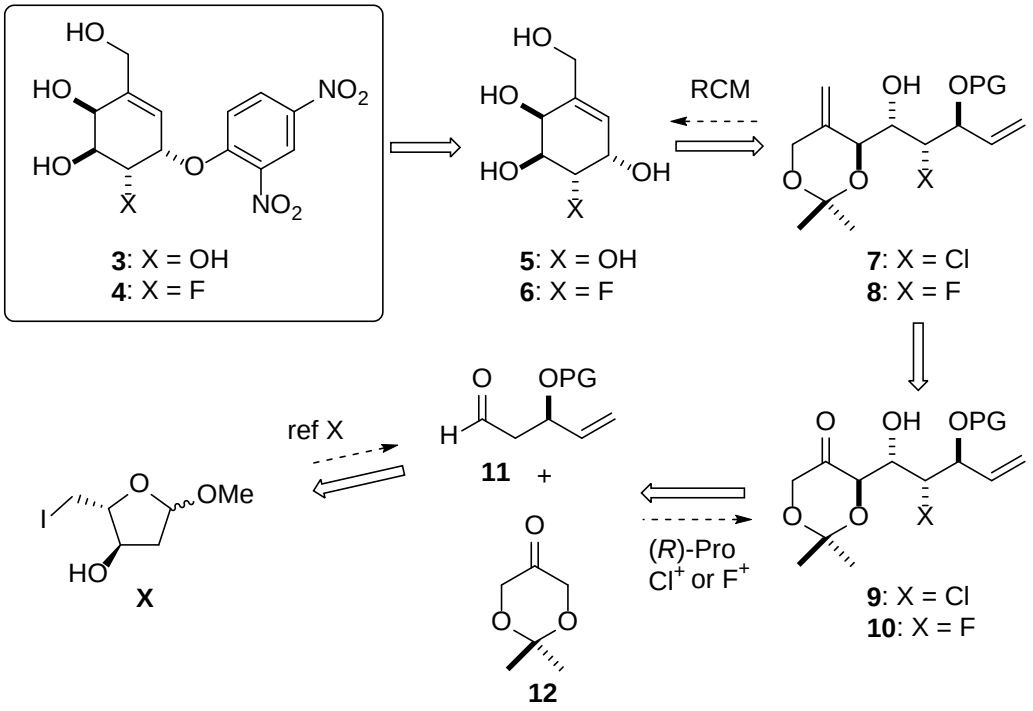
We have previously reported that the allylic carbasugar **1** is a covalent inhibitor of yeast α-glucosidase (GH13 family), and that its reaction coordinate likely passes through a $^4H_3$ glucopyranosylium ion-like TS.[13,14,36] As a result, we hypothesized that a *galacto*-configured analogue would covalently label a GH36 □-galactosidase in which the TS for glycosylation has a presumed $^4H_3$ conformation (based on a homology with GH27).[14,19] Although members from families GH13, GH27, and GH36 bind bicyclo[4.1.0]heptyl amines (with the appropriate stereochemistry) tightly,[37,38] these retaining families are reported to have different pyranosyl ring conformations of the covalent glycosyl-enzyme intermediate. That is, for GH13 enzymes the structure of the enzyme bound intermediate is a $^4C_1$ chair,[39] while that for GH27 enzymes (member of GH clan-D, as is the GH36 family) is a $^1S_3$ skew-boat.[40] The kinetic scheme for this type of covalent inhibitor is shown in Scheme 1.



Scheme 1. Kinetic scheme for the covalent inhibition of α-glycosidases by allylic carbasugar analogues.

Previously, we were unable to identify the amino acid that was covalently labeled during the reaction of yeast α-glucosidase with **1** (the ratio of the two first-order rate constants k_{inact}/k_{react} ~ 75) using mass spectroscopic techniques.[7] In an effort to structurally characterize the covalent intermediate, here we aimed to enhance the leaving group ability of the pseudo aglycone. That is, we targeted the *galacto* configured 2,4-dinitrophenyl ether **3**, and the corresponding 2-fluorinated analogue **4**, with the expectation of an increased lifetime for the fluorine-containing covalent intermediate.[21,23]

Unfortunately, our previous synthesis of allylic D-*gluco*-carbasugar analogues **1**[7] and that of Danby and Withers to make **2**,[8] which used glucose or gluconolactone as starting materials, could not be readily modified to provide access to the parent or fluorinated D-*galacto*-analogues **3** and **4**. We therefore initiated separate *de novo* syntheses of the D-*galacto*-configured 2-hydroxy and 2-fluoro carabasugar analogues **3** and **4**, respectively. As depicted in Scheme 2, we expected that both compounds could be accessed through the arylation of a suitably protected allylic alcohol (e.g., **5** or **6**) derived from the ring closing metathesis of halohydrins **7** or **8**, respectively.

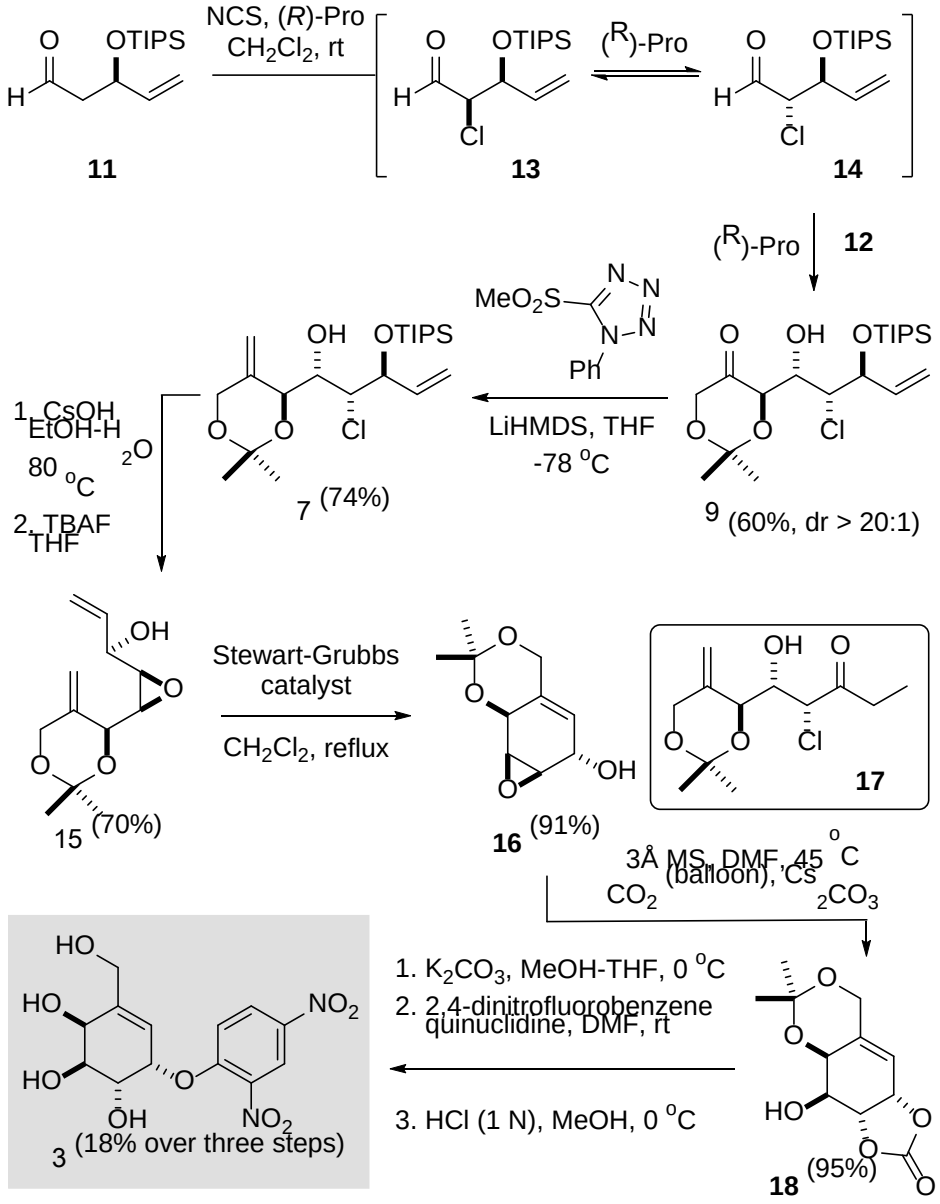


Scheme 2. A concise synthetic plan for carbasugars 3 and 4 relies on ring closing metathesis reactions to access the carbocyclic core and an α-halogenation/aldol reaction to introduce the correct absolute stereochemistry of the functional groups

The 2-hydroxy function in carbasugar **3** could then be installed through a double inversion process involving epoxide formation followed by regioselective hydroxy-directed epoxide opening. Notably, synthesis of both the fluorohydrin **10** and chlorohydrin **9** would take advantage of one-pot organocatalytic α-halogenation/dynamic kinetic asymmetric aldol reactions to establish three of the required stereocenters and introduce the necessary halide in a single operation (e.g., **11** + **12** → **9**). The precursor for the common aldehyde starting material **11** can be prepared from 2-deoxy-L-ribose following an established process.[24]

Synthesis of the 2-hydroxycarbasugar **3** started with 2-deoxy-L-ribose-derived aldehyde **11**, which was subjected to a proline-catalyzed α-chlorination-aldol reaction (Scheme 3). Here, the diastereomeric α-chloroaldehydes **13** and **14** are formed *in situ* through the proline-catalyzed α-chlorination of **11**. Importantly, proline also catalyzes the epimerization of the chloromethine center (i.e., equilibration of **13** and **14**) and

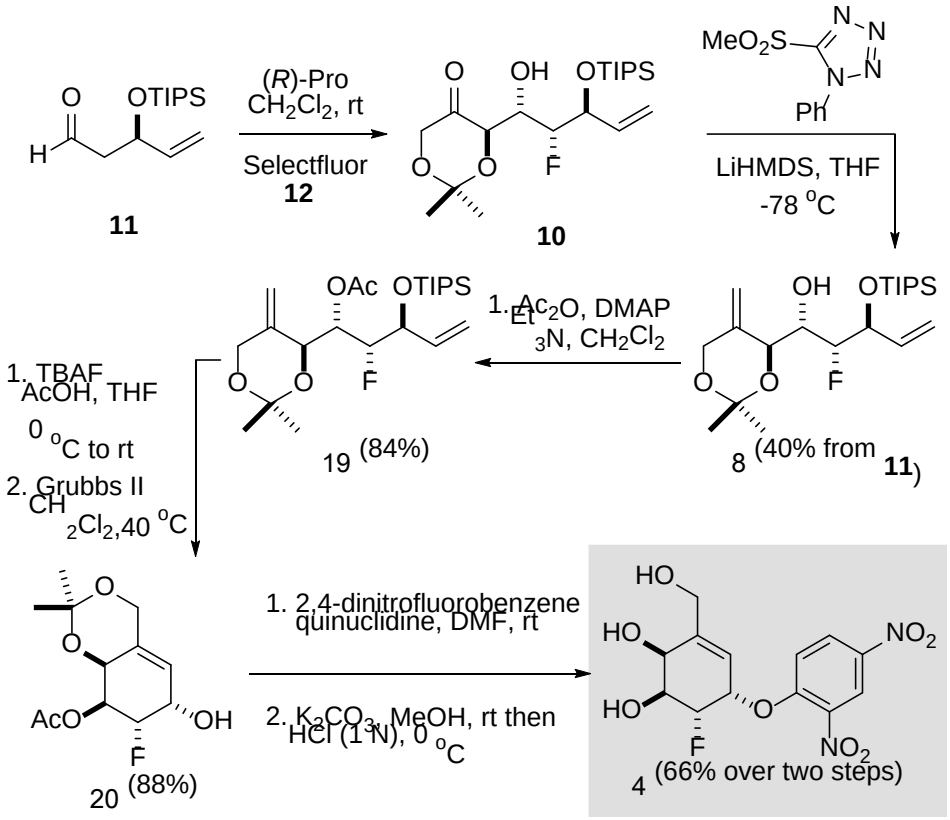
consequently, the subsequent proline catalyzed aldol reaction with dioxanone **12** effects a dynamic kinetic resolution that delivers the *anti*-aldol-*syn*-chlorohydrin **9** in good yield and excellent diastereoselectivity (dr >20:1).



Scheme 3. Synthesis of the carbagalactose analogue 3.

Here, we found that it was critical to employ the bulky TIPS group to protect the secondary alcohol in **11**, as we noted that equivalent reactions involving *para*-methoxybenzyl, benzyl or TBS protected alcohols were complicated by elimination of the alkoxy or silyloxy group from the starting material to give an α,β-unsaturated aldehyde. With chlorohydrin **9** in hand, several olefination methods were examined from which we found that optimally a Julia-Kocienski reaction using the lithium anion derived from methylsulfonyl phenyltetrazole delivered the diene **7** in excellent yield without requiring protection of the newly-formed secondary alcohol function. Unfortunately, attempts to effect a ring closing metathesis (RCM) on this material using various catalysts returned only starting material. Also, removal of the TIPS protecting group (TBAF) followed by treatment with Grubb's first or second-generation catalyst resulted primarily in isomerization to ethyl ketone **17**. In an effort to circumvent these problems and facilitate the desired RCM reaction, we converted chlorohydrin **7** into the corresponding *syn*-epoxide **15** by treatment with CsOH in hot EtOH-H_2O. While the TIPS-protected epoxyalcohol was not a productive substrate for a RCM, following the removal of the silyl protecting group the corresponding alcohol **15** was readily transformed into cyclohexene **16** using either Stewart-Grubb's or Grubb's second generation catalyst. From here, conversion into the protected carbasugar simply involved reaction of the epoxy alcohol **16** with CO_2 and $CsCO_3$ in warm DMF, which afforded the carbonate **18** in excellent yield. Notably, following this optimized route, the densely functionalized carbasugar **18** was accessed in only 6 steps (~25% overall yield) from the readily available aldehyde **11**. Finally, removal of the carbonate and reaction of the resulting triol with 2,4-dinitrofluorobenzene afforded a mixture of arylated carbasugars from which the targeted dinitrophenyl adduct **3** could be isolated by preparative thin layer chromatography.

The corresponding 2-fluorocarbasugar **4** was accessed following a similar route to that described above (Scheme 4). Here, however, we modified our previously reported α-chlorination aldol reaction by first effecting a proline-catalyzed α-fluorination of aldehyde **11** using Selectfluor in DMF [RB-ref] followed directly by reaction of the resulting unstable α-fluoroaldehyde (not shown) with dioxanone **12** in CH_2Cl_2. This subsequent proline-catalyzed aldol reaction proceeded smoothly in affording fluorohydrin **10** as a single diastereomer. A full discussion of the scope and utility of this potentially powerful transformation will be presented elsewhere.



Scheme 4. Synthesis of the 2-deoxy-2-fluorocarbagalactose analogue 4.

Surprisingly, while the chlorohydrin **9** could be isolated and stored indefinitely, the analogous fluorohydrin **10** decomposed readily during purification by flash column chromatography and upon storage in a freezer. Analysis of the degradation products from **10** suggested that decomposition occurred primarily via a retro aldol reaction that was followed by subsequent reactions of the α-fluoroaldehyde. To mitigate degradation, the crude fluorohydrin was subjected directly to the optimized Julia-Kocienski olefination conditions, which afforded the diene **8** as a stable and isolable intermediate. The yield for the fluorohydrin **8** (40%) over these two steps was comparable to that of the chlorohydrin **7** (44%) following the sequence described above. From here, acetylation of the free alcohol and subsequent removal of the TIPS protecting group provided the allylic alcohol **19**, which smoothly underwent RCM using Grubb's second-generation catalyst to afford protected 2-fluorocarbasugar **20** in excellent overall yield. Finally, arylation of the free

alcohol function and global deprotection afforded the target 2-fluorocarbasugar **4** in good yield. Notably, this robust 7-step sequence proceeds in good overall yield (~20%).

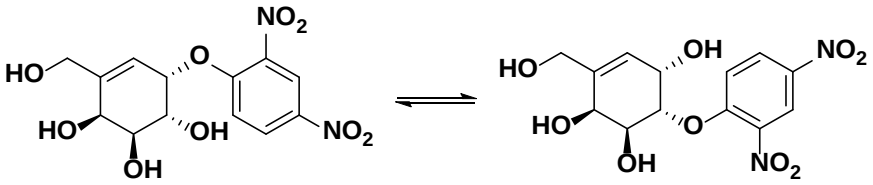
4.3.2. Evaluation of 3 and 4 as Covalent Inhibitors of a GH36 α-Galactosidase.

With the target molecules in hand, we assayed the GH36 family α-galactosidase from *T. maritima* (*Tm*GalA) with activated allylic carbasugar **3**. In contrast to the 4-nitrophenyl allylic glucoside analogue **1**, *galacto* analogue **3** is turned over within 10 minutes and therefore maybe better considered as a poor substrate for *Tm*GalA (Table 1). Unfortunately, we also observed, by ^{1}H NMR spectroscopy in D_2O, that compound **3** undergoes an isomerization reaction to give the isomeric allylic alcohol **21** (Scheme 5). This reaction reached an equilibrium (in D_2O) of ~2:1 in favor of **3** (Supporting Information Figures S23 and S24). We also noted that the rate of equilibration increased in buffered solutions (HEPES-d_{18}, data not shown), thus suggesting that this isomerization reaction is general-base catalyzed and that the rate constant for equilibration will be larger in H_2O than it is in D_2O. As a result, our measured kinetic parameters for the hydrolysis of **3** by *Tm*GalA (k_{cat} and k_{cat}/K_m Table 1) are presumed to be lower limits if, as expected, the enzyme does not bind isomer **21**.

Table 1. Kinetic Parameters for the *Tm*GalA-Catalyzed Reactions with 3 and 4. Conditions for all experiments are *T* = 37 °C in 50 mM HEPES buffer, pH 7.4.

Experiment	k_{cat} (s^{-1})	k_{cat}/K_m or k_{inact}/K_i (M^{-1} s^{-1})	K_m or K_i (μM)
Hydrolysis of 3	$(3.78 \pm 0.06) \times 10^{-2}$ [a]	$(2.25 \pm 0.19) \times 10^4$	1.68 ± 0.14
Hydrolysis of 4	$(2.29 \pm 0.11) \times 10^{-4}$ [b]	56 ± 13	4.1 ± 0.9
Covalent Inhibition by4	id[c]	143 ± 5	id[c]

[a] $k_{cat} = (9.7 \pm 2.3) \times 10^{-2}$ s^{-1} corrected value based on the inactivation data for **4**, which does not assume 100% activity. [b] $k_{cat} = (5.9 \pm 1.3) \times 10^{-4}$ s^{-1} calculated from the inactivation data for **4**. [c] id = indeterminable.



Scheme 5. Equilibration observed by NMR spectroscopy for the intramolecular migration of the 2,4-dinitrophenyl group in 3.

We subsequently measured the kinetic parameters for turnover of **4** by *Tm*GalA (Table 1) and Figure 2a shows a standard Michaelis–Menten plot for this hydrolysis reaction. The simplest kinetic scheme for covalent labeling of *Tm*GalA by **4** involves rapid formation of the covalent intermediate (k_3), from the Michaelis complex (E:I), and that this is slowly hydrolyzed (k_5) to regenerate active enzyme (Scheme 1). The appropriate equations for the enzyme-catalyzed turnover of **4** are:

$$k_{cat} = \frac{k_3 k_5}{k_3 + k_5}, \quad K_m = \frac{k_2 k_5}{k_1 k_3 + k_1 k_5} \quad and \quad \frac{k_{cat}}{K_m} = \frac{k_1 k_3}{k_2}$$

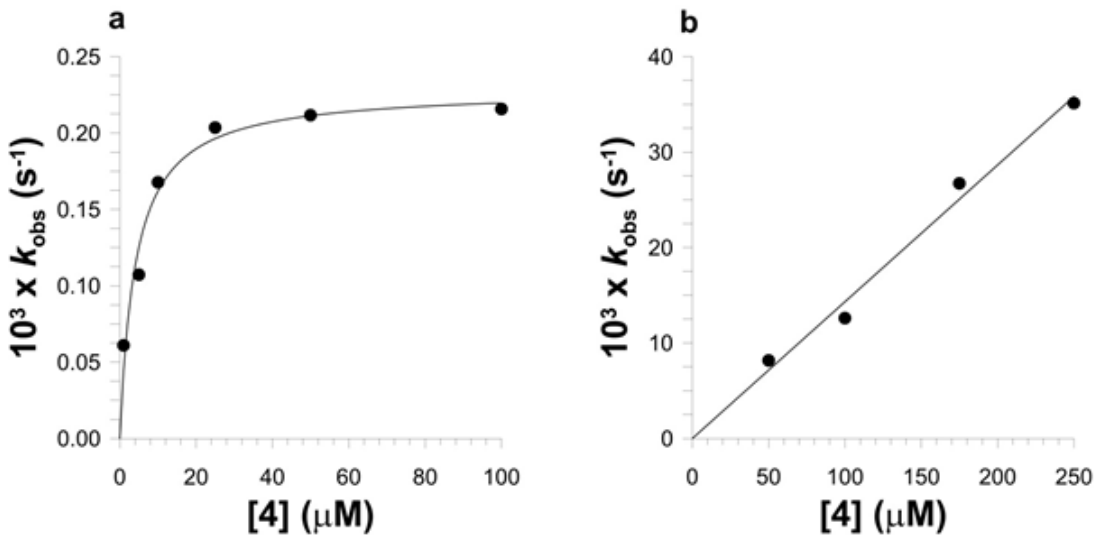


Figure 2. Kinetics for the inactivation of TmGalA by compound 4. a) Michaelis-Menten plot for the TmGalA-catalyzed hydrolysis of 4. b) A plot of the first-order rate constants for loss of TmGalA activity as a function of concentration for the covalent inhibitor 4. Conditions for all experiments are T = 37 °C in 50 mm HEPES buffer, pH 7.4.

We noted that the apparent binding constant (K_m) is in the low μM range, suggesting that pseudo-deglycosylation of **4** is rate-limiting.[8] Consequently, we undertook a sequential mixing stopped-flow experiment to measure the rate constants for the pseudo-glycosylation reaction.[41] The appropriate equations for the covalent-labeling of *Tm*GalA by **4** are:

$$k_{inact} = k_3, \quad K_i = \frac{k_2}{k_1} \quad and \quad \frac{k_{inact}}{K_i} = \frac{k_1 k_3}{k_2}$$

Figure S25 (Supporting Information) shows plots of α-galactosidase activity against incubation time for four concentrations of **4.** Unfortunately, we were unable to measure rate constants for inactivation at higher concentrations of **4** than 0.25 mM due to the large quantities of inhibitor needed for these sequential mixing stopped-flow experiments. Nevertheless, it is clear from the data shown in Figure 2b that we can only calculate the second-order rate constant k_{inact}/K_i from these covalent labeling experiments. Gratifyingly, our two calculated second-order rate constants k_{cat}/K_m and k_{inact}/K_i, which in

theory should be identical, are similar (Table 1). It must be remembered that the value of k_{cat}/K_m is calculated based on an assumption that the enzyme is 100% active. Given that no curvature is discernible in Figure 2b we are forced to conclude that $k_{cat} \approx k_5$ (Scheme 1) for the turnover of **4**, and likely **3**. Thus, once we correct k_{cat} values using the second-order stopped-flow rate constant for covalent labeling and the active *Tm*GalA concentration (based on the stopped-flow and hydrolysis kinetic data), the approximate half-lives of the two covalent intermediates formed during the reactions of **3** and **4** with *Tm*GalA are approximately 7 seconds and 20 minutes, respectively.

Introduction of the 2-fluoro group retards hydrolysis of the covalent intermediate by a factor of about 165, whereas the second-order rate constant for covalent labelling by **3** is 400-fold larger than that for **4**. That is, the fluorine substitution shows a 2-fold larger rate reducing effect on pseudo-glycosylation than on dealkylation.

4.3.3. Structural insights into the conformational itinerary adopted by *Tm*GalA during hydrolysis of 4.

We undertook structural studies of *Tm*GalA in complex with **4** to trap different points in the conformational itinerary. Given the rapid formation of the covalent intermediate of **4** with *Tm*GalA observed in the kinetic experiments, we chose to use a mutated form of *Tm*GalA, where Asp387, the acid/base residue, is mutated to alanine, in an effort to obtain a complex with intact **4**. Co-crystallisation of mutant *Tm*GalA D387A with **4** produced crystals that diffracted to 1.72 Å resolution and the structure was solved using molecular replacement with apo *Tm*GalA (PDB 5M0X) as the search model. The active site contained electron density consistent with an intact molecule of **4** bound (Figure 3a), which provides insights into interactions made at the substrate Michaelis complex. The carbasugar moiety displays a $^2H_3$ half chair conformation, as observed in the Michaelis complex with a bicyclo[4.1.0]heptyl carbasugar analogue[5]. The active site interactions are also largely unchanged; the C6-OH interacts with Asp221, the C4-OH with Trp257, Asp220 and Lys325, the C3-OH with Lys325, Arg383 and Tyr191. Due to the absence of Asp387, the hydrogen bonds formed with the C2-OH (which is a fluoro group in **4**) and with the glycosidic oxygen, which were reported previously, are not possible. The dinitrophenyl leaving group of **4** interacts only with water molecules. Superimposition of the carbasugar moiety from **4** with the bicyclo[4.1.0]heptyl compound from shows a small

downward displacement of **4** at C1 and C2; the pseudo-anomeric carbon atoms are 0.37 Å apart.

To obtain structural information on the interactions made between **4** and wild type *Tm*GalA, crystals were grown and then soaked with the compound for different lengths of time. Following a 1 h soak with **4**, the structure, solved using data to 1.42 Å resolution, revealed electron density in the active site of *Tm*GalA consistent with only the carbasugar **6** (Figure 3b). This demonstrated that within this timeframe **4** had been turned over, with retention of stereochemistry, and thus was essentially mimicking a product complex. Crystals that were soaked for a short amount of time with **4** (around 2 min) led to a structure, solved using data to 1.77 Å resolution, where the electron density showed the covalent adduct of the carbasugar with the nucleophile of *Tm*GalA (Figure 3c). The hydrolyzed carbasugar **6** binds in a similar position and same conformation ($^2H_3$) as that observed for the pseudo-glycone of **4**. There is, however, an upward displacement of **6** compared to the carbasugar moiety of **4** at C1, C2, and the carbon in place of the endocyclic oxygen, which at its greatest is 0.39 Å apart. Identical interactions are made with active site residues as described for the complex with **4**, except for the addition of interactions between the C2 fluoro group and the acid/base residue Asp387 and C1-OH with Asp387 and a water molecule. Other water molecules fill the void left by the dinitrophenyl leaving group. Comparison of the structure of *Tm*GalA in complex with **6** compared with the cyclopropyl inhibitor[5] shows there is a significant deviation at C5 (0.40 Å) and the carbon atom in place of the endocyclic oxygen (0.60 Å), while C1, C2, C3, and C4 atoms overlap. A difference in this region of the compounds is perhaps not surprising given the restraints imposed in this part of the molecule by either the double bond or cyclopropyl moiety, but such a difference was not evident in the superposition between the equivalent intact molecules.

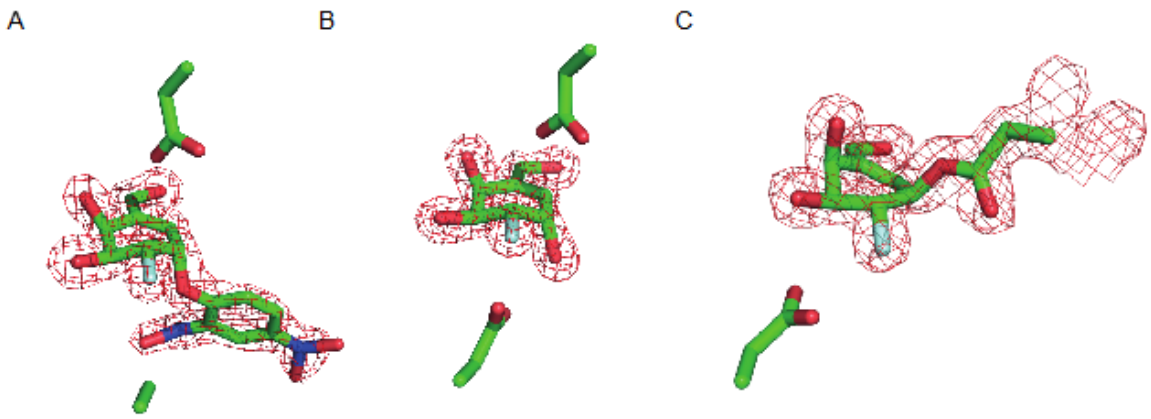


Figure 3. Structure of TmGalA in complex with 4. (A) Structure of TmGalA mutant D387A in complex with intact 4. (B) Structure of TmGalA in complex with hydrolyzed inhibitor 6. (C) Structure of TmGalA in complex with 2-deoxy-2-fluorocarbagalactose fragment of 4 covalently bound to the nucleophile Asp327. The electron density 2Fobs-Fcalc map is shown at 1.5 sigma in all cases. The catalytic residues Asp327 and Asp387 (or D387A) residues are shown.

There was unambiguous electron density demonstrating the covalent linkage of the pseudo-glycone of **4** with Asp327, the nucleophile, of *Tm*GalA with inversion of stereochemistry with respect to the substrate. The carbasugar binds in a $^{2}H_{3}$ conformation. Although there is significant movement of C1, C2 and the carbon atom in the position of the endocyclic oxygen, and to a lesser extent C3, C4 and C5, to accommodate the formation of the covalent bond when compared to intact **4**, all interactions with active site residues for the carbasugar are the same. The pseudo-anomeric carbon moves by a significant 2.04 Å between the Michaelis complex and the covalent intermediate. Unlike the superposition of the intact and hydrolyzed compounds, the covalent adduct of **4** and of the bicyclo[4.1.0]heptyl carbasugar[5] are largely identical.

4.4. Summary

In summary, we have reported the first insights into the conformational itinerary for an allylic mechanism-based inhibitor with an alpha-galactosidase from GH family 36. We synthesized allylic carbasugars that mimic galactose, and have determined their kinetic parameters for turnover and inhibition with *Tm*GalA. In addition, we have provided

structural insights into the conformations of the inhibitors at each step through the catalytic cycle. Together these findings provide glimpses into the subtleties of GH catalysis, which will aid exploration of GH mechanism and direct development of inhibitors.

4.5. References

1. Lombard, V.; Ramulu, H. G.; Drula, E.; Coutinho, P. M.; Henrissat, B. *Nucleic Acids Res.* **2014**, *42*, D490-D495.

2. Schramm, V. L. *J. Biol. Chem.* **2007**, *282*, 28297-28300.

3. Schramm, V. L. *Annu. Rev. Biochem.* **2011**, *80*, 703-732.

4. Schramm, V. L. *ACS Chem. Biol.* **2013**, *8*, 71-81.

5. Adamson, C.; Pengelly, R.; Shamsi Kazem Abadi, S.; Chakladar, S.; Draper, J.; Britton, R.; Gloster, T.; Bennet, A. J. *Angew. Chem., Int. Ed. Engl.* **2016**, *55*, 14978–14982.

6. Chakladar, S.; Wang, Y.; Clark, T.; Cheng, L.; Ko, S.; Vocadlo, D. J.; Bennet, A. J. *Nat. Commun.* **2014**, *5*, 5590.

7. Shamsi Kazem Abadi, S.; Tran, M.; Yadav, A. K.; Adabala, P. J. P.; Chakladar, S.; Bennet, A. J. *J. Am. Chem. Soc.* **2017**, *139*, 10625–10628.

8. Danby, P. M.; Withers, S. G. *J. Am. Chem. Soc.* **2017**, *139*, 10629-10632.

9. Wolf, S.; Zismann, T.; Lunau, N.; Meier, C. *Chem. Eur. J.* **2009**, *15*, 7656-7664.

10. Wolfenden, R. *Chem. Rev.* **2006**, *106*, 3379-3396.

11. Wolfenden, R.; Lu, X.; Young, G. *J. Am. Chem. Soc.* **1998**, *120*, 6814-6815.

12. Schwarz, J. C. P. *J. Chem. Soc., Chem. Commun.* **1973**, 505-508.

13. Vocadlo, D. J.; Davies, G. J. *Curr. Opin. Chem. Biol.* **2008**, *12*, 539-555.

14. Davies, G. J.; Planas, A.; Rovira, C. *Acc. Chem. Res.* **2012**, *45*, 308-316.

15. Mhlongo, N. N.; Skelton, A. A.; Kruger, G.; Soliman, M. E. S.; Williams, I. H. *Proteins* **2014**, *82*, 1747-1755.

16. Sinnott, M. *Comprehensive biological catalysis: A mechanistic reference*; Academic Press: San Diego, 1998.

17. Sinnott, M. *Carbohydrate Chemistry and Biochemistry: Structure and Mechanism*; 2nd ed.; RSC Publishing: Cambridge, 2013.

18. Sinnott, M. L. *Chem. Rev.* **1990**, *90*, 1171-1202.

19. Comfort, D. A.; Bobrov, K. S.; Ivanen, D. R.; Shabalin, K. A.; Harris, J. M.; Kulminskaya, A. A.; Brumer, H.; Kelly, R. M. *Biochemistry* **2007**, *46*, 3319-3330.

20. Hosie, L.; Sinnott, M. L. *Biochem. J.* **1985**, *226*, 437-446.

21. Withers, S. G.; Street, I. P.; Bird, P.; Dolphin, D. H. *J. Am. Chem. Soc.* **1987**, *109*, 7530-7531.

22. Withers, S. G.; Warren, R. A. J.; Street, I. P.; Rupitz, K.; Kempton, J. B.; Aebersold, R. *J. Am. Chem. Soc.* **1990**, *112*, 5887-5889.

23. Vocadlo, D. J.; Davies, G. J.; Laine, R.; Withers, S. G. *Nature* **2001**, *412*, 835-838.

24. Skaanderup, P. R.; Poulsen, C. S.; Hyldtoft, L.; Jorgensen, M. R.; Madsen, R. *Synthesis* **2002**, 1721-1727.

25. Winter, G. *J Appl Crystallogr* **2010**, *43*, 186-190.

26. Kabsch, W. *Acta Crystallogr D* **2010**, *66*, 125-132.

27. Evans, P. R.; Murshudov, G. N. *Acta Crystallogr D* **2013**, *69*, 1204-1214.

28. Collaborative Computational Project Number 4 *Acta Cryst. D* **1994**, *50*, 760-763.

29. Vagin, A.; Teplyakov, A. *J Appl Crystallogr* **1997**, *30*, 1022-1025.

30. Mccoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J Appl Crystallogr* **2007**, *40*, 658-674.

31. Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. *Acta Crystallogr D* **2011**, *67*, 355-367.

32. Emsley, P.; Cowtan, K. *Acta Crystallogr D* **2004**, *60*, 2126-2132.

33. Joosten, R. P.; Long, F.; Murshudov, G. N.; Perrakis, A. *Iucrj* **2014**, *1*, 213-220.

34. Bienfait, B.; Ertl, P. *J Cheminformatics* **2013**, *5*.

35. Schuttelkopf, A. W.; van Aalten, D. M. F. *Acta Crystallogr D* **2004**, *60*, 1355-1363.

36. Speciale, G.; Thompson, A. J.; Davies, G. J.; Williams, S. J. *Curr. Opin. Struct. Biol.* **2014**, *28*, 1-13.

37. Tanaka, K. S. E.; Winters, G. C.; Batchelor, R. J.; Einstein, F. W. B.; Bennet, A. J. *J. Am. Chem. Soc.* **2001**, *123*, 998-999.

38. Wang, Y.; Bennet, A. J. *Org. Biomol. Chem.* **2007**, *5*, 1731-1738.

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

40. Guce, A. I.; Clark, N. E.; Salgado, E. N.; Ivanen, D. R.; Kulminskaya, A. A.; Brumer, H.; Garman, S. C. *J. Biol. Chem.* **2010**, *285*, 3625-3632.

41. Khazaei, K.; Yeung, J. H. F.; Moore, M. M.; Bennet, A. J. *Can. J. Chem.* **2015**, *93*, 1207-1213.

4.6. Supporting Information

Figure S1: ^{1}H NMR spectrum for **11** in $CDCl_3$.

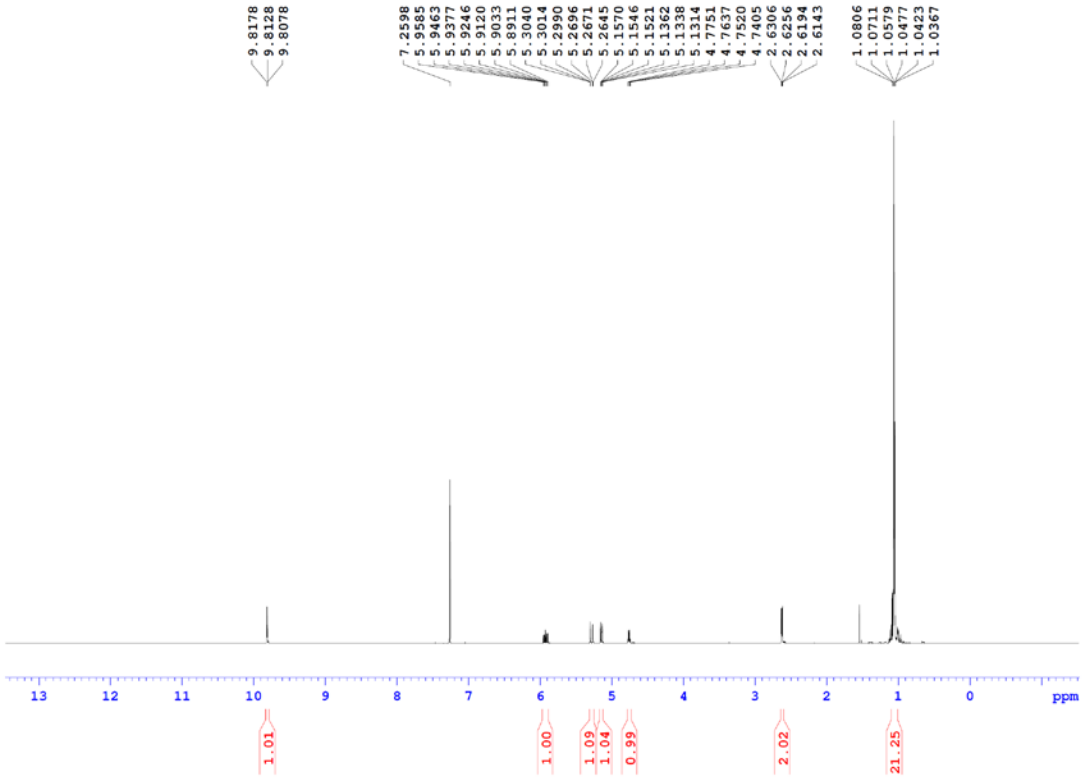


Figure S2: ^{13}C NMR spectrum for **11** in $CDCl_3$.

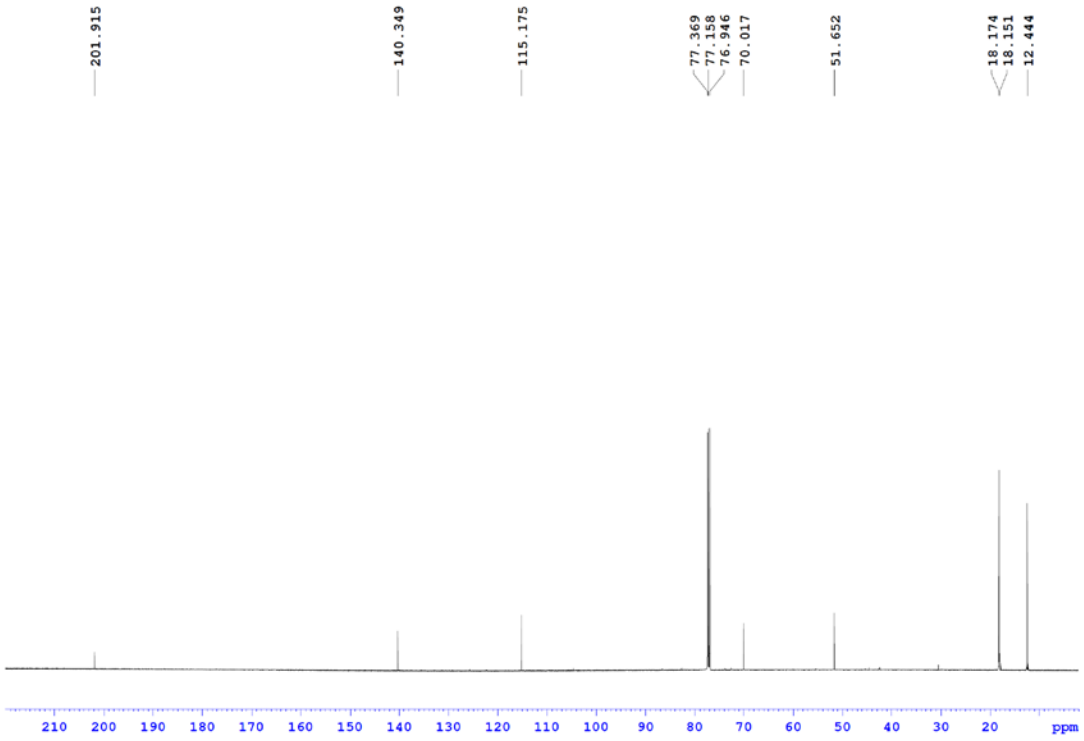


Figure S3: ^{1}H NMR spectrum for **9** in $CDCl_3$.

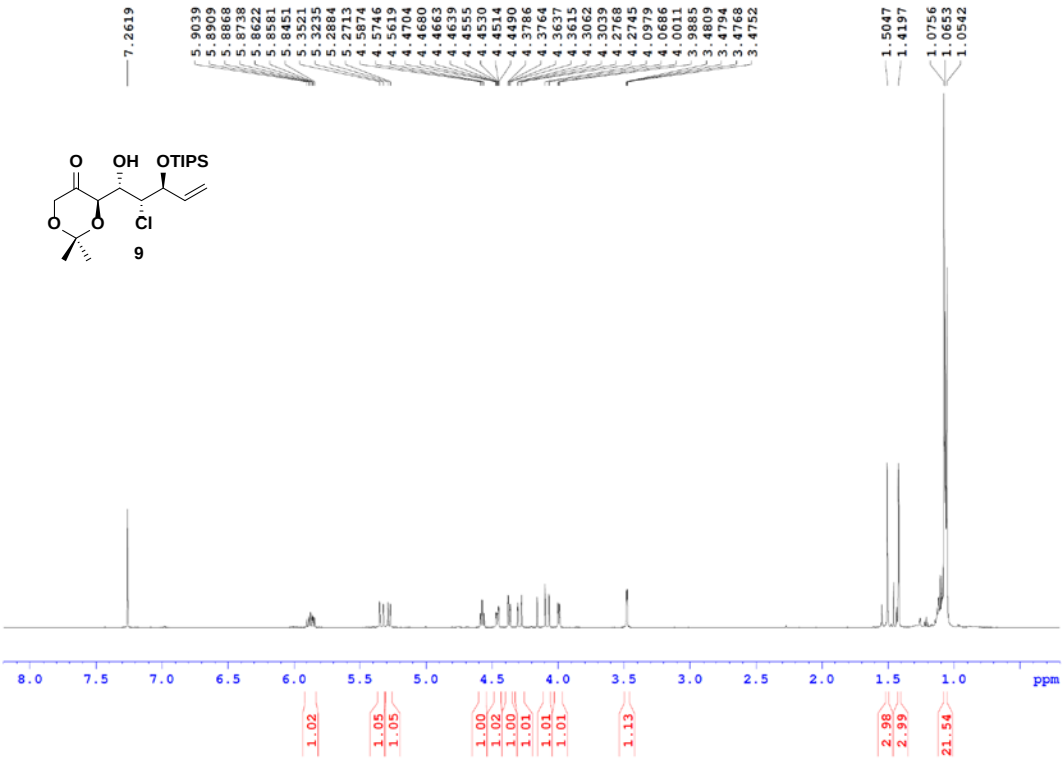


Figure S4: ^{13}C NMR spectrum for **9** in $CDCl_3$.

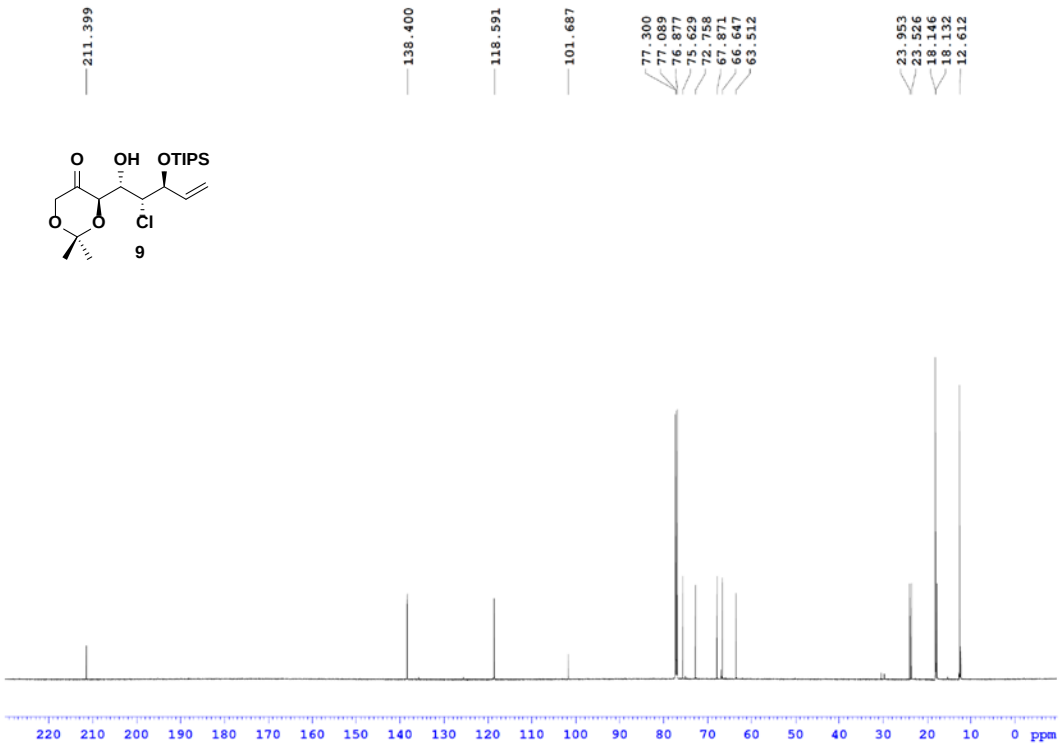


Figure S5: ^{1}H NMR spectrum for **7** in $CDCl_3$.

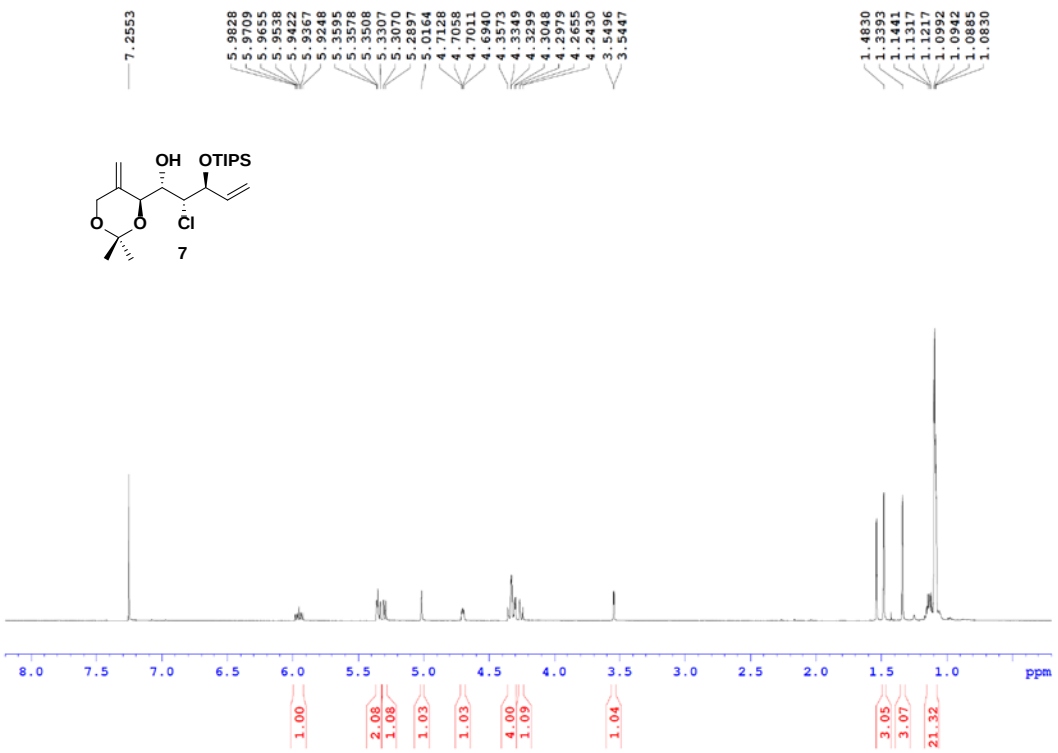


Figure S6: ^{13}C NMR spectrum for **7** in $CDCl_3$.

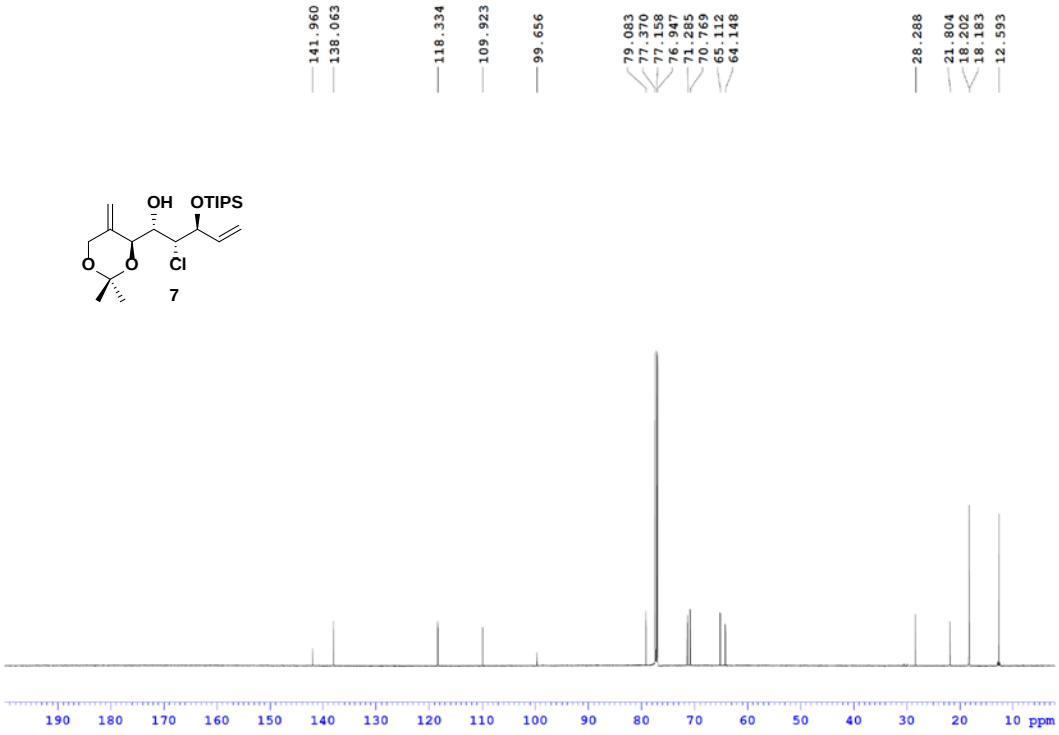


Figure S7: ^{1}H NMR spectrum for **15** in $CDCl_3$.

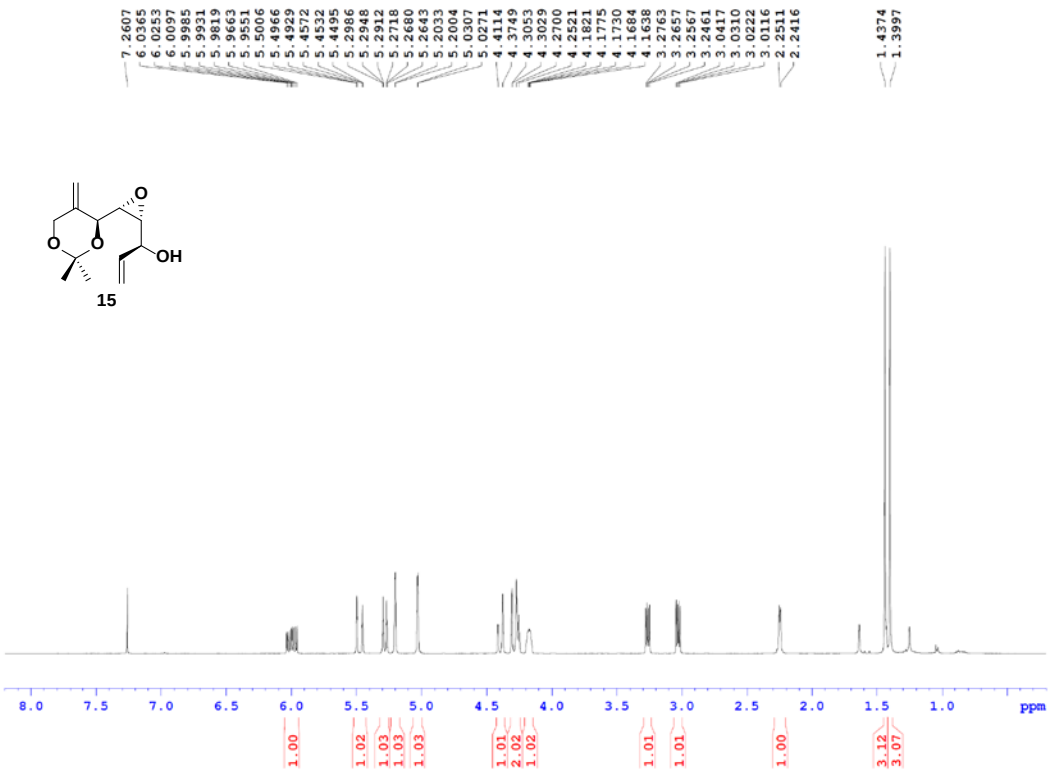


Figure S8: ^{13}C NMR spectrum for **15** in $CDCl_3$.

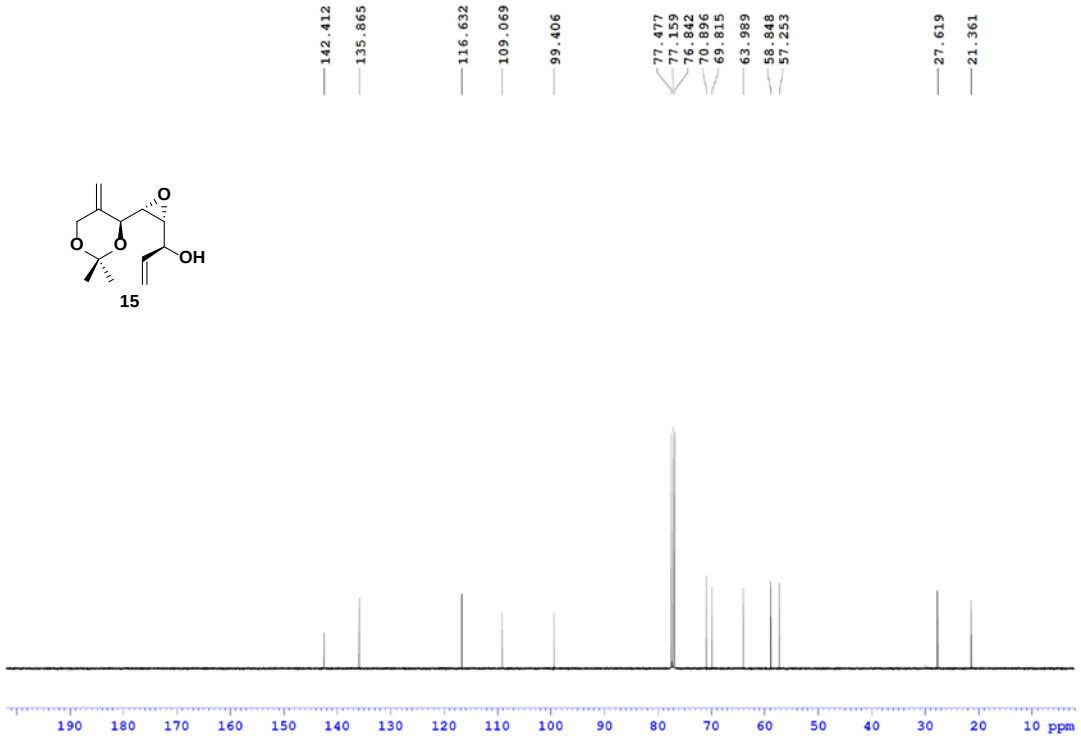


Figure S9: ^{1}H NMR spectrum for **16** in $CDCl_3$.

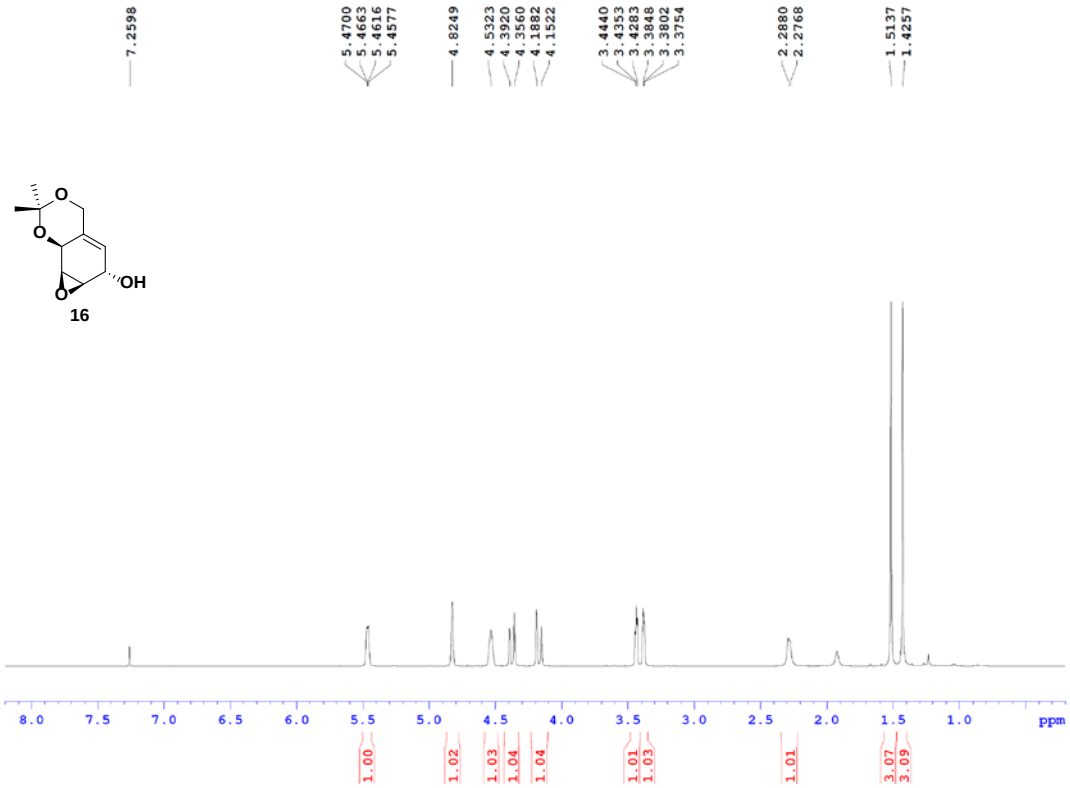


Figure S10: ^{13}C NMR spectrum for **16** in $CDCl_3$.

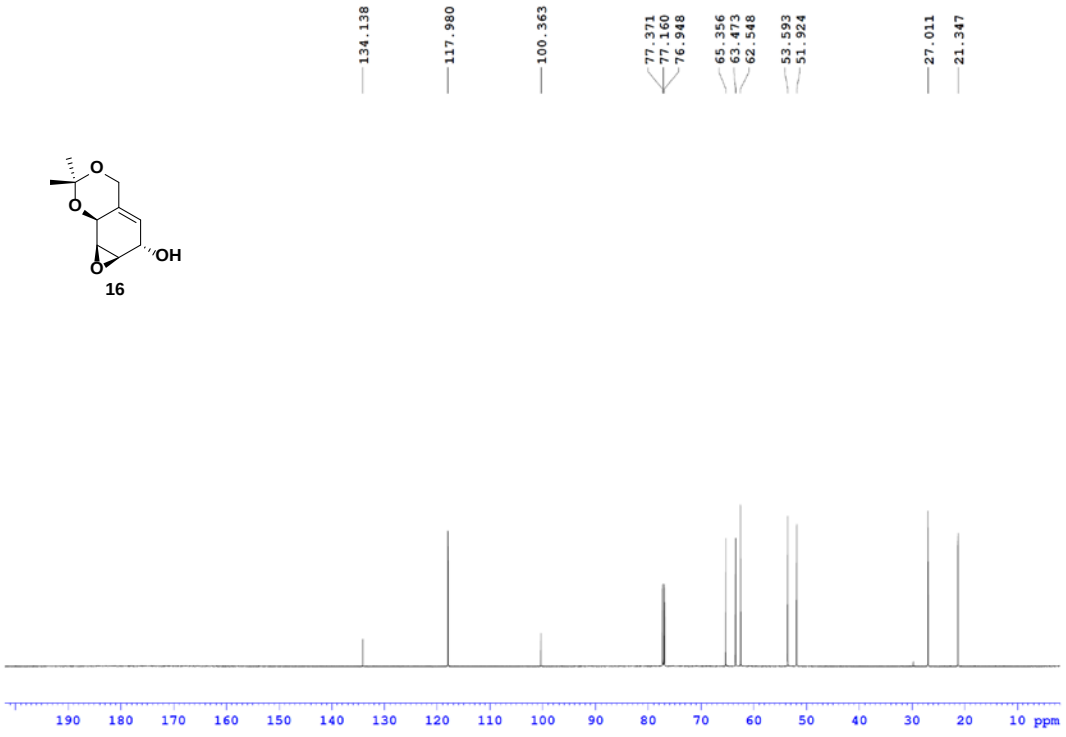


Figure S11: ^{1}H NMR spectrum for **18** in $CDCl_3$.

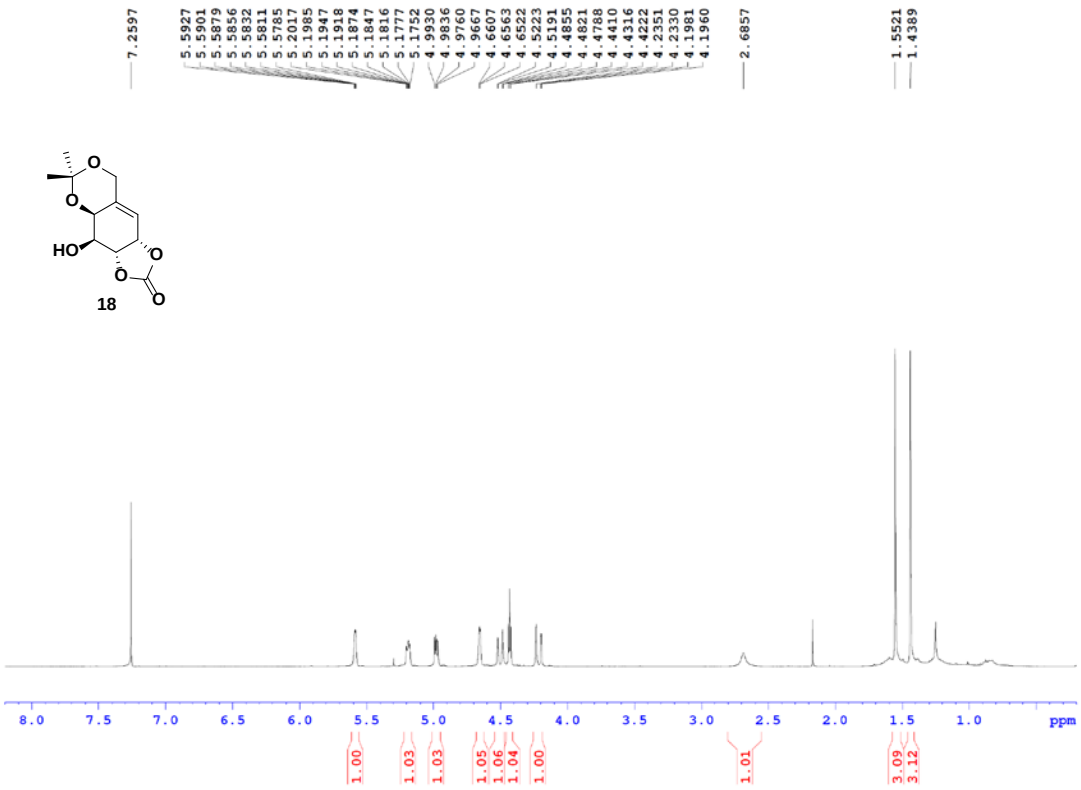


Figure S12: ^{13}C NMR spectrum for **18** in $CDCl_3$.

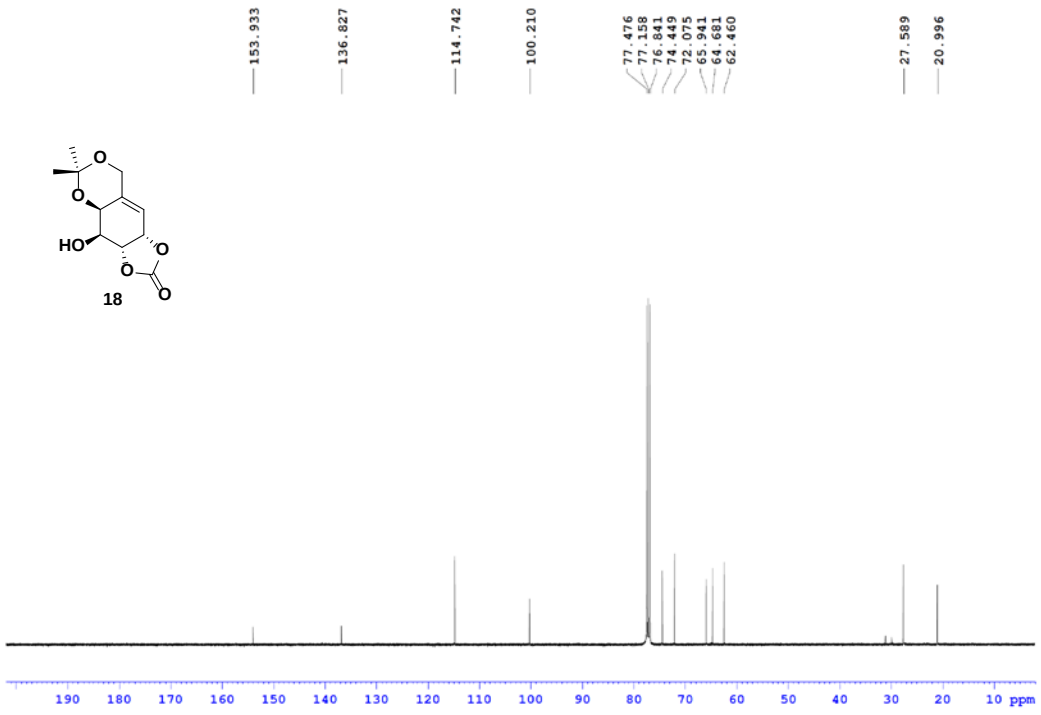


Figure S13: ^{1}H NMR spectrum for **3** in CD_3OD.

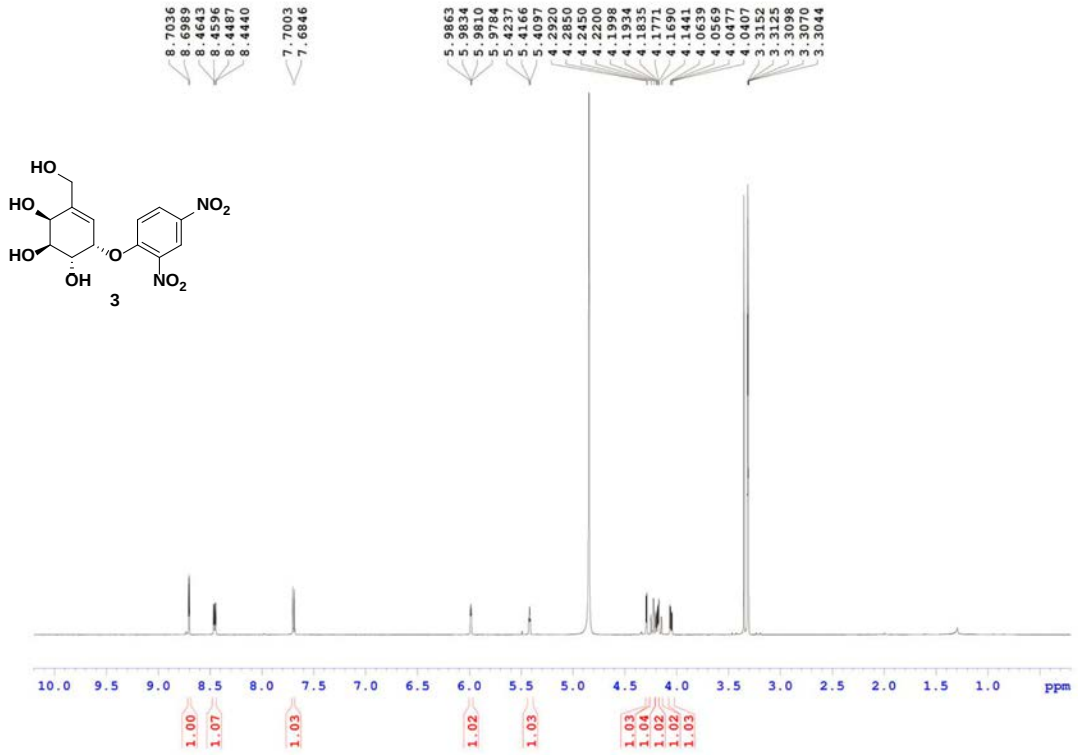


Figure S14: ^{13}C NMR spectrum for **3** in CD_3OD.

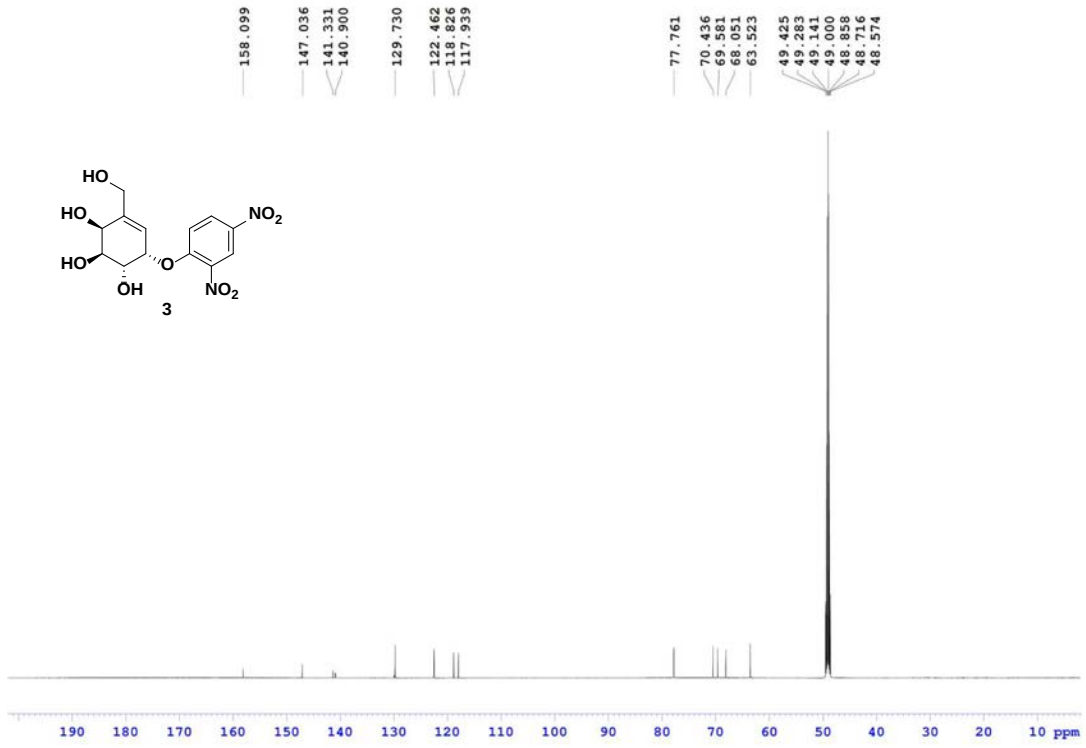


Figure S15: ^{1}H NMR spectrum for **8** in $CDCl_3$.

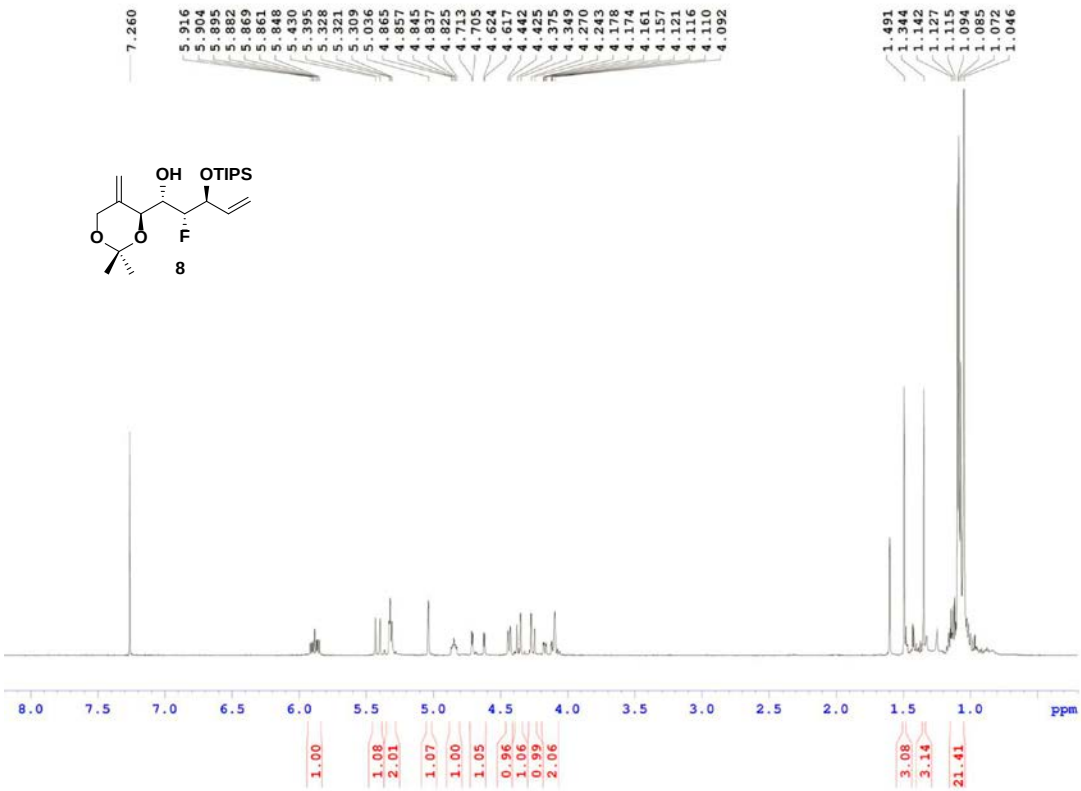


Figure S16: ^{13}C NMR spectrum for **8** in $CDCl_3$.

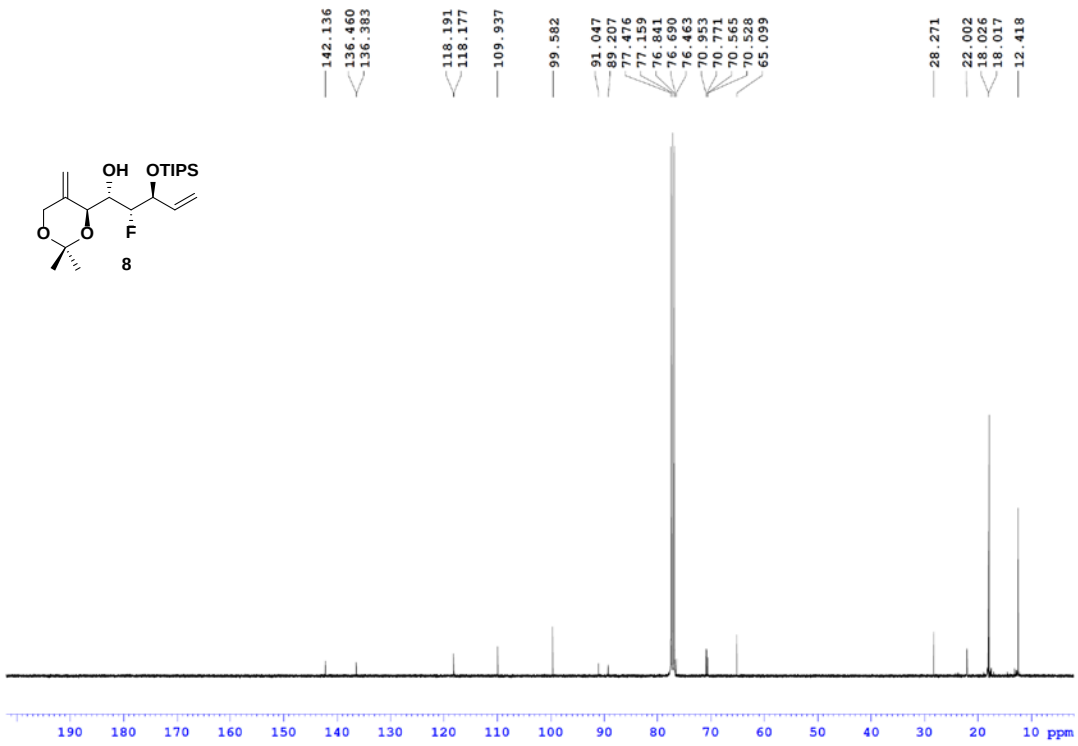


Figure S17: ^{1}H NMR spectrum for **19** in $CDCl_3$.

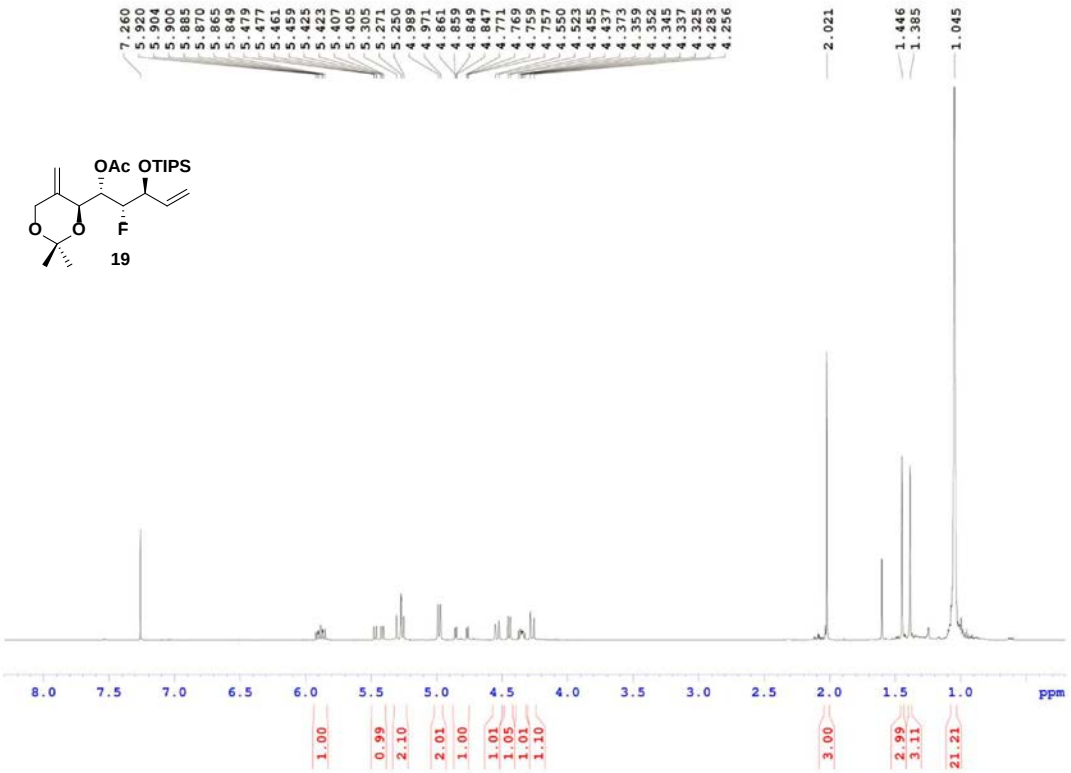


Figure S18: ^{13}C NMR spectrum for **19** in $CDCl_3$.

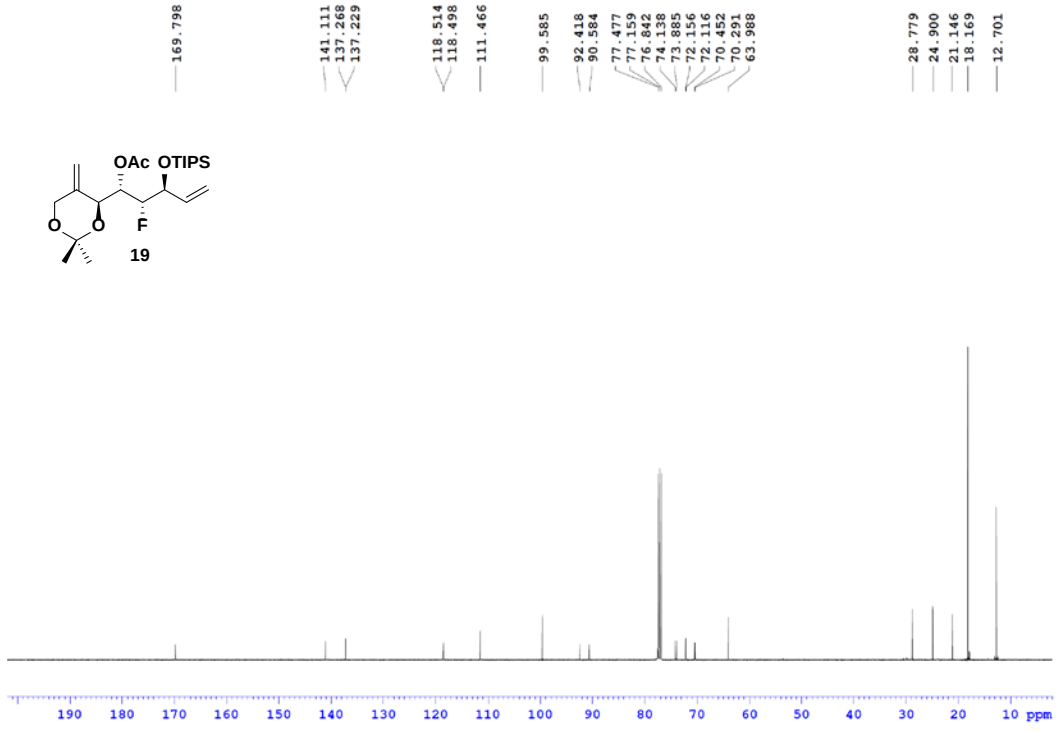


Figure S19: ^{1}H NMR spectrum for **20** in $CDCl_3$.

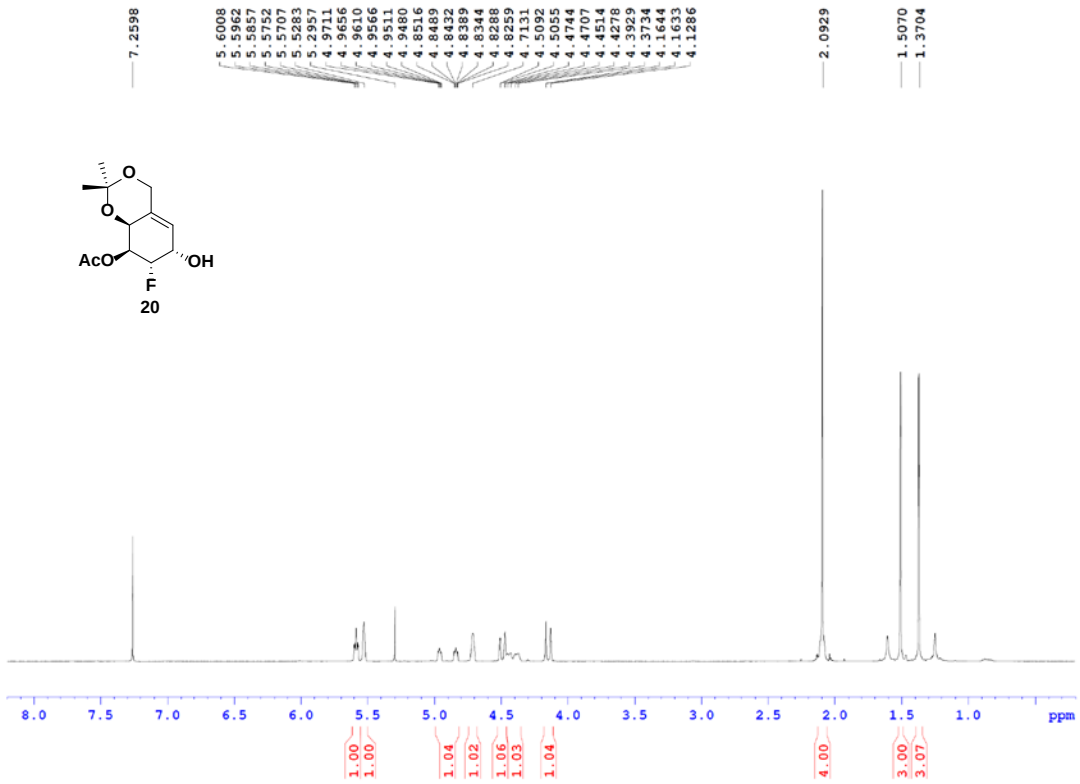


Figure S20: ^{13}C NMR spectrum for **20** in $CDCl_3$.

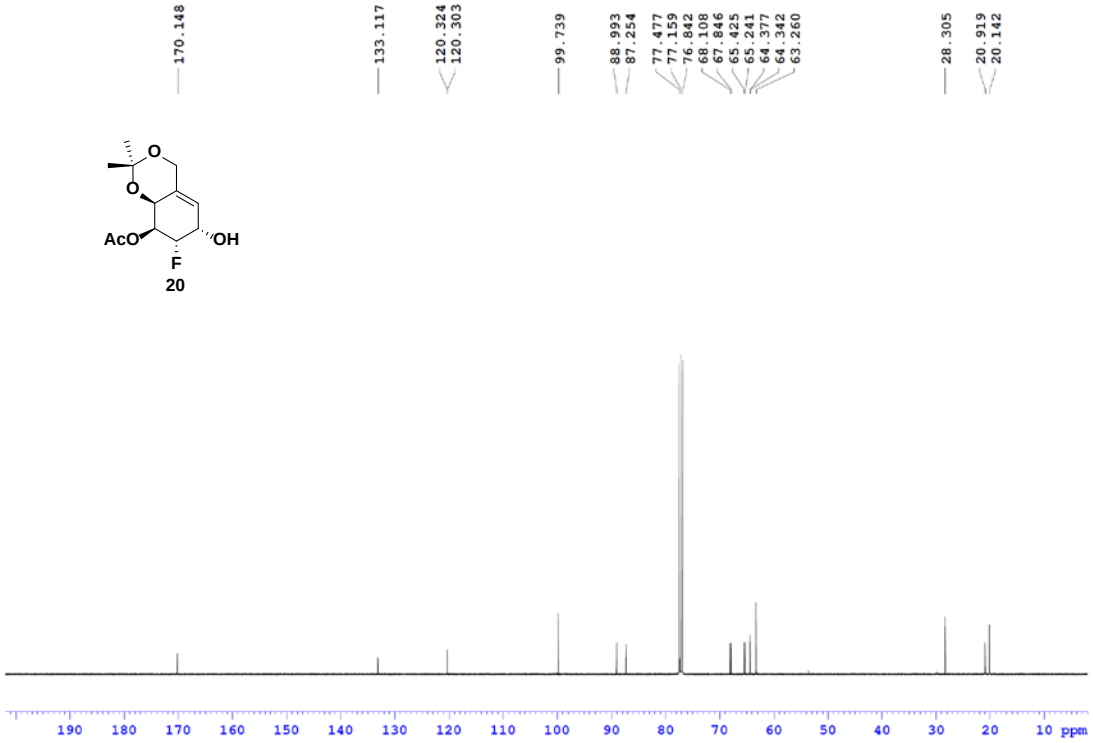


Figure S21: ^{1}H NMR spectrum for **4** in CD_3OD.

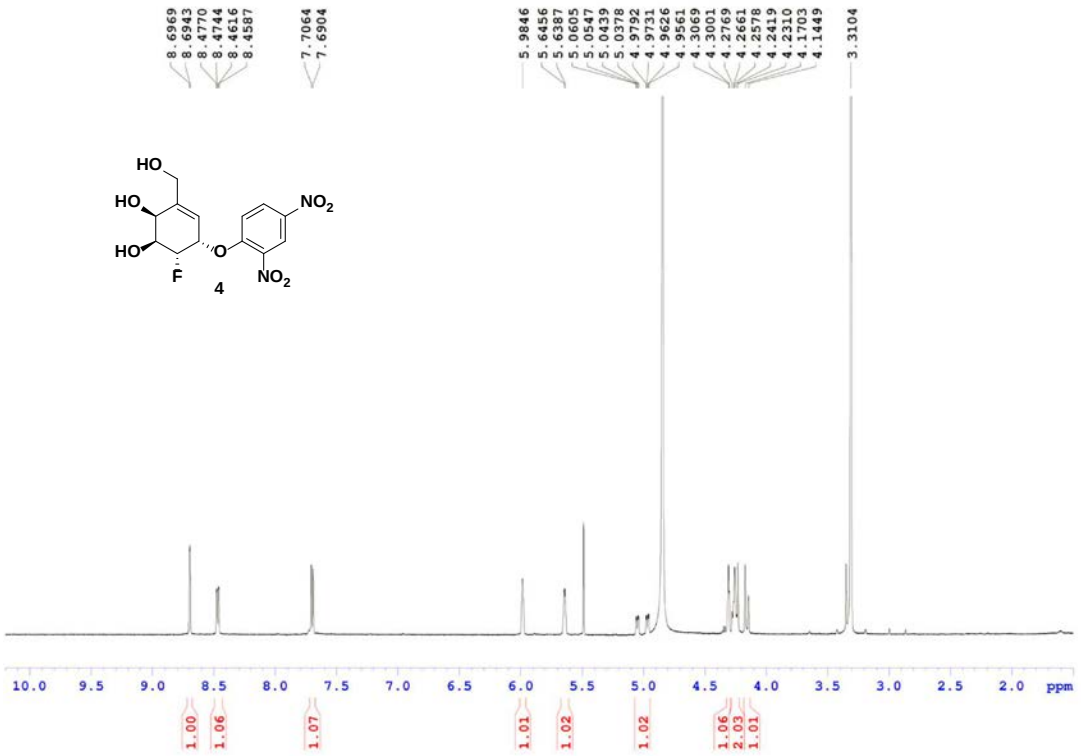


Figure S22: ^{13}C NMR spectrum for **4** in CD_3OD.

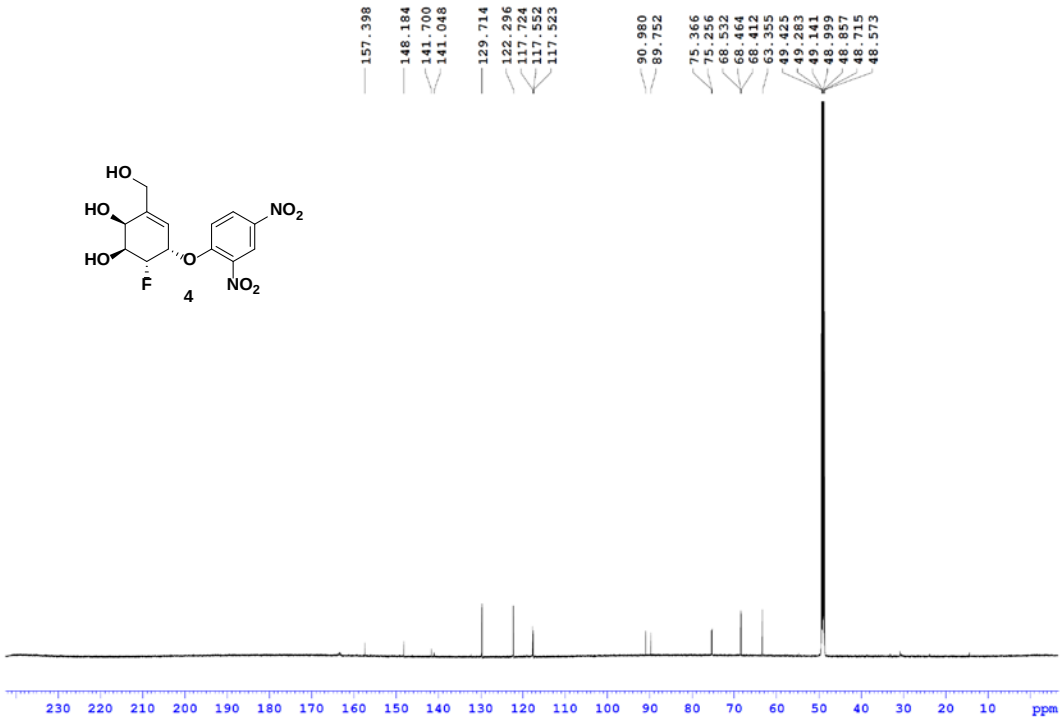


Figure S23: Initial ^{1}H NMR spectrum for equilibration of compounds **3** and **3'** (in D_2O at rt)

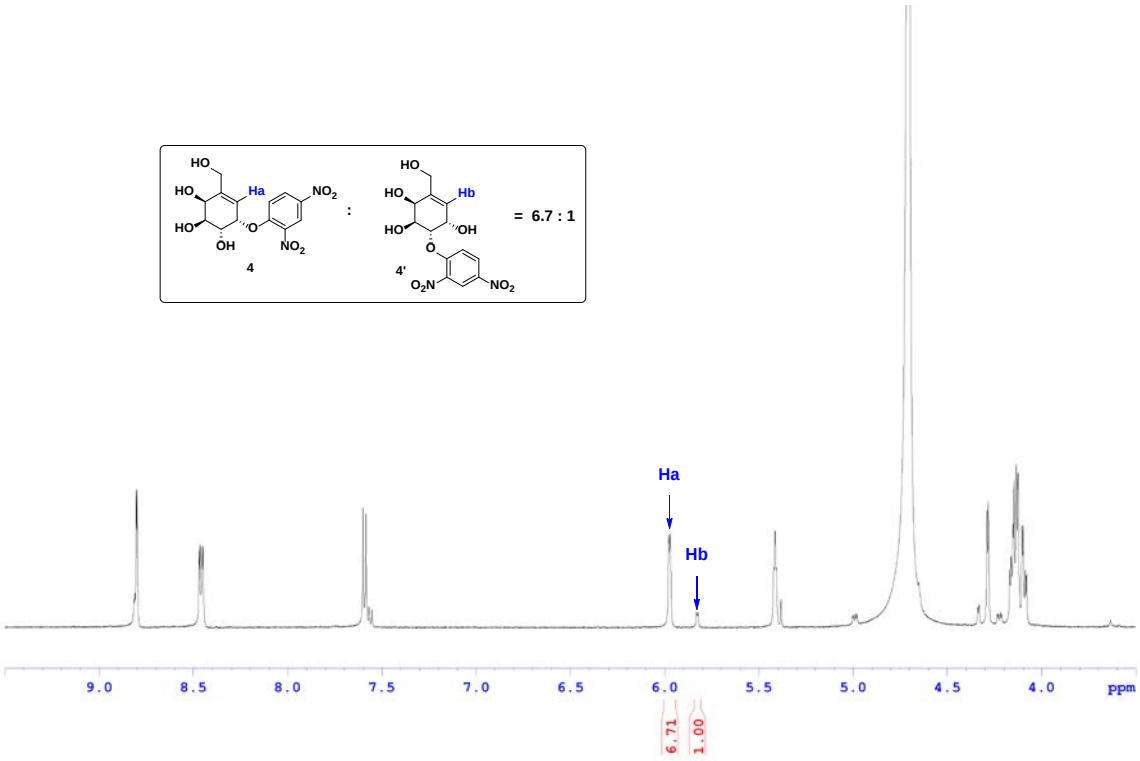
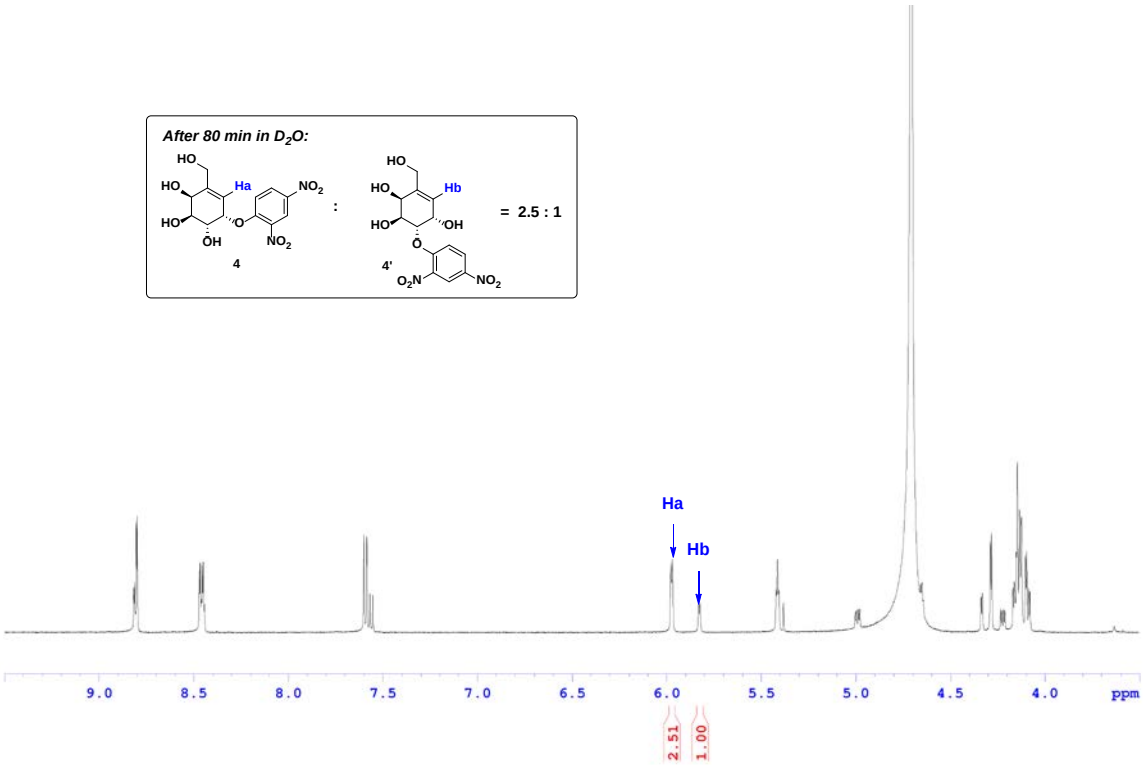


Figure S24: ^{1}H NMR spectrum for the equilibration of **3** and **3'** (in D_2O at rt) after 3 hrs.



Chapter 5. A Mechanistic Study on the α-N-acetylgalactosaminidase from *E. meningosepticum*: A Family 109 Glycoside Hydrolase

*Saswati Chakladar, Saeideh Shamsi Kazem Abadi and Andrew J. Bennet**

Departments of Chemistry and Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC V5A 1S6, Canada.

* To whom correspondence should be addressed. E-mail: *bennet@sfu.ca*

Reproduced from "A mechanistic study on the α-N-acetylgalactosaminidase from *E. meningosepticum* : a family 109 glycoside hydrolase. Chakladar, Saswati ; Shamsi Kazem Abadi, Saeideh ; Bennet, Andrew J., MedChemComm, 2014, Vol.5(8), pp.1188-1192" with permission from the Royal Society of Chemistry.

5.1. Abstract

A recombinant glycoside hydrolase family 109 α-*N*-acetylgalactosaminidase from the pathogenic bacteria *E. meningosepticum* catalyses the hydrolysis of aryl 2-acetamido-2-deoxy-α-D-galactopyranosides. The sensitivities to leaving group abilities (β_{lg} values) on *V* and *V/K* are –0.08 ± 0.06 and –0.31 ± 0.12, respectively. These results are consistent with an E2 elimination following hydride transfer from C3.

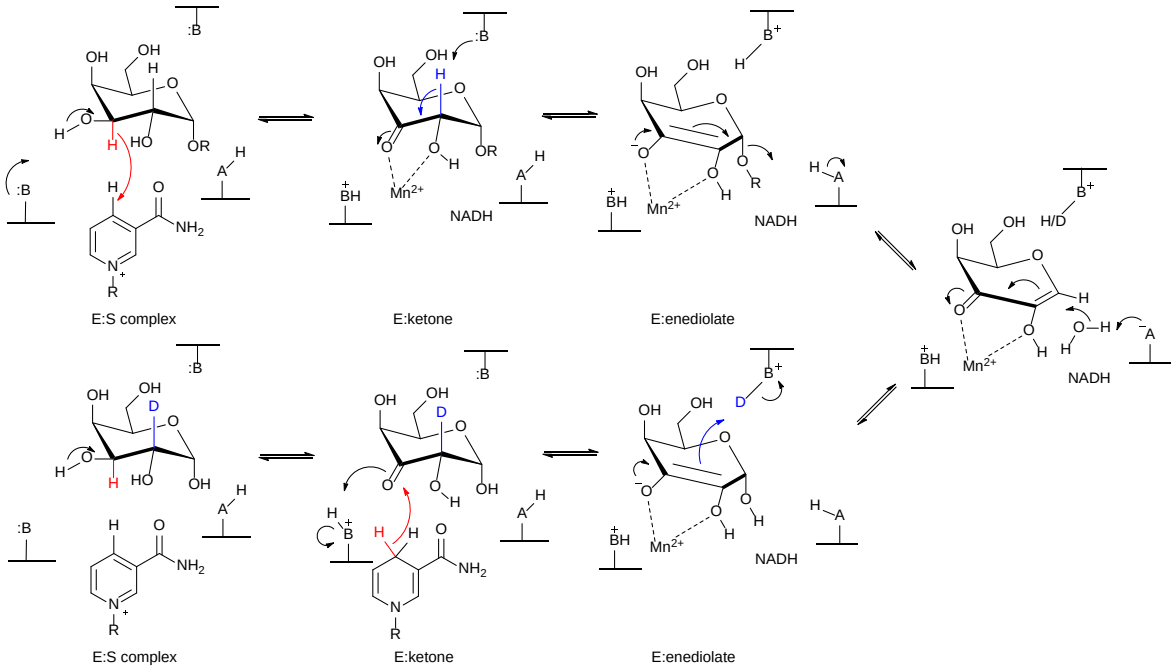
5.2. Introduction

There is a growing awareness in the scientific community that post-translational protein glycosylation is a complex modification, found throughout nature,[1] that individualizes biomolecules and imparts physical properties that define biomolecular roles in a range of physiological and biological recognition events. α-*N*-Acetylgalactosamine (αGalNAc) residues are commonly found as constituents of mucin glycoproteins,[2, 3] and they are critical antigenic components of blood group A[4] and of the sensory nerve structures of humans.[5] Enzymatic removal of the antigenic αGalNAc residue from blood group A cells has been studied as a potential route for the production of universal red blood cells (RBCs) for transfusion.[6-8] Glycosidases (glycoside hydrolases: GHs) constitute a superfamily of enzymes that hydrolyse glycosidic bonds found in a wide range of glycoconjugates. Based on sequence alignment, these enzymes are classified into more than 130 different families.[9-11] α-*N*-Acetylgalactosaminidases (α-NAGALs) are carbohydrate-processing enzymes that cleave terminal non-reducing αGalNAc residues from glycoconjugates. α-NAGALs have been classified into GH families 27, 36, 109 and 129.[8, 12-14] Of these, GH27 and GH36 share a common ancestral gene and are members of the glycoside hydrolase clan GH-D, which possess a commonly found triosephosphate isomerase (TIM) barrel structure.[9]

Most GHs[15] catalyze simple nucleophilic substitution reactions that occur on the anomeric carbon atom of a carbohydrate. The enzymatic transition state for this reaction involves the development of positive charge on the anomeric carbon, which then can delocalize onto the ring oxygen atom.[16, 17] Indeed, three out of the four glycoside hydrolase

families that have α-NAGAL enzymes as members (GH27, GH36 and GH129) are proposed to operate via such nucleophilic substitution reactions.[12, 14]

In 2007, Liu *et al.* reported the X-ray crystal structure of an α-NAGAL enzyme from *E. meningosepticum,* which is a member of the exclusively prokaryotic GH109 family.[8] The X-ray diffraction structure revealed several similarities between GH4 glycosidases and GH109 α-NAGALs, as well as a remote relationship to the sequences of oxidoreductases.[8, 18] An unusual feature of the GH4 family is that various cofactors are required for activity, in particular, NAD^+, a divalent metal ion (Mn^{2+}) and a reducing agent (such as DTT or TCEP).[19-22] Scheme 1 shows the currently accepted mechanism for GH4 NAD^+-dependent glycosidases, using α-galactosidase as an example; the NAD^+-dependent glycosidase reaction mechanism is atypical of glycosidases. In Scheme 1, the enzyme:ketone complex is shown as a bona fide intermediate,[20-22] however, based on kinetic isotope effect evidence it has been proposed that the E:S complex undergoes a concerted reaction to give the E:enediolate intermediate directly.[19] Key observations that underpin this mechanism include: (i) a requirement for NAD^+ for activity, (ii) the first formed product has the same anomeric configuration as the substrate (retaining glycoside hydrolase), and (iii) when the reaction is performed in D_2O the hydrolysis product contains a deuterium on C-2.[20-22] Given that Liu et al. reported similar observations for their *E. meningosepticum* GH109 α-NAGAL it is clear that in addition to sharing a structural similarity, GH4 and GH109 enzymes are mechanistically comparable.[8, 23] Lastly, the GH4 family contains enzymes that catalyze the hydrolysis of both α- and β-glycoside substrates,[18, 24] an unusual situation that is not seen in GH109 where all members characterized to date are α-NAGALs that show narrow substrate specificity.[8]



Scheme 1 Proposed mechanism for GH4 glycoside hydrolases showing all possible intermediates in the catalytic cycle. Hydride transfers are shown in red, while proton transfers to and from C2 are shown in blue.19-22 The required divalent cation (Mn2+) is not shown in all structures in order to avoid overlap with mechanistic arrows. Also shown is the incorporation of deuterium at C2 in the product when the reaction is performed in D2O.

In contrast to GH4 glycosidases, the GH109 family enzyme *E. meningosepticum* α-*N*-acetylgalactosaminidase requires no external cofactors for activity. That is, the required NAD^+ is embedded in the enzyme environment where it remains tightly bound to the protein.[8] Because the *E. meningosepticum* enzyme does not require a divalent cation, the pK_a of the C2 proton of the intermediate (E:ketone, Scheme 1) should be higher than that of the corresponding species bound to the GH4 enzyme. In contrast, the electrophilic stabilization provided by the bound divalent cation (Mn^{2+}) in GH4 enzymes should provide a dramatic lowering of the free energy of the E:enediolate intermediate (Scheme 1). This major difference in the GH4 and GH109 enzyme active sites piqued our interest and provided incentive for undertaking an in-depth mechanistic study on a GH109 family member. In this report, we detail the cloning of the Flv109 gene from *E. meningosepticum*

(ATCC 51720D) as well as the expression, purification, and characterization of the encoded protein (α-NAGAL).

5.3. Enzyme Production

We amplified the genomic DNA of *E. meningosepticum*—using a different strain than that reported previously in the literature[8] and used this material to generate a truncated α-*N*-acetylgalactosaminidase lacking the first 17 *N*-terminal amino acid residues. In addition, our construct differs in sequence from the enzyme used by Liu et al[8] by six amino acid residues. Full experimental details and protein and DNA sequences are given in the supporting information section. We mapped the six amino acid changes onto the published structure of the *E. meningosepticum* α-*N*-acetylgalactosaminidase (Fig. 1) and found that none are within the active site of the enzyme.

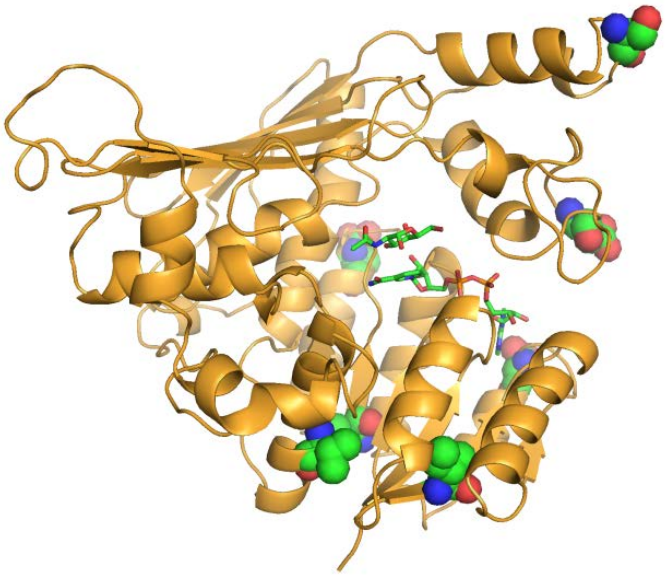
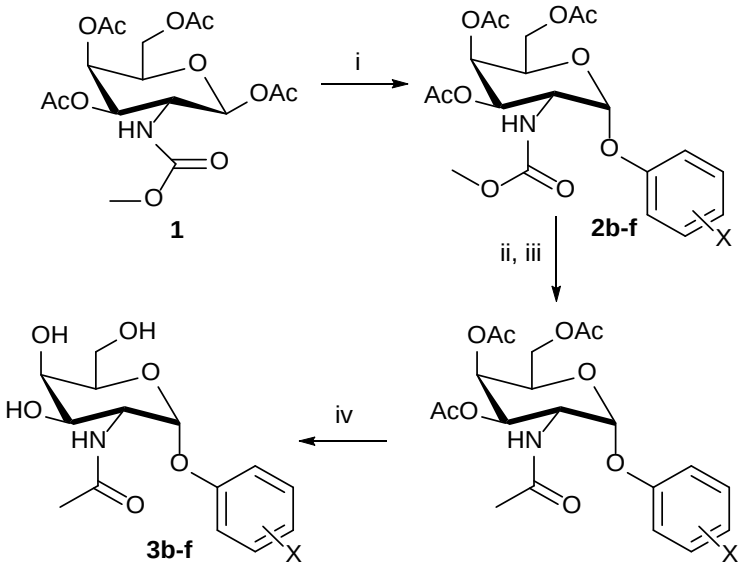


Figure 1. The structure for a single monomer of the *E. meningosepticum* α-*N*-acetylgalactosaminidase, where the NAD^+ and GalNAc bound to the active site are shown as stick structures. The positions of the six amino acid residues that differ between the α-NAGAL used in the current study and that reported by Liu et al. are shown using space-filling structures.

5.4. Substrate Synthesis

A panel of five substituted aryl 2-acetamido-2-deoxy-α-D-galactopyranosides were synthesized and used in combination with the commercially available 4-nitrophenyl substrate (3a) to assess the effect of leaving group ability on the two catalytic constants (k_{cat} and k_{cat}/K_m). We used stannic chloride promoted glycosylation of methylcarbamate 1 for the synthesis of the protected aryl α-galactosaminides 2b–f (Scheme 2). Standard reactions gave access to substrates 3b–f. Full experimental details are given in the supporting information section.



Scheme 2 Reagents and conditions for the syntheses of aryl 2-acetamido-2-deoxy-α-d-galactopyranoside substrates: (i) ArOH, SnCl4, CH2Cl2, rt; (ii) TBAF, THF, reflux; (iii) Ac2O, pyridine, rt; (vi) NaOMe, MeOH, 0 °C

5.5. Kinetic and Product Studies.

In order to verify mechanistic commonality between the enzymes of glycoside hydrolase families 4 and 109 (GH4 and GH109) and to confirm that the six amino acid changes to our enzyme are mechanistically unimportant we performed a product study for the α-NAGAL-catalyzed hydrolysis of 4-nitrophenyl 2-acetamido-2-deoxy-α-D-galactopyranoside (PNPαGalNAc) in the presence of methanol. The product of this

reaction contained an anomeric proton with a chemical shift of 5.22 ($J_{1,2}$ = 3.8 Hz), thereby confirming that the reaction proceeded with retention of anomeric configuration to give methyl 2-acetamido-2-deoxy-α-D-galactopyranoside. We also confirmed that α-NAGAL-catalyzed hydrolysis of PNPαGalNAc in D_2O gave the reaction product 2-acetamido-2-deoxy-D-galactose which was completely deuterated at C-2 as indicated by the appearance of the anomeric protons for both anomers as singlets in the ^{1}H NMR spectrum.

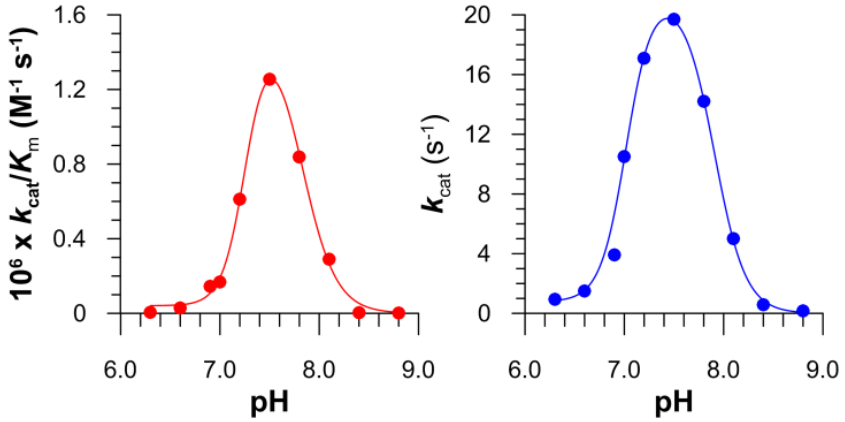


Figure 2. pH activity profiles: kcat/Km (left in red) and kcat (right in blue) for the hydrolysis of PNPαGalNAc by α-NAGAL.

The measured pH versus rate profiles for k_{cat}/K_m and k_{cat} (Fig. 2) for the *E. meningosepticum* GH109 α-NAGAL-catalyzed hydrolysis of PNPαGalNAc are bell-shaped curves. In theory, a bell-shaped curve for the pH profile of k_{cat} means that the enzyme active site has two ionisable groups (one protonated and the other deprotonated during the catalytic cycle).[25] In the case of our enzyme, which exists in solution as a functional dimer, the experimental data only fit to a pH-rate equation that incorporated Hill coefficients on both of the ionization events (Table 1). Because the derived Hill coefficients for the changes in α-NAGAL activity as a function of pH are greater than 2, the two active sites in the dimeric enzyme—one from each monomeric unit—are not functionally independent of one another.[25] Our GH109 α-NAGAL enzyme exhibits a narrow peak of activity in a pH range of 7.0–8.0 (Fig. 2), and the two apparent pK_a values that characterize its activity are around 7.1 and 7.9 (Table 1).

The maximal rate constants measured in our current study are similar to those reported by Liu et al. (Table 1).[8] Thus, we conclude that the six amino acid differences between the two enzymes do not significantly affect their catalytic activities.

Table 1. Kinetic Parameters for GH109 *E. meningosepticum* α -*N*-acetylgalactosaminidase-catalyzed hydrolysis of PNP α GalNAc.

parameter	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})
maximal rate constants	19.7 ± 1.0[a]	$(1.25 \pm 0.05) \times 10^6$[a]
pK_{a1}	7.0 ± 0.1	7.3 ± 0.1
pK_{a2}	7.9 ± 0.1	7.8 ± 0.1
h_1	3.4 ± 0.2	3.7 ± 1.4
h_2	2.8 ± 0.2	2.5 ± 0.8

[a] Rate constants reported by Liu et al. $k_{cat} = 9.84 \pm 0.16$ s^{-1} $k_{cat}/K_m = (1.28 \pm 0.10) \times 10^6$ M^{-1} s^{-1}.[8]

Next, we evaluated the sensitivity of the enzymatic rate constants k_{cat} and k_{cat}/K_m to a change in the leaving group ability. This was accomplished by measuring the change in the logarithm of the rate constant as a function of the pK_a of the leaving group aglycone's conjugate acid.[26] Fig. 3 presents Brønsted plots of the data; the derived sensitivities (β_{lg} values) on k_{cat} and k_{cat}/K_m are 0.08 ± 0.06 and –0.31 ± 0.12, respectively.

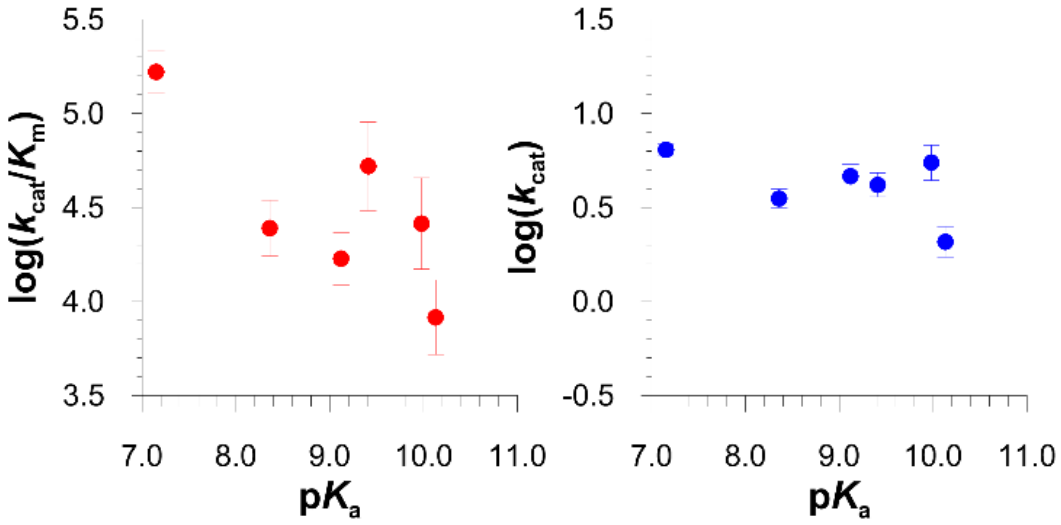


Figure 3. Brønsted plots. Effect of leaving group ability on k_{cat}/K_m (red circles) and k_{cat} (blue circles) for α-*N*-acetylgalactosaminidase-catalyzed hydrolyses. All experiments were performed at 37 □C and pH 7.5. Leaving group abilities represented as pK_a (ArOH) are as follows: 4-nitrophenol (7.15); 3-nitrophenol (8.36); 3-chlorophenol (9.12); 4-chlorophenol (9.41); phenol (9.98); and 4-methoxyphenol (10.13).

The mechanism of action of GH4 and GH109 enzymes involves a hydride transfer from C3 to a bound NAD^+ co-factor to give a non-covalently bound ketone (Scheme 1). Following the formation of this intermediate,[20, 22] or during hydride transfer,[19] a proton is transferred from C2 of the sugar to an enzymatic tyrosine residue.[8, 18, 24] This transfer occur either simultaneously with (i.e., an E2 elimination mechanism) or prior to (i.e., an elimination reaction that proceeds via a conjugate base, $E1_{CBirr}$) aglycone departure to give a bound ketone-glycal intermediate (Scheme 1, E:glycal-ketone). The measured β_{lg} value on k_{cat}/K_m (–0.31 ± 0.12), is consistent with cleavage of the anomeric C–O bond being kinetically significant, although scatter in the Brønsted plot makes it difficult to come to a more definitive conclusion. If glycosidic bond cleavage is partially rate-determining then breaking this bond is likely to be part of a concerted E2 elimination reaction which by virtue of the good leaving groups used in this study, would occur without acid-catalysis. Mechanistic studies on GH4 enzymes report smaller β_{lg} values[20, 22] than our values for the GH109 enzyme; the GH4 results are interpreted to mean that the elimination occurs via an $E1_{CBirr}$ ($D_HAxh^‡ + D_N$)[27] reaction. For the GH4 enzymes deprotonation to form the E:enediolate intermediate is partially rate limiting for k_{cat}/K_m (Scheme 1).

Keeping in mind that GH4 enzymes require a divalent cation for activity, we suggest that their active site environment is such that the C-2 proton of the substrate is made more acidic as the enediolate intermediate undergoes electrophilic stabilization by the metal cofactor. In contrast, the GH109 α-NAGAL active site promotes binding with the C2 substituent via hydrophobic and hydrogen-bonding interactions with methionine and tyrosine residues, respectively (Fig. 4). We also note that the structure of α-*N*-acetylgalactosaminidase does not have an appropriately placed acidic residue to assist in aglycone departure.[8]

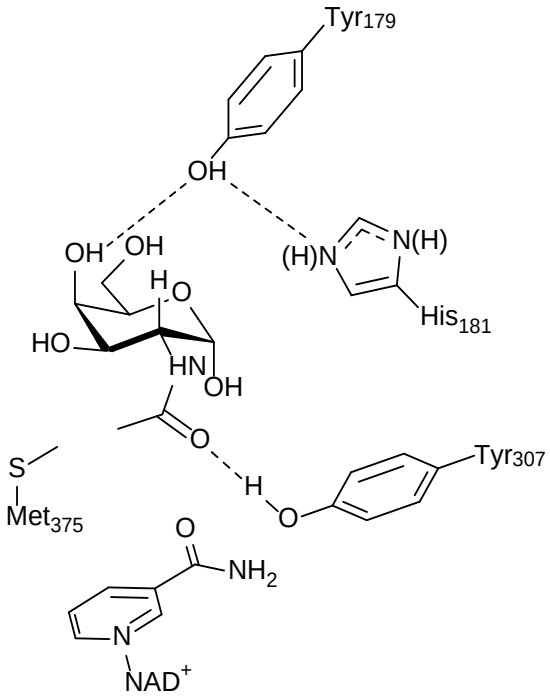


Figure 4. Cartoon depiction of the active site of GH109 E. meningosepticum α-N-acetylgalactosaminidase (modified from references 8, 23). Shown is the most likely base for the abstraction of the C2-proton (Tyr179), presumably activated by the proximal His181). Also, shown are the amino acids residues that interact with the C2-acetamido group: Tyr301 and Met375 via hydrogen-bonding and hydrophobic interactions, respectively.

Based on the β_{lg} value on k_{cat} of close to zero and that the Brønsted β_{lg} value on k_{cat}/K_m is non-zero, the kinetically significant step for k_{cat} must occur after glycosidic bond cleavage. This kinetically significant step could involve either the Michael addition of water to the glycal-ketone intermediate—which would likely occur through a network of solvent

molecules due to the lack of an acid/base catalyst proximal to the glycosidic oxygen—or the subsequent transfer of either a proton or a hydride.

Given that both GH4 and GH109 enzymes are only produced by prokaryotes, the design of selective inhibitors for these two families of glycoside hydrolases is attractive from an antimicrobial therapeutics perspective.

5.6. Conclusions

The *E*. *meningosepticum* α-*N*-acetylgalactosaminidase hydrolyses 2-acetamido-2-deoxy-α-D-galactopyranosides by a NAD^+-mediated oxidation followed by an α,β-elimination to give a Michael acceptor intermediate. This intermediate undergoes hydration along with proton and hydride transfer to generate 2-acetamido-2-deoxy-α-D-galactose as the reaction product. The hydration reaction limits k_{cat}, and the E2 elimination is at least partially rate-limiting for k_{cat}/K_m. More detailed mechanistic conclusions concerning the hydride transfer to NAD^+ from C3 will require the development of new synthetic procedures to make C1, C2 and C3 deuterated substrates.

5.7. Notes

[a] *Department of Chemistry, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. Fax: +1-778-782-3765; Tel: +1-778-782-8814; E-mail: bennet@sfu.ca*
[b] *Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. Fax: +1-778-782-5583*
† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

5.8. References

1. K. W. Moremen, M. Tiemeyer and A. V. Nairn, *Nat. Rev. Mol. Cell. Biol.*, 2012, **13**, 448.
2. I. Brockhausen, *EMBO Rep.*, 2006, **7**, 599.
3. B. Weissman and D. F. Hinrichs, *Biochemistry*, 1969, **8**, 2034.
4. S. I. Patenaude, N. O. L. Seto, S. N. Borisova, A. Szpacenko, S. L. Marcus, M. M. Palcic and S. V. Evans, *Nat. Struct. Biol.*, 2002, **9**, 685.

5. H. Wiegandt, *Behav. Brain Res.*, 1995, **66**, 85.
6. H. Tuppy and Staudenb.Wl, *Biochemistry*, 1966, **5**, 1742.Fre
7. M. L. Olsson, C. A. Hill, H. de la Vega, Q. Y. P. Liu, M. R. Stroud, J. Valdinocci, S. Moon, H. Clausen and M. S. Kruskall, *Transfus. Clin. Biol.*, 2004, **11**, 33.
8. Q. Y. P. Liu, G. Sulzenbacher, H. P. Yuan, E. P. Bennett, G. Pietz, K. Saunders, J. Spence, E. Nudelman, S. B. Levery, T. White, J. M. Neveu, W. S. Lane, Y. Bourne, M. L. Olsson, B. Henrissat and H. Clausen, *Nat. Biotechnol.*, 2007, **25**, 454.
9. P. M. Coutinho, E. Deleury, G. J. Davies and B. Henrissat, *J. Mol. Biol.*, 2003, **328**, 307.
10. G. J. Davies, T. M. Gloster and B. Henrissat, *Curr. Opin. Struct. Biol.*, 2005, **15**, 637.
11. B. L. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard and B. Henrissat, *Nucleic Acids Res.*, 2009, **37**, D233.
12. D. A. Comfort, K. S. Bobrov, D. R. Ivanen, K. A. Shabalin, J. M. Harris, A. A. Kulminskaya, H. Brumer and R. M. Kelly, *Biochemistry*, 2007, **46**, 3319.
13. S. C. Garman, L. Hannick, A. Zhu and D. N. Garboczi, *Structure*, 2002, **10**, 425.
14. M. Kiyohara, T. Nakatomi, S. Kurihara, S. Fushinobu, H. Suzuki, T. Tanaka, S. Shoda, M. Kitaoka, T. Katayama, K. Yamamoto and H. Ashida, *J. Biol. Chem.*, 2012, **287**, 693.
15. V. Lombard, H. G. Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, *Nucleic Acids Res.*, 2014, **42**, D490.
16. G. Davies, M. L. Sinnott and S. G. Withers, in *Comprehensive Biological Catalysis*, ed. M. L. Sinnott, Academic Press, San Diego, 1 edn., 1998, pp. 119.
17. D. J. Vocadlo and G. J. Davies, *Curr. Opin. Chem. Biol.*, 2008, **12**, 539.
18. S. S. Rajan, X. J. Yang, F. Collart, V. L. Y. Yip, S. G. Withers, A. Varrot, J. Thompson, G. J. Davies and W. F. Anderson, *Structure*, 2004, **12**, 1619.
19. S. Chakladar, L. Cheng, M. Choi, J. Liu and A. J. Bennet, *Biochemistry*, 2011, **50**, 4298.
20. V. L. Y. Yip, J. Thompson and S. G. Withers, *Biochemistry*, 2007, **46**, 9840.
21. V. L. Y. Yip, A. Varrot, G. J. Davies, S. S. Rajan, X. J. Yang, J. Thompson, W. F. Anderson and S. G. Withers, *J. Am. Chem. Soc.*, 2004, **126**, 8354.
22. V. L. Y. Yip and S. G. Withers, *Biochemistry*, 2006, **45**, 571.

23. G. Sulzenbacher, Q. P. Liu, E. P. Bennett, S. B. Levery, Y. Bourne, G. Ponchel, H. Clausen and B. Henrissat, *Biocatal. Biotransfor.*, 2010, **28**, 22.
24. A. Varrot, V. L. Y. Yip, Y. S. Li, S. S. Rajan, X. J. Yang, W. F. Anderson, J. Thompson, S. G. Withers and G. J. Davies, *J. Mol. Biol.*, 2005, **346**, 423.
25. J. R. Knowles, *CRC Crit. Rev. Biochem.*, 1976, **4**, 165.
26. T. H. Lowry and K. S. Richardson, *Mechanism and theory in organic chemistry*, Harper & Row, New York, 1987.
27. R. Guthrie and W. P. Jencks, *Acc. Chem. Res.*, 1989, **22**, 343

5.9. Supporting Information

Supporting Information Contents:

5.9.1. List of Abbreviations

app apparent

br Broad

BSA Bovine serum albumin

c concentration

CHES N-Cyclohexyl-2-aminoethanesulfonic acid

d Doublet

DTT Dithiothreitol

GH Glycosyl hydrolase

HEPES 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

LB Luria Broth

m multiplet

MES 2-(N-morpholino)ethanesulfonic acid

NAD^+ Nicotinamide adenine dinucleotide

NMR Nuclear magnetic resonance

PCR Polymerase chain reaction

PNPαGalNAc 4-nitrophenyl 2-acetamido-2-deoxy-α-D-galactopyranoside

t triplet

TB Terrific Broth

TBAF Tetrabutyl ammonium fluoride

TCEP Tris-(carboxyethyl)phosphine

TLC Thin-layer chromatography

TMS Tetramethylsilane

Trunc truncated

5.9.2. General Information

All chemicals were of analytical grade or better and were purchased from Sigma-Aldrich unless noted otherwise. Milli-Q water (18.2 MΩ cm^{-1}) was used for all kinetic experiments. All pH values were measured using a standard pH electrode attached to a VWR pH meter. All NMR spectra were acquired on either a Bruker 400, 500 or 600 MHz spectrometer. Chemical shifts are reported in parts per million downfield from signals for TMS. The signal residues from deuterated chloroform and external TMS salts (D_2O) were used for ^1H NMR spectral references; for ^{13}C NMR spectra, natural abundance signals from $CDCl_3$ and external TMS salts (D_2O) were used as references. Coupling constants (*J*) are reported in hertz. Melting points were determined on a Gallenkamp melting point apparatus and are not corrected. Optical rotations were measured on a Perkin-Elmer 341 polarimeter and are reported in units of deg cm^2 g^{-1} (concentrations reported in units of g per 100 mL). PNPαGalNAc (3a) was purchased from Sigma-Aldrich and used without further purification.

5.9.3. Experimental Details

Typical procedure for the synthesis of substrates: all aryl 2-acetamido-2-deoxy-α-D-galactopyranosides were synthesized from 2-deoxy-2-(methoxycarbonylamino)-1,3,4,6-tetra-*O*-acetyl-β-D-galactopyranoside using $SnCl_4$ as activator in a solvent of CH_2Cl_2. In a typical procedure, 2-deoxy-2-(methoxycarbonylamino)-1,3,4,6-tetra-*O*-acetyl-β-D-galactopyranoside **1** (1 g, 2.4 mmol) and the appropriate phenol (4.9 mmol) were dissolved in anhydrous CH_2Cl_2 (50 mL) and then $SnCl_4$ (0.3 mL, 2.4 mmol) was added, and the reaction mixture was stirred at ambient temperature under an inert atmosphere for 48 h. Following the addition of water (40 mL) the reaction mixture was neutralized by adding saturated $NaHCO_3$ (30 mL). The crude product was extracted from the aqueous layer using CH_2Cl_2 (3 × 40 mL), and the combined organic layers were washed with brine (2 × 25 mL), dried (Na_2SO_4) and concentrated under reduced pressure (typical α:β ratios 4:1 to 5:1). This crude material was purified by column chromatography using EtOAc-Hexane (35:65, EtOAc:Hexane) as the eluent to obtain the pure α-galactoside (yields 55–

65%). Removal of the carbamate was accomplished using TBAF according to a reported procedure to give crude aryl 2-amino-2-deoxy-α-D-galactopyranosides (yields 45–50%).[1] The resultant crude product was acetylated under standard conditions: pyridine (10 mL) and acetic anhydride (10 mL). The peracetylated product was extracted from the aqueous layer using CH_2Cl_2 (3 × 35 mL), and the combined organic layers were washed with brine (2 × 20 mL), dried (Na_2SO_4) and concentrated under reduced pressure and the excess pyridine was removed azeotropically using toluene. Purification of the final product was achieved by performing a column chromatography in 5% MeOH-CH_2Cl_2, to obtain the pure per-acetylated product (quantitative yields). Finally, the substrates were obtained by deprotection of the peracetylated material using Zemplen conditions (NaOMe/MeOH) followed by neutralization using Amberlite (H^+) resin to obtain the pure substrates in the series of aryl 2-acetamido-2-deoxy-α-D-galactopyranosides in 20–30% overall yield over six steps (individual overall yields are reported with the physical data for compounds 3b–f). Characterization data for the panel of substrates used in this study are given below:

3-Nitrophenyl-1,3,4,6-tetra-O-acetyl-2-deoxy-2-(methoxycarbonylamino)-α-d galactopyranoside (**2b**) $[\alpha]^{20}_D$ = +29.3 (*c* 1.0, $CHCl_3$) ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.95 (s, 1H, Ar), 7.92 (d, *J* = 8.1, 1H, Ar), 7.47 (t, 1H, Ar), 7.39 (d, *J* = 8.2, 1H, Ar), 5.67 (d, *J* = 3.1, 1H, H-1), 5.45 (app.d, 1H, H-4), 5.34 (dd, *J* = 11.4, 2.9, 1H, H-3), 4.97 (d, *J* = 9.5, 1H, NH), 4.48 - 4.52 (m, 1H, H-2), 4.26 (t, *J* = 6.5, 1H, H-5), 4.11 (m, 2H, H-6, 6′), 3.68 (s, 3H, OCH_3, $NHCOOCH_3$), 2.17 (s, 3H), 2.03 (s, 3H), 1.91 (s, 3H) (3× $OCOCH_3$); ^{13}C NMR (151 MHz, $CDCl_3$) δ 170.70, 170.25, 170.08, 156.52, 149.10, 130.25 (Ar C), 122.89 (Ar C), 117.91(Ar C), 111.87 (Ar C), 97.23 (C-1), 67.98 (C-5), 67.81(C-3), 66.98 (C-4), 61.60 (C-6), 52.48 (OCH_3, $NHCOOCH_3$), 49.55 (C-2), 20.63, 20.58, 20.38 (3 × OCH_3).

3-Chlorophenyl-1,3,4,6-tetra-O-acetyl-2-deoxy-2-(methoxycarbonylamino)-α-D-galactopyranoside (**2c**) $[\alpha]^{20}_{D}$ = +22.9 (*c* 0.8, $CHCl_3$) ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.23 (d, *J* = 8.2, 1H, ArH), 7.13 (s, 1H, ArH), 7.06 (d, *J* = 8.0, 1H, ArH), 6.96 (dd, *J* = 8.1, 1.7, 1H, ArH), 5.58 (d, *J* = 3.0, 1H, H-1), 5.45 (app d, 1H, H-4), 5.32 (dd, *J* = 11.3, 2.8, 1H, H-3), 4.94 (d, *J* = 9.6, 1H, NH), 4.44 - 4.48 (m, 1H, H-2), 4.26 (t, *J* = 6.3, 1H, H-5), 4.14 – 4.02 (m, 2H, H-6, 6′), 3.67 (s, 3H, OCH_3, $NHCOOCH_3$), 2.19 (s, 3H), 2.04 (s, 3H), 1.97 (s, 3H) (3 × $OCOCH_3$); ^{13}C NMR (126 MHz, $CDCl_3$) δ 170.77, 170.38, 170.21(3×CO, 3×$OCOCH_3$), 156.58, 130.45 (ArC), 123.34(ArC), 117.23 (ArC), 115.06 (ArC), 96.76 (C-1), 68.11 (C-3), 67.80 (C-5), 67.18 (C-4), 61.70 (C-6), 52.53 (OCH_3, $NHCOOCH_3$), 49.63 (C-2), 20.74, 20.68, 20.55 (3×OCH_3).

4-Chlorophenyl-1,3,4,6-tetra-O-acetyl-2-deoxy-2-(methoxycarbonylamino)-α-D-galactopyranoside (**2d**) $[\alpha]^{20}_{D}$ = +20.9 (*c* 0.32, $CHCl_3$) ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.26 (d, *J* = 8.9, 2H, ArH), 7.00 (d, *J* = 9.0, 2H, ArH), 5.56 (d, *J* = 3.4, 1H, H-1), 5.44 (d, 1H, H-4), 5.32 (dd, *J* = 11.3, 2.9, 1H, H-3), 4.98 (d, *J* = 9.8, 1H, NH), 4.45 (m, 1H, H-2), 4.24 (t, *J* = 6.6, 1H, H-5), 4.07 (m, 2H, H-6, 6′), 3.66 (s, 3H, OCH_3, $NHCOOCH_3$), 2.18 (s, 3H), 2.04 (s, 3H), 1.94 (s, 3H) (3× $OCOCH_3$); ^{13}C NMR (126 MHz, $CDCl_3$) δ 170.76, 170.26, 170.18 (3×CO, 3×$OCOCH_3$), 156.53, 154.57, 129.57 (ArC), 117.97 (ArC), 96.92 (C-1), 68.10 (C-3), 67.67 (C-5), 67.14 (C-4), 61.59 (C-6), 52.49 (OCH_3, $NHCOOCH_3$), 49.61(C-2), 20.72, 20.66, 20.53 (3×OCH_3).

Phenyl 1,3,4,6-tetra-O-acetyl-2-deoxy-2-(methoxycarbonylamino)-α-D-galactopyranoside (**2e**) $[\alpha]^{20}_{D}$ = +19.5 (*c* 1.12, $CHCl_3$) ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.29 - 7.33 (m, 2H, Ar), 7.05 - 7.08 (m, 3H, Ar) 5.60 (d, *J* = 3.4, 1H, H-1), 5.45 (app.d, 1H, H-4), 5.35 (dd, *J* = 11.4, 3.0, 1H, H-3), 5.0 (d, *J* = 9.8, 1H, NH), 4.48 - 4.43 (m, 1H, H-2), 4.28 (t, *J* = 6.5, 1H, H-5), 4.03 - 4.13 (m, 2H, H-6, 6′), 3.65 (s, 3H, OCH_3, $NHCOOCH_3$), 2.18 (s, 3H), 2.03 (s, 3H), 1.92 (s, 3H) (3× $OCOCH_3$); ^{13}C NMR (126 MHz, $CDCl_3$) δ 170.75, 170.30, 170.23 (3 × CO), 156.56, 156.02, 129.65 (ArC), 116.57 (ArC), 96.64(C-1), 68.29 (C-5), 67.50 (C-3), 67.23 (C-4), 61.57 (C-6), 52.44 (OCH_3, $NHCOOCH_3$), 49.64(C-2), 20.74, 20.68, 20.53 (3×OCH_3).

4-Methoxyphenyl-1,3,4,6-tetra-O-acetyl-2-deoxy-2-(methoxycarbonylamino)-α-D-galactopyranoside (**2f**) $[\alpha]^{20}_{D}$ = +20.5 (*c* 0.58, $CHCl_3$) ^{1}H NMR (600 MHz, $CDCl_3$) δ 6.99

(d, *J* = 9.2, 2H, ArH), 6.83 (d, *J* = 9.1, 2H, ArH), 5.47 (d, *J* = 3.3, 1H, H-1), 5.45 (app.d, *J* = 2.8, 1H, H-4), 5.32 (dd, *J* = 11.3, 3.0, 1H, H-3), 5.01 (d, *J* = 9.9, 1H, NH), 4.43-4.442 (m, 1H, H-2), 4.32 (t, *J* = 6.6, 1H, H-5), 4.13-4.04 (m, 2H, H-6, 6′), 3.77 (s, 3H, OCH_3), 3.66 (s, 3H, OCH_3, $NHCOOCH_3$), 2.17 (s, 3H), 2.03 (s, 3H), 1.96 (s, 3H) (3× $OCOCH_3$); ^{13}C NMR (151 MHz, $CDCl_3$) δ 170.74, 170.30, 170.24 (3×CO, 3×$OCOCH_3$), 156.58, 155.50, 150.06, 118.00 (ArC), 114.67 (ArC), 97.63 (C-1), 68.34 (C-3), 67.40 (C-5), 67.31(C-4), 61.69 (C-6), 55.64 (C, OCH_3), 52.42 (OCH_3, $NHCOOCH_3$), 49.69 (C-2), 20.72, 20.67, 20.57 (3×OCH_3).

3-Nitrophenyl 2-acetamido-2-deoxy-α-D-galactopyranoside (**3b**) Yield 25% (6 steps); Mpt = 159–161 °C, $[\alpha]^{20}_D$ =+271.4 (*c* 0.29, H_2O); ^1H NMR (600 MHz, D_2O) δ 7.97 (t, *J* = 2.2, 1H, ArH), 7.95 (d, *J* = 7.8, 1H, ArH), 7.56 (t, *J* = 8.2, 1H, ArH), 7.51 (dd, *J* = 8.3, 2.3, 1H, ArH), 5.73 (d, *J* = 3.6, 1H, H-1), 4.37 (dd, *J* = 11.0, 3.6, 1H, H-2), 4.15 (dd, *J* = 11.1, 3.2, 1H, H-3), 4.05- 4.02 (m, 2H, H-4, H-5), 3.71 (m, 2H, H-6, 6′), 2.02 (s, 3H, $NHCOCH_3$); ^{13}C NMR (151 MHz, D_2O) δ 174.77 (CO, $NHCOCH_3$), 156.35 (ArC), 148.67 (ArC), 130.51(ArCH), 123.84 (ArCH), 117.85 (ArCH), 111.93 (ArCH), 96.26 (C-1), 72.16 (C-5), 68.38 (C-4), 67.58 (C-3), 61.06 (C-6), 49.59 (C-2), 21.91 (OCH_3, $NHCOCH_3$); HRMS expected for $C_{14}H_{18}N_2O_8$ is 365.0961 ($M^+ + Na^+$): Found 365.0953.

3-Chlorophenyl 2-acetamido-2-deoxy-α-D-galactopyranoside (**3c**) Yield 25% (6 steps); Mpt = 180-181 °C (decomp.), $[\alpha]^{20}_D$ =+210.5 (*c* 0.63, H_2O) ^1H NMR (600 MHz, D_2O) δ 7.32 (t, *J* = 8.0, 1H, ArH), 7.20 (t, *J* = 2.2, 1H, ArH), 7.14 – 7.10 (m, 1H, ArH), 7.06 (dd, *J* = 8.4, 2.4, 1H, ArH), 5.61 (d, *J* = 3.7, 1H, H-1), 4.33 (dd, *J* = 11.1, 3.7, 1H, H-2), 4.12 (dd, *J* = 11.0, 3.2, 1H, H-3), 4.05 – 4.02 (m, 2H, H-4, H-5), 3.74 – 3.68 (m, 2H, H-6, 6′), 2.02 (s, 3H, $NHCOCH_3$); ^{13}C NMR (151 MHz, D_2O) δ 174.77 (CO, $NHCOCH_3$), 156.83 (ArC), 134.30 (ArC), 130.78 (ArCH), 122.99 (ArCH), 117.35 (ArCH), 115.49 (ArCH), 96.22 (C-1), 71.94 (C-5), 68.42 (C-4), 67.60 (C-3), 61.06 (C-6), 49.67 (C-2), 21.89 (OCH_3, $NHCOCH_3$); HRMS expected for $C_{14}H_{18}ClNO_6$ 354.0720 ($M^+ + Na^+$): Found 354.0714 ($M^+ + Na^+$).

4-Chlorophenyl 2-acetamido-2-deoxy-α-D-galactopyranoside (**3d**) Yield 20% (6 steps); Mpt = 218–220 °C, $[\alpha]^{20}_D$ =+190.5 (*c* 0.8, H_2O) ^1H NMR (600 MHz, D_2O) δ 7.40 (d, *J* = 9.0, 2H), 7.14 (d, *J* = 9.0, 2H), 5.63 (d, *J* = 3.6, 1H, H-1), 4.37 (dd, *J* = 11.1, 3.7, 1H, H-2), 4.16 (dd, *J* = 11.1, 3.2 Hz, H-3), 4.09 (m, 2H, H-4, H-5), 3.79 – 3.72 (m, 2H, H-6, 6′), 2.07 (s,

3H, OCH3, $NHCOCH_3$); ^{13}C NMR (151 MHz, D_2O) δ 174.25 (CO, $NHCOCH_3$), 154.35 (ArC), 129.02 (ArC), 126.82 (ArCH), 118.08 (ArCH), 95.85 (C-1), 71.37 (C-5), 67.90 (C-4), 67.11(C-3), 60.56 (C-6), 49.20 (C-2), 21.40 (OCH_3, $NHCOCH_3$); HRMS expected for $C_{14}H_{18}ClNO_6$ 354.0720 ($M^+ + Na^+$): Found 354.0717 ($M^+ + Na^+$).

Phenyl 2-acetamido-2-deoxy-α-D-galactopyranoside (**3e**) Yield 21% (6 steps); Mpt = 249–250 °C (decomp.), $[\alpha]^{20}_D$ =+220.0 (*c* 0.56, H_2O) ^1H NMR (600 MHz, D_2O) δ 7.43 (t, *J* = 8.0, 2H), 7.17 (dd, *J* = 12.6, 7.7, 3H), 5.66 (d, *J* = 3.6, 1H, H-1), 4.37 (dd, *J* = 11.1, 3.5, 1H, H-2), 4.19 (dd, *J* = 11.2, 3.2, 1H, H-3), 4.13 (t, *J* = 6.1, 1H, H-5), 4.10 (app d, 1H, H-4), 3.75-3.77 (m, 2H, H-6, 6′), 2.07 (s, 3H, OCH_3, $NHCOCH_3$); ^{13}C NMR (151 MHz, D_2O) δ 174.27 (CO, $NHCOCH_3$), 155.69, 129.39 (ArC), 122.64 (ArC), 116.70 (ArC), 95.73 (C-1), 71.26 (C-5), 67.94 (C-4), 67.15 (C-3), 60.56 (C-6), 49.28 (C-2), 21.40 (OCH_3, $NHCOCH_3$); HRMS expected for $C_{14}H_{19}NO_6$ 320.1110 ($M^+ + Na^+$): Found 320.1103 ($M^+ + Na^+$).

4-Methoxyphenyl 2-acetamido-2-deoxy-α-D-galactopyranoside (**3f**) Yield 30% (6 steps); Mpt = 175–177 °C, $[\alpha]^{20}_D$ =+200.8 (*c* 0.35, H_2O) ^1H NMR (600 MHz, D_2O) δ 7.05 (d, *J* = 9.1, 2H, ArH), 6.92 (d, *J* = 9.1, 2H, ArH), 5.43 (d, *J* = 3.7, 1H, H-1), 4.26 (dd, *J* = 10.9, 3.5, 1H, H-2), 4.09 – 4.05 (m, 2H, H-3, H-5), 4.01 (d, *J* = 2.5, 1H, H-4), 3.74 (s, 3H, OCH_3), 3.68 (app. d, 2H, H-6, 6′), 2.00 (s, 3H, $NHCOCH_3$); ^{13}C NMR (151 MHz, D_2O) δ 174.26 (CO, $NHCOCH_3$), 154.16 (ArC), 150.03 (ArC), 118.40 (ArCH), 114.56 (ArCH), 96.80 (C-1), 71.21(C-5), 67.96 (C-4), 67.13 (C-3), 60.59 (C-6), 55.31(OCH_3), 49.33 (C-2), 21.41(OCH_3, $NHCOCH_3$); HRMS expected for $C_{15}H_{21}NO_7$ 350.1216 ($M^+ + Na^+$): Found 350.1210 ($M^+ + Na^+$).

5.9.4. Cloning of *E. meningosepticum* α-N-acetylgalactosaminidase (α-NAGAL)

A truncated version of *E. meningosepticum* α-*N*-acetylgalactosaminidase (EMBL Nucleotide Sequence Database accession number AM039444) lacking the first 17 amino acids, which is the same truncation as that reported by Lui et al.,[2] was amplified by PCR from genomic DNA (ATCC No. 51720D) using the Thermo Scientific Phusion High-Fidelity PCR system with primers NAGAL-EcoRI: 5'CCGGAATTCAAAAAGGTAAGAATAGCTTTTT3' and NAGAL-HindIII:

5'CCCAAGCTTGTAGTCGTCATTTATTGCAAATG3' introducing the EcoRI and HindIII restriction sites in the forward and reverse primer respectively. The PCR fragment was digested with *EcoRI* and *HindIII* and inserted into correspondingly digested pET28a vector (Strategene) for expression of a C-termius His_6 tagged protein. The construct was used to transform *Escherichia coli* BL21 (DE3 Gold) (Novagen) competent cells. Transformants were plated onto LB plates containing 100 μg/mL kanamycin. Plasmid from a single colony was purified and DNA sequencing by Macrogen using T7 promoter and T7 terminator primers provided the nucleotide composition of the amplified gene. The nucleotide sequence was aligned with AM039444, which was amplified from genomic DNA (ATCC 13253),[2] using BioEdit and the amino acid differences were identified (Figure S1).

5.9.5. Expression and purification of *E. meningosepticum* α-*N*-acetylgalactosaminidase (α-NAGAL)

The tagged enzyme was grown in TB supplemented with 100 μg/ml kanamycin. Expression was induced at OD_{600} ~ 0.6 by addition of isopropyl β-D-1-thiogalactoside (IPTG) at a final concentration of 1 mM. Cell pellet from 1 L culture was resuspended in 50 mL of lysis buffer, 40 mM $NaPO_4$, pH 6.8, 100 mM NaCl, 1% lysozyme and a protease inhibitor cocktail tablet (Roche Diagnostics) followed by sonication (20 sec ON/ 60 sec OFF at 50% capacity). The lysate was centrifuged for 30 min at 13000 rpm and the supernatant was loaded onto a HisTrap column (GE Life Sciences). The protein was eluted by increasing the concentration of imidazole in the elution buffer. The fractions containing enzyme as determined by SDS-PAGE were dialyzed against Tris buffer (40 mM containing NaCl 100 mM) overnight at 4 °C. These fractions were then concentrated using a 10 kDa Micorcon centrifugal filter (Millipore). The protein was then concentrated by centrifugation through a 10 kDa filter and its concentration was assessed (Bradford Assay).

5.9.6. Typical Conditions for the Measurement of Michaelis–Menten Parameters.

The concentration of α-NAGAL was chosen such that less than 10% of the total substrate was consumed during the assay. For each assay, the enzyme was incubated in

the appropriate buffer at 37 °C for 5 min. After which the reaction was initiated by the addition of substrate. The initial rate of hydrolysis was followed spectrophotometrically at the wavelength of maximal absorbance change. Typically, the substrate concentration was varied between 40 µM and 500 mM and the measured initial rate versus concentration data were fit to a standard Michaelis-Menten equation.

To determine the effect of pH on enzymatic activity, kinetics parameters were measured over a pH range of 6.3–8.8. The buffers used were MES (20 mM, MES-NaOH, pH 6.0–6.7), HEPES (20 mM, pH 6.5–8.2), and CHES (20 mM, pH 8.5–9.5). Typical assay conditions were: α-NAGAL (final concentration of 0.23 µg/mL) was incubated at 37 °C with the appropriate buffer containing NaCl (50 mM) and BSA (0.1 % w/v) for 5 min prior to addition of substrate PNPαGalNAc and the hydrolysis reaction was monitored by UV-vis spectroscopy. The difference in extinction coefficients ($\Delta\varepsilon$) for PNPαGalNAc and the released 4-nitrophenolate was determined at each pH value and, the initial rate measurements were fit to a standard Michaelis-Menten equation. The so obtained values for k_{cat} and k_{cat}/K_m and the associated pH values were fit to the equation given below using the computer program prism.

$$k = \frac{k_L}{1 + \left(K_1/[H^+]\right)^{h1} + \left(K_1K_2/[H^+]^2\right)^{h2}} + \frac{k_M}{1 + \left([H^+]/K_1\right)^{h1} + \left(K_2/[H^+]\right)^{h2}} + \frac{k_H}{1 + \left([H^+]^2/K_1K_2\right)^{h1} + \left([H^+]/K_2\right)^{h2}}$$

Where k is the observed rate constant (k_{cat} or k_{cat}/K_m), k_L and k_H are the rate constants at low and high pH, while k_M is the rate constant for the active form of the enzyme. K_1 and K_2 are the two apparent equilibrium constants for protonation and *h1* and *h2* are the Hill coefficients. In both fits (k_{cat} vs. pH and k_{cat}/K_m vs. pH) the fitted value for k_H was within error equal to zero.

5.9.7. Kinetic Investigation of Cofactor Dependence *E. meningosepticum* α-*N*-acetylgalactosaminidase (α-NAGAL)

A sample of α-NAGAL (final concentration of 0.23 µg/mL) in buffer (50 mM HEPES, pH 7.5) was pre-incubated with NAD^+ (5–100 µM) at 37 °C for 5 min. PNPαGalNAc was added to the solution to initiate the hydrolysis reaction at 400 nm. Similar experiment was

carried out for the dependence on reducing agent DTT, (5–100 μM) and divalent metal ion Mn^{2+} (10–100 μM). The enzyme showed no enhancement in catalytic rate of hydrolysis upon addition of either of the cofactors.

5.9.8. Product Studies

^1H NMR spectroscopy (500 MHz) was employed to identify the stereochemical course of the enzyme-catalyzed reaction. The reaction conditions involved incubation of enzyme α-NAGAL (3.0 μg/ mL) in buffer (50 mM HEPES, pH 7.5) containing CD_3OD (5 M) at 37 °C,. After addition of PNPαGalNAc (2.1 mg) the reaction was allowed to proceed at 37 °C for ~10 h until TLC analysis (1:4 v/v MeOH:EtOAc) showed no remaining starting material. Removal of the enzyme by centrifugal ultra-filtration (10KDa filtration unit) at 4 °C was followed by lyophilization of the resultant solution. The resultant white solid was dissolved D_2O and a ^1H NMR spectrum was acquired.

5.9.9. Linear Free Energy Relationship – Brønsted Analysis

A series of substrates with varying leaving groups were synthesized to perform a Brønsted analysis. Full Michaelis-Menten curves were measured, in buffer (50 mM HEPES, pH 7.5) at 37 °C, for each substrate using the above protocol.

Figure S1: Protein sequence alignment for the *E. meningosepticum* α-*N*-acetylgalactosaminidase used in this study and that reported by Liu et al. (ATCC 13253).[2]

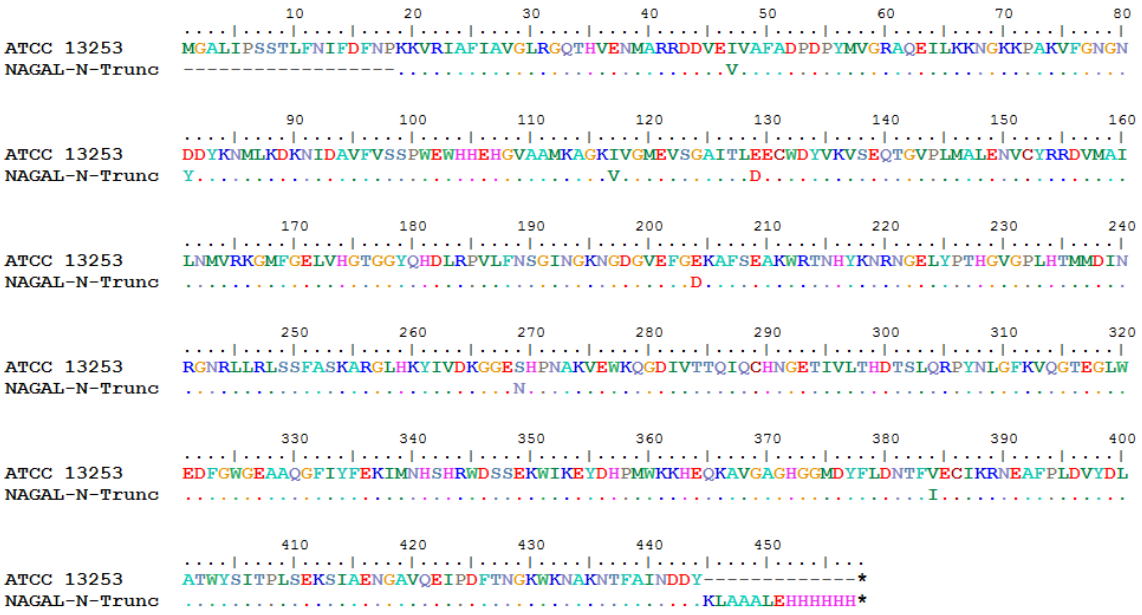


Figure S2: DNA sequence alignment for the *E. meningosepticum* α-*N*-acetylgalactosaminidase gene used in this study and that reported by Liu et al. (ATCC 13253).[22]

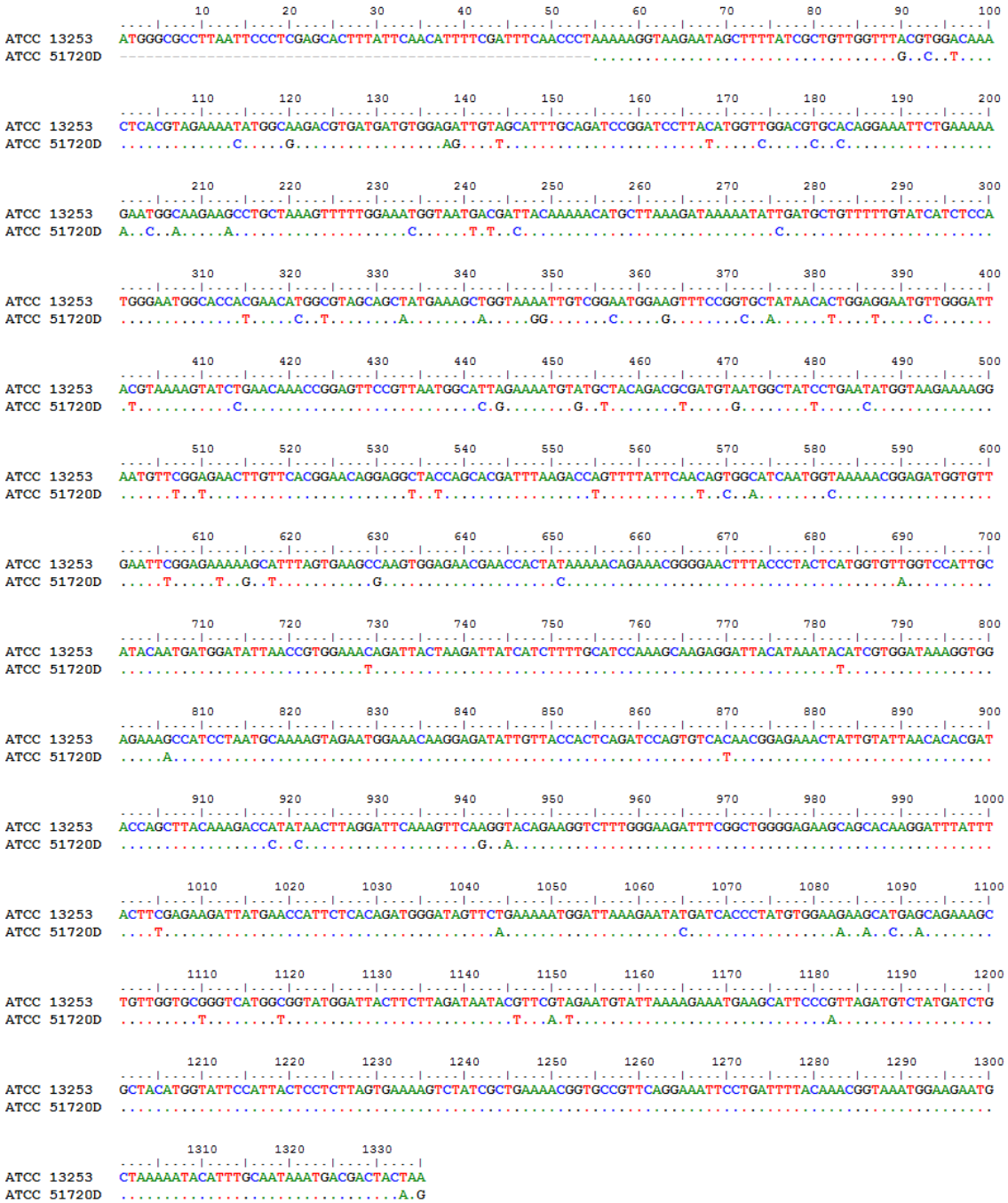


Fig. S3 Comparison of the active sites for GH4 (left) and GH109 (right) enzymes. Key features of the GH4 enzyme are the Mn^{2+} cation, tyrosine 265, and aspartic acid 172, which acidify the C-2 proton, abstract the C-2 proton and catalyse aglycone departure, respectively. Key features of the GH109 enzyme are: methionine 375 and tyrosine 307, which bind to the *N*-acetyl group and tyrosine 179 that is the closest possible base to the C-2 proton.

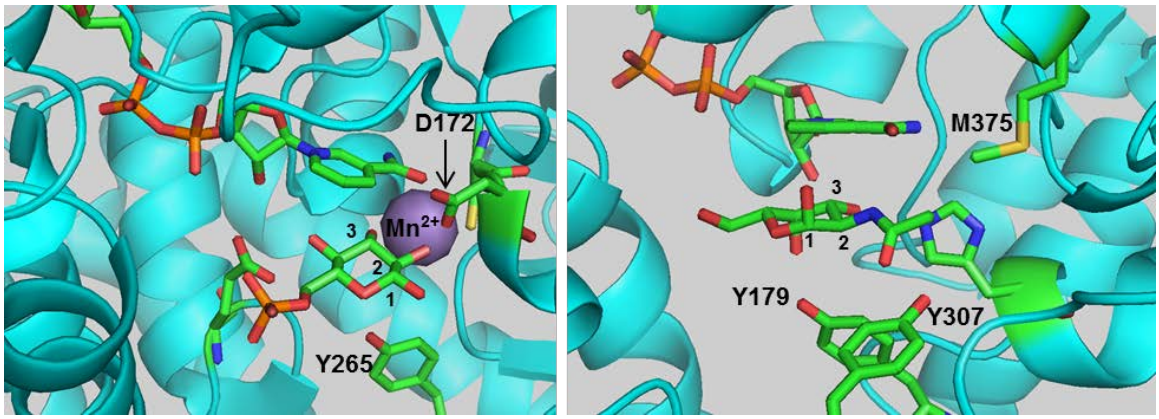


Figure S4. 3-Chlorophenyl 2-acetamido-2-deoxy-α-D-galactopyranoside (600 MHz, D_2O).

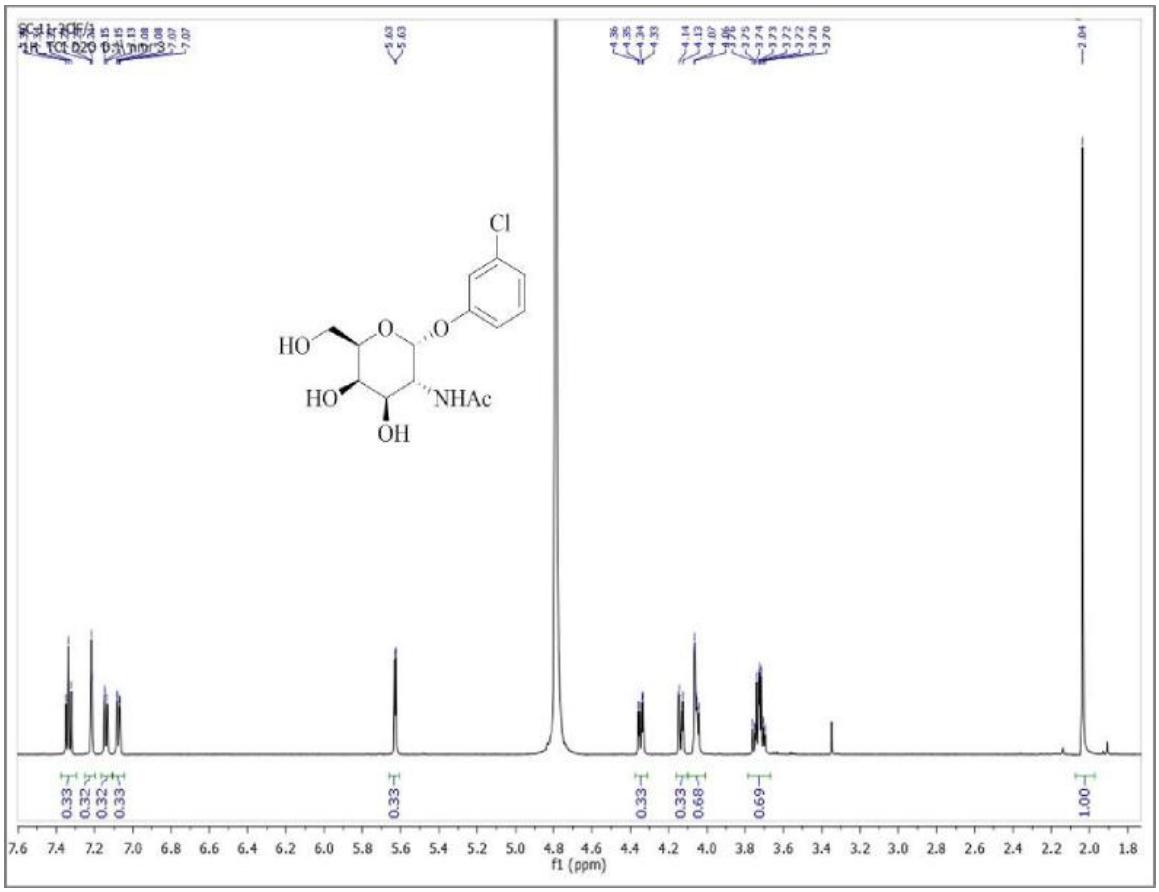


Figure S5. 3-Nitrophenyl 2-acetamido-2-deoxy-α-D-galactopyranoside (600 MHz, D_2O).

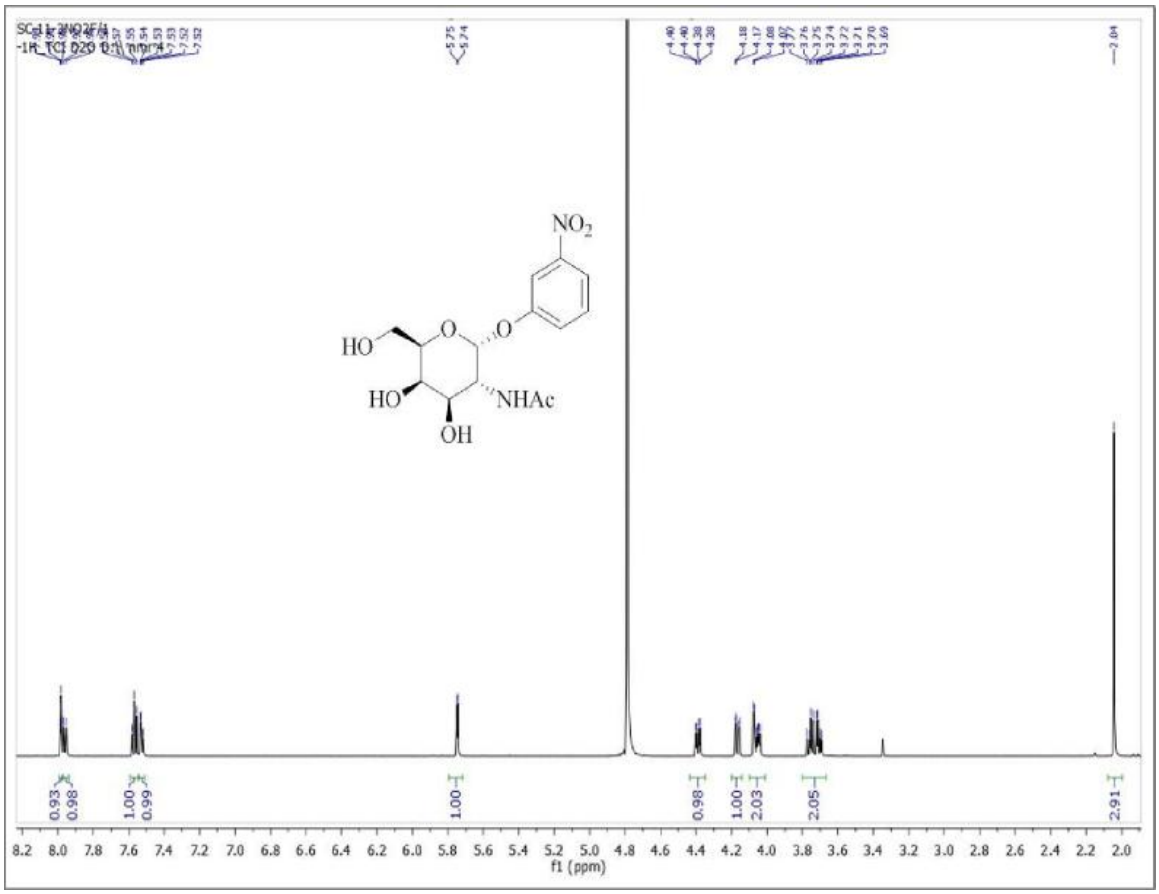


Figure S6. 4-Methoxyphenyl 2-acetamido-2-deoxy-α-D-galactopyranoside (600 MHz, D_2O).

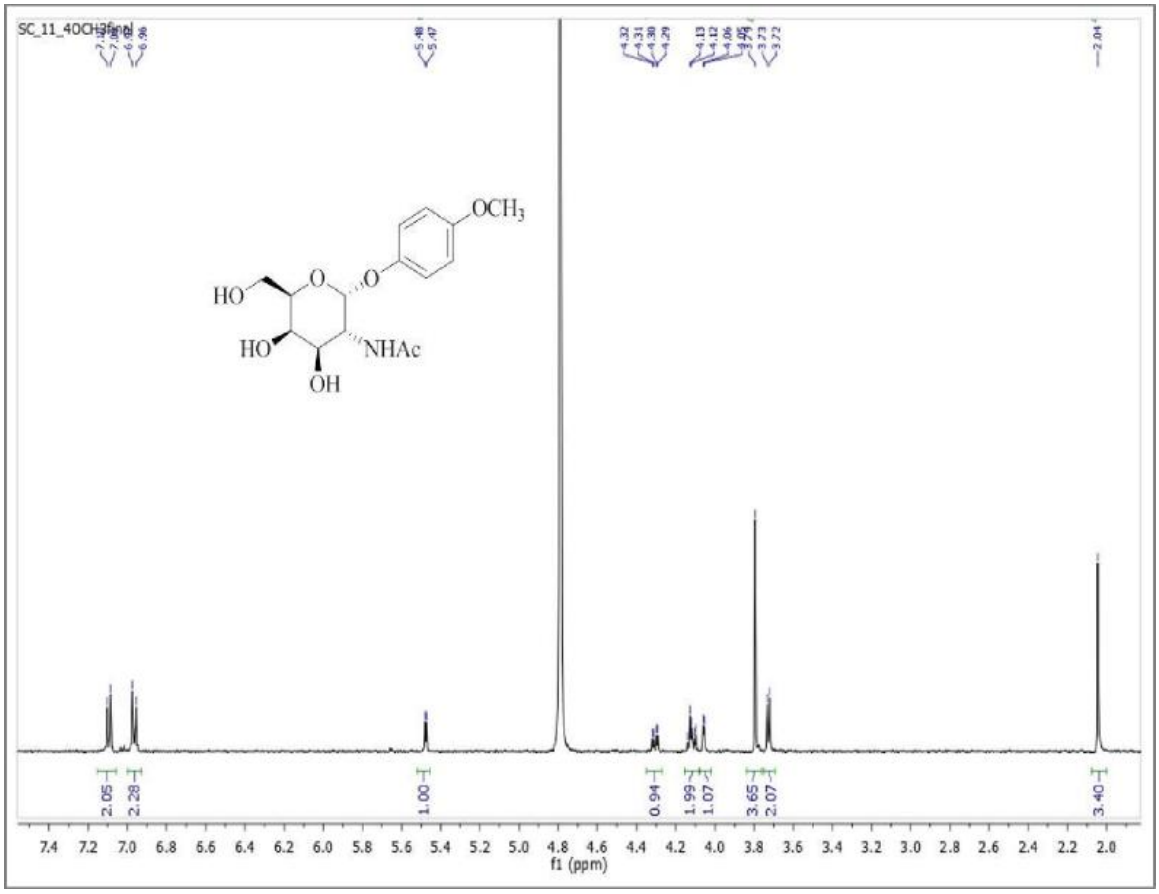
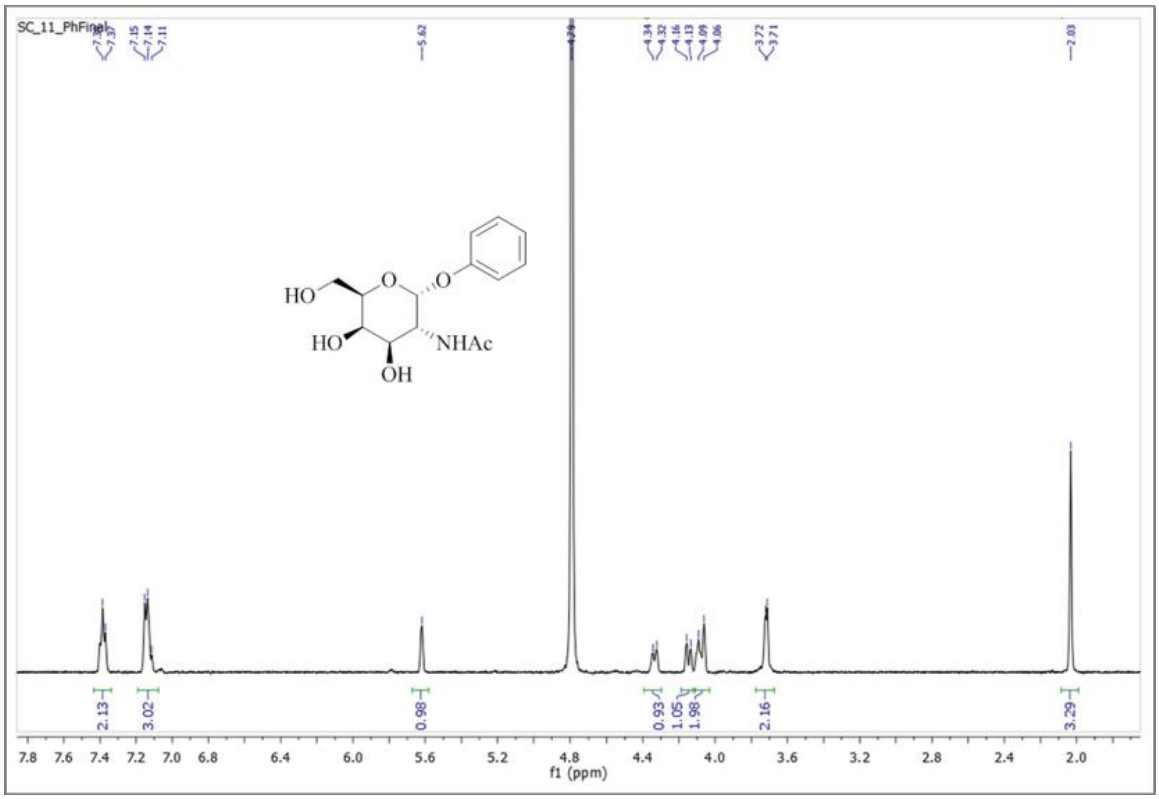


Figure S7. Phenyl 2-acetamido-2-deoxy-α-D-galactopyranoside (600 MHz, D_2O).



5.9.10. Reference

1. U. Jacquemard, V. Beneteau, M. Lefoix, S. Routier, J. Y. Merour and G. Coudert, *Tetrahedron*, 2004, **60**, 10039.

2. Q. Y. P. Liu, G. Sulzenbacher, H. P. Yuan, E. P. Bennett, G. Pietz, K. Saunders, J. Spence, E. Nudelman, S. B. Levery, T. White, J. M. Neveu, W. S. Lane, Y. Bourne, M. L. Olsson, B. Henrissat and H. Clausen, *Nat. Biotechnol.*, 2007, **25**, 454.

Chapter 6. Future Work and Directions

6.1. Directed Evolution of the Sialidase from *Micromonospora viridifaciens* into a Kdnase

Following isolation of selected enzymes with improved Kdnase activity, the aim is to introduce further random and site directed mutations along the gene and select for clones with further improvement of Kdnase activity and reduction of sialidase activity.

A viable tactic is to evolve a generation of the genetic library based on previous successful clones. Also, we plan to use our positive and negative selection protocols on our current library against 2-3 linked Kdn and Neu5Ac glycosides.

In parallel to this project, we aim to generate enzymes capable of hydrolyzing *O*-acetylated sialoglycans, which despite their documented role in bacterial pathogenicity and numerous human cancers such as human colorectal cancer, are scantily understood[109]. The mutagenesis library is prepared and the chemical compounds, which have *N*-acyl groups rather than *O*-acyl groups in order to be hydrolytically stable for the selection, have been synthesized. The next step is to run the positive and negative selection panels and based on the isolated clones, perform further enhancements.

On cell surfaces, terminal sialic acids are attached in a linkage specific bond to the penultimate sugars such as a galactose. Most *exo*-sialidases exhibit substrate specificity regarding sialic acid linkage. Generally, α2-3 linkages are hydrolyzed more readily than α2-6 bonds. These properties are both biologically and practically significant. From a glycobiology point of view there is great interest in isolation or evolution of a sialidase which would exclusively hydrolyze α2-6 linkages as there is an abundant presence of α2-6 linked sialic acids on human cell surfaces. To study the complex biological effects that result from the presentation of various sialic acids on cell surfaces, there is a need to generate sialidases with high specific activities for individual surface modifications.

A subgroup of sialidases, called *trans*-sialidases, transfer sialic acid directly from one glycosidic linkage onto another sugar residue instead of a water molecule. A 3D

overlap of the structure of *Mv*NA with one such *trans*-sialidase, *Trypanosoma cruzi trans*-sialidase (*Tc*TS), the agent of Chagas' disease in humans and animals, shows a high degree of structural similarity with the exception of presence of an additional 4 amino acid residues in a Trp containing loop in *Tc*TS. We have created *Mv*NA mutants where the additional amino acids are included in loop. We have also created a genetic library of saturation mutations of W loop and would like to monitor the hydrolysis activity of mutant enzymes by running a selection scheme similar to the Kdnase screen/selection protocols. The chemical compounds α-Neu5Ac-(2→3)-β-D-galactopyranosyl chloramphenicol and α-Neu5Ac-(2→6)-β-D-Gal-Tyrosine are synthesized (Figure 6.1) and the next step would be to isolate individual clones with enhanced 2,6 cleavage activity. Following identification of mutants capable of selective cleavage of α2-6 linkages, we would then perform a stringent selection screen to identify the mutants specifically hydrolyzing α-2-6 linkages. A second generation of random mutation along the entire length of the *Mv*NA sequence would allow for identification of highly active enzymes.

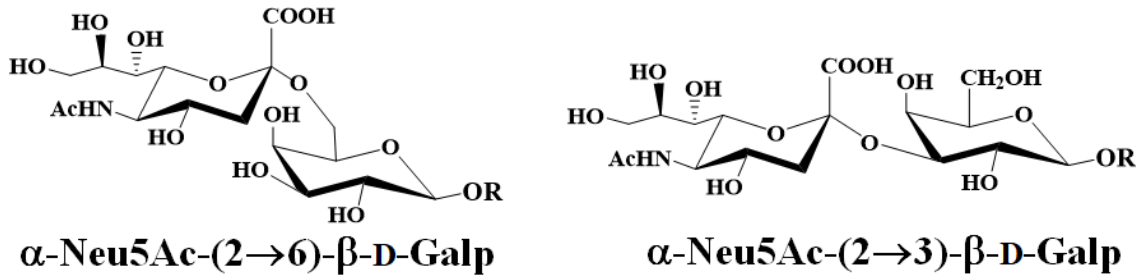


Figure 6.1 Compound synthesized for the positive and negative selection protocol.

6.2. A New Class of Glycoside Hydrolase Mechanism-Based Covalent Inhibitors: Glycosylation Transition State Conformations

The two classes of two classes of carbasugar analogue covalent inhibitors discussed in our paper will be useful research tools for biological studies as unlike other covalent inactivators, they show a time dependent loss and return of enzymatic activity. Our two classes of reversible covalent inhibitors could be used to monitor cellular responses to time-dependent changes in glycoside hydrolase activity.

As a continuation of our study of reversible covalent inhibition of glycoside hydrolases, we wanted to determine the leaving group dependency of enzymatic inhibition of yeast α-glucosidase by covalent inhibitors of both allylic and bicyclo nature. To this end we have synthesized five allylic inhibitors with different leaving groups. Figure 6.2 shows the structures of these inhibitors.

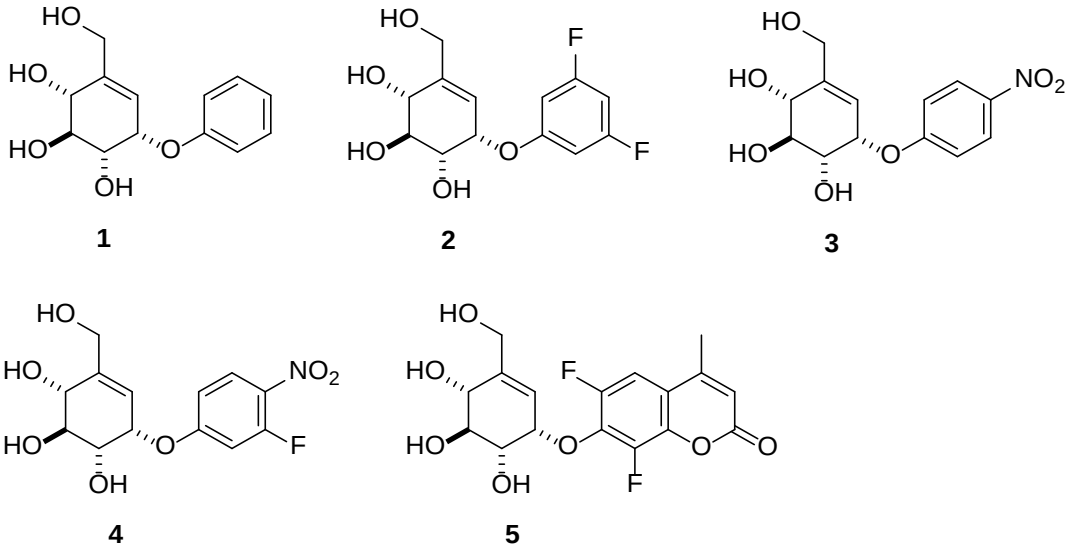


Figure 6.2 Structure of allylic inhibitors.

The measured kinetic parameters of inhibition of the yeast α-glucosidase by these allylic inhibitors, listed in Table 6.1, shows that there is little or no dependency on the leaving group ability for covalent labelling of the enzyme. This suggests that rate-limiting step for inactivation is a non-chemical step, for example a conformation change.

Table 6.1 Kinetic parameters of inhibition assays with allylic inhibitors.

Inactivator Compounds	K_i (μM)	k_{inact} (s^{-1})	k_{inact}/K_i (M^{-1} s^{-1})	pK_a (ArOH)
1	1600 ± 1100	$(6.3 \pm 3.5) \times 10^{-3}$	3.8 ± 0.4	9.98
2	92 ± 13	$(2.84 \pm 0.12) \times 10^{-3}$	30.7 ± 4.5	8.26
3 [a]	570 ± 90	$(9.05 \pm 0.63) \times 10^{-4}$	1.59 ± 0.27	7.15
4	210 ± 50	$(1.59 \pm 0.13) \times 10^{-3}$	7.7 ± 1.9	6.10
5	10.5 ± 2.3	$(2.31 \pm 0.18) \times 10^{-3}$	220 ± 51	4.70

a: Data taken from "New Class of Glycoside Hydrolase Mechanism-Based Covalent Inhibitors: Glycosylation Transition State Conformations". Saeideh Shamsi Kazem Abadi, Michael Tran, Anuj K. Yadav, Pal John Pal Adabala, Saswati Chakladar, and Andrew J. Bennet. Journal of the American Chemical Society 2017 139 (31), 10625-10628

We also synthesized two bicyclo inhibitors with different leaving groups, Figure 6.3, and performed inhibition assays. The data is listed in Table 6.2.

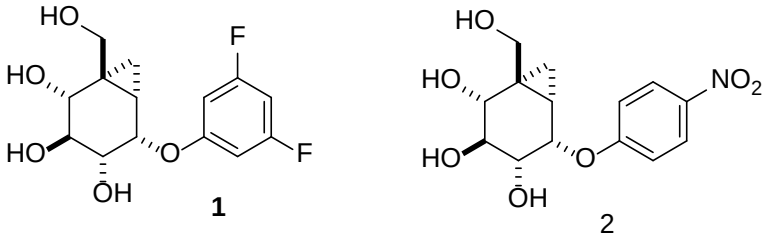


Figure 6.3 Structures of covalent Inhibitors that are bicyclo[4.1.0]heptyl glycoside mimics.

Table 6.2 Kinetic parameters of inhibition assays with bicyclo inhibitors.

Inactivator Aglycone	K_i (μM)	k_{inact} (s^{-1})	k_{inact}/K_i (M^{-1} s^{-1})	pK_a (ArOH)
3,5-Difluoro	180 ± 20	$(7.03 \pm 0.14) \times 10^{-5}$	0.391 ± 0.044	8.26
4-NO_2 [a]	285 ± 45	$(1.82 \pm 0.14) \times 10^{-3}$	6.4 ± 1.1	7.15

a: Data taken from "New Class of Glycoside Hydrolase Mechanism-Based Covalent Inhibitors: Glycosylation Transition State Conformations". Saeideh Shamsi Kazem Abadi, Michael Tran, Anuj K. Yadav, Pal John Pal Adabala, Saswati Chakladar, and Andrew J. Bennet. Journal of the American Chemical Society 2017 139 (31), 10625-10628

Based on the limited data available for the bicyclo inhibitors, it seems likely that, in contrast to the allylic covalent inhibitors, a large effect on the inhibition rate constants occurs on changing the leaving group of the bicyclo compounds. A result that suggests

the alkylation step is the rate limiting step. A definitive conclusion will have to wait until more bicyclo compounds, with various leaving groups, have been synthesized and tested for their inhibitory activity.

6.3. Characterization of Reaction Coordinates for Covalent Labeling of Glycosidase by Cyclohexene Inhibitors

We will be exploring the mechanism of inhibition more thoroughly by synthesizing compounds with various leaving groups and measuring kinetic parameters of the inhibition.

6.4. A Mechanistic Study on the α-N-acetylgalactosaminidase from *E. meningosepticum*: A Family 109 Glycoside Hydrolase

The *E*. *meningosepticum* α-*N*-acetylgalactosaminidase hydrolyses 2-acetamido-2-deoxy-α-D-galactopyranosides by a NAD^+-mediated oxidation followed by an α,β-elimination to give an enone Michael acceptor intermediate. This intermediate undergoes hydration along with proton and hydride transfer to generate 2-acetamido-2-deoxy-α-D-galactose as the reaction product. A step after elimination likely limits k_{cat}, and the elimination is at least partially rate-limiting for k_{cat}/K_m. More detailed mechanistic conclusions concerning the hydride transfer to NAD^+ from C3 will require the development of new synthetic procedures to make C1, C2 and C3 deuterated substrates.