University of Alberta

X-ray Crystallographic Studies of Human and Bacterial β-Hexosaminidase: Understanding the Molecular Basis of Tay-Sachs and Sandhoff Disease

by

Brian L. Mark

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Biochemistry

Edmonton, Alberta

2003

Your file Votre référence
ISBN: 0-612-82138-2
Our file Notre référence
ISBN: 0-612-82138-2

University of Alberta

Library Release Form

Name of Author: Brian L. Mark

Title of Thesis: X-ray Crystallographic Studies of Human and Bacterial β-Hexosaminidase: Understanding the Molecular Basis of Tay-Sachs and Sandhoff Disease

Degree: Doctor of Philosophy

Year this Degree Granted: 2003

Brian L. Mark
10815-80th Ave
Edmonton, Alberta
CANADA, T6E 1V9

Date Submitted to the FGSR: Jan 27 /2003

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **X-ray Crystallographic Studies of Human and Bacterial β-Hexosaminidase: Understanding the Molecular Basis of Tay-Sachs and Sandhoff Disease** submitted by Brian L. Mark in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Michael James

Dr. Michael N.G. James

Dr. J.N. Mark Glover

Dr. Howard Young

Monica Palcic

Dr. Monica M. Palcic

Dr. Brian W. Matthews

Date of thesis approval: Jan 21 2003

For Adrienne.

Abstract

Glycoside hydrolases are enzymes that catalyze glycosidic bond hydrolysis. They have been divided into more than 80 sequence-related families. Family 20 includes β-hexosaminidases, enzymes that catalyze the removal of β(1→4) linked *N*-acetyl-β-hexosaminides (βGlcNAc or βGalNAc) from the non-reducing end of oligosaccharides and glycoconjugates. Humans express two major β-hexosaminidase isoenzymes: Hex A and Hex B. Hex A is a heterodimer of α and β subunits (60% identity), whereas Hex B is a homodimer of β subunits. The heritable deficiency of human Hex activity results in various forms of G_{M2}-gangliosidosis, including Tay-Sachs and Sandhoff disease; these are lysosomal storage disorders resulting from the abnormal accumulation of G_{M2}-ganglioside in the brain and peripheral nervous system. It has been proposed that family 20 β-hexosaminidases use a substrate-assisted catalytic mechanism involving neighbouring C2-acetamido group participation and the formation of a cyclic oxazolinium ion intermediate. Insight into this catalytic mechanism was achieved by determining the X-ray crystal structure of a family 20 β-hexosaminidase from *Streptomyces plicatus* (*Sp*HEX) alone and in complex with the mechanism-based inhibitors GalNAc-isofagomine, and NAG-thiazoline. Furthermore, the catalytic role of a conserved aspartate residue (*Sp*HEX Asp313) was investigated by determining wild type *Sp*HEX and two variants thereof (Asp313Ala and Asp313Asn) in complex with the product, *N*-acetyl-D-glucosamine. These structural studies help define the role of Asp313 in the catalytic mechanism of *Sp*HEX, and provide evidence for substrate-assisted catalysis and the formation

of a cyclic intermediate in the mechanism of family 20 β-hexosaminidases. Direct insight into the structural basis of Tay-Sachs and Sandhoff disease was gained by determining the crystal structure of human Hex B, alone and in complex with GalNAc-isofagomine or NAG-thiazoline. From these structures, and the known X-ray structure of the G_{M2}-activator, a protein that solubilizes G_{M2}-ganglioside for presentation to Hex A, a model of Hex A in complex with the activator and ganglioside substrate was built. Together, these crystallographic and modeling data of human Hex demonstrate how α and β subunits dimerize to form either Hex A or Hex B, how these isoenzymes hydrolyze diverse substrates, and how many documented point mutations cause Sandhoff disease (β-subunit mutations) and Tay-Sachs disease (α-subunit mutations).

Acknowledgements

I would like to thank Michael James for his outstanding supervision, guidance and generosity. He has made my education at the University of Alberta a very exciting and challenging experience.

Adrienne, my wife and friend, like always, has been a source of comfort and support to me, and I thank her.

My family, without their support and encouragement, I would not have been able to obtain education I have today. I am forever grateful.

Drs. David Vocadlo and Stephen Withers (University of British Columbia) and Dr. Don Mahuran (University of Toronto) have been outstanding collaborators on many of the projects described herein, and I thank them their participation.

Everyone who has been in Michael's laboratory during my PhD studies has contributed to my education. Without the exchange of ideas and knowledge between coworkers in the laboratory, science would not be the exciting discipline that it is. I thank them.

While living in Edmonton, I made many wonderful friends. I thank them for their all their help and support during my studies, and for making my time in Alberta an enjoyable experience.

Finally, I must thank my uncle Barry Taylor for those many unforgettable ski trips in the great Canadian Rocky Mountains.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Å	Ångstroms (10^{-10} meters)
APS	Advanced Photon Source
aa	amino acid
DTT	dithiothreitol
bp	base pair
|F|	diffracted X-ray structure factor amplitude
Gal	galactose
Glu	glucose
GlcNAc	*N*-acetyl-D-glucosamine
GalNAc	*N*-acetyl-D-galactosamine
G_{M2}-ganglioside	GalNAc-β(1,4)-[*N*-acetylneuraminic acid 2,3)-]-Gal-β(1,4)-Glc-ceramide
HEW lysozyme	hen egg white lysozyme
Hex A	human β-hexosaminidase isoform A (αβ heterodimer)
Hex B	human β-hexosaminidase isoform B (ββ homodimer
Hex S	human β-hexosaminidase isoform S (αα homodimer)
K_{cat}	turnover number (second $^{-1}$)
kDa	kiloDaltons
K_i	inhibition constant (moles/liter)
K_m	Michaelis constant (Molar)
λ	wavelength
MAD	Multiwavelength Anomalous Diffraction
4-MUGS	4-methylumbelliferyl-2-acetamido-2-deoxy-β-glycopyranoside-6-sulphate
4-MUG	4-methylumbelliferyl-2-acetamido-2-deoxy-β-D-glycopyranoside
NAM	*N*-acetylmuramic acid
NMR	Nuclear Magnetic Resonance
ORF	Open Reading Frame

rms	root mean square
σ	standard deviation
Sf9	*Spodoptera frugiperda* cell line 9
Sf21	*Spodoptera frugiperda* cell line 21
*Sp*HEX	Streptomyces plicatus β-hexosaminidase
*Sm*CHB	*Serratia marcescens* chitobiase
SSRL	Stanford Sychrotron Radiation Laboratory
PCR	Polymerase Chain Reaction
URL	Uniform Resource Locator
V_{max}	Maximal initial velocity

Chapter 1
Introduction

A. Glycoside hydrolases

Carbohydrates are involved in a variety of biological functions including cell structural integrity, energy storage, pathogen defense and invasion mechanisms, viral penetration and cellular signaling. Therefore, a large number of enzymes dedicated to carbohydrate processing have evolved. Enzymes specifically responsible for carbohydrate degradation via glycosidic bond hydrolysis are collectively referred to as *O*-glycoside hydrolases (glycosidases) (EC 3.2.1.-). The reaction catalyzed by these enzymes is a nucleophilic substitution reaction that takes place at the carbon atom of the anomeric center, and can result in either retention or inversion of anomeric configuration [1] (Figure 1.1). These alternative stereochemical outcomes require different catalytic mechanisms and accordingly, glycosidases exist in one of two possible mechanistic groups: those that catalyze bond hydrolysis with net retention of anomeric configuration and those that invert the anomeric configuration [1, 2]. When considering these two mechanistic groups in combination with the anomeric configuration of the scissile glycosidic bond (axial (α) or equatorial (β)), four basic mechanistic categories of glycosidases arise: α-retaining, α-inverting, β-retaining and β-inverting [1] (Figure 1.1).

A wealth of structural, functional and sequence information is now available for hundreds of glycosidases, and a useful way of organizing this information has been to classify the enzymes into sequence-related families using standard sequence-alignment algorithms as well as hydrophobic cluster analyses [3-5]. Considering that amino acid sequence dictates three-dimensional structure, it is not surprising to find that although individual sequence-related families may contain glycosidases with distinct substrate specificities, the families tend to be comprised of glycosidases that have highly similar three-dimensional structures. Moreover, in spite of sharing little sequence

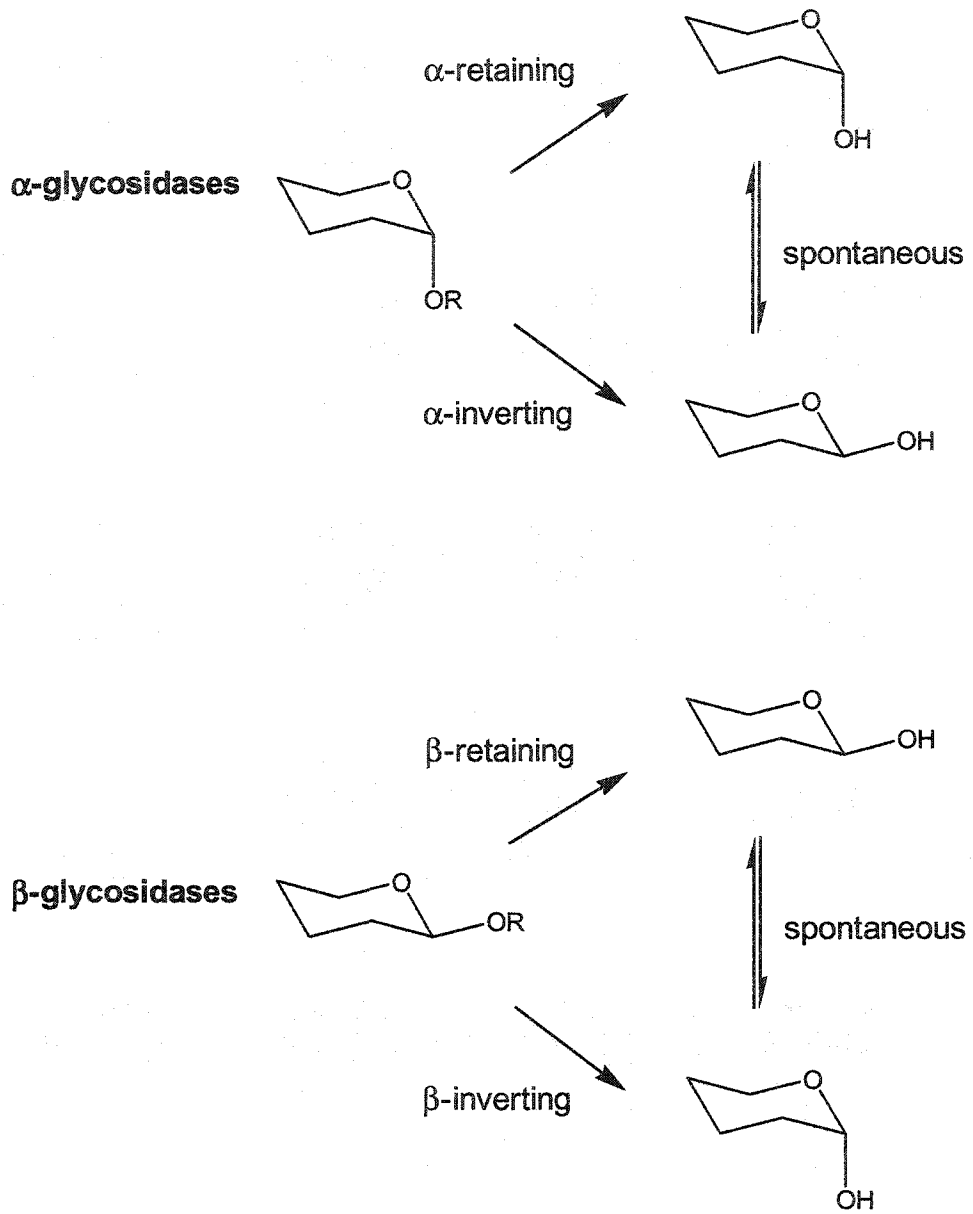


Figure 1.1. Glycosidases are categorized into two mechanistic groups: configuration retaining or configuration inverting glycosidase, and each of the two mechanistic groups can be further subdivided into α– or β-glycosidases depending on the anomeric configuration of the preferred substrate [1]. The reducing end of the product spontaneously changes configuration (mutarotation) so that there is a 2:3 mixture of α and β anomers present at equilibrium.

similarity, many families have been found to have a common protein fold; thus, when structural information is available, glycosidase families can be grouped into clans based solely on structural similarity [6].

More than 84 sequence-related glycosidase families currently exist, and at least 42 families contain members that have had their three-dimensional structures determined. Such efforts in structural biology have revealed that the glycosidases are a structurally diverse group of enzymes; this diversity has necessitated the creation of 13 separate clans. Clan GH-A is the largest, and contains 16 families all sharing a conserved $(\beta/\alpha)_8$ fold. The URL http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html points to a frequently updated database of the sequence-related glycosidase families. All members of an individual sequence-related family catalyze hydrolyitic reactions having the same stereochemical outcome (retaining or inverting) [7, 8]. Thus, the current classification system not only provides insight into the evolutionary relationships among glycosidases, it can be used to derive mechanistic information as well.

B. General β-glycosidase mechanisms

B.1. Inverting mechanism.

The vast majority of glycosidases have two key active site carboxyl groups (Glu or Asp) that are intimately involved in the catalytic mechanisms of both inverting and retaining glycosidases [9, 10]. Inverting glycosidases use a single-displacement mechanism for which the two carboxyl groups are spaced approximately 10.5 Å apart, thereby allowing a substrate molecule and a water molecule to bind between them [1, 2, 10]. One of the carboxyl groups promotes general-acid catalyzed cleavage of the glycosidic bond, whereas the other acts as a general base, catalyzing the nucleophilic attack of a water at the anomeric center in a single nucleophilic reaction that occurs via an oxocarbenium ion-like transition state [2, 9, 11, 12]. The result of this reaction is a hemiacetal product having inverted anomeric stereochemistry from the substrate (Figure 1.2a).

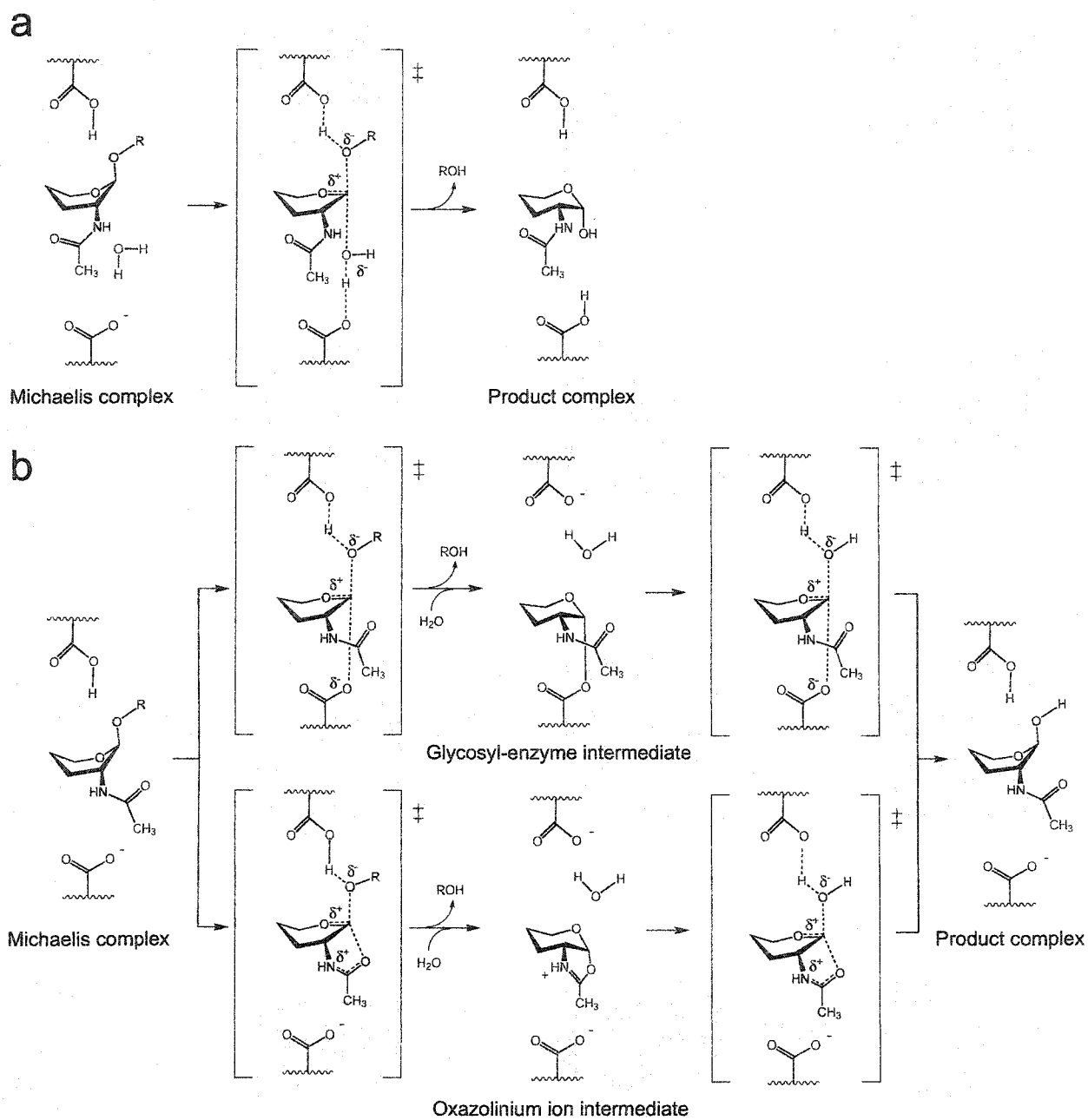


Figure 1.2. Proposed catalytic mechanisms for inverting (a) and retaining (b) glycosidases. Hydroxyl groups and C6 have been removed from the sugar ring for clarity and, ignoring the C2-acetamido group, the mechanism shown in panels a and b, *upper pathway* can be used to describe the catalytic mechanisms of inverting and retaining glycosidases in general. No attempt has been made to indicate the true positions of the enzymic residues, and the carboxylate illustrated beneath the cyclized intermediate of panel b, *lower pathway*, is not positioned to act as a nucleophile. Instead, this carboxylate is positioned to stabilize the positive charge that develops on nitrogen (N2) of the C2-acetamido group upon cyclization.

B.2. Retaining mechanism using an enzymic nucleophile.

Retaining glycosidases use a double-displacement mechanism. The majority of retaining glycosidases contain two enzymic carboxyl groups, spaced approximately 5.5 Å apart, such that only the substrate molecule can fit between them [1, 2, 10]. One carboxyl group promotes general-acid catalyzed cleavage of the glycosidic bond, whereas the other acts as a nucleophile, directly attacking the anomeric center to form a covalent glycosyl-enzyme intermediate. After departure of the leaving group, the intermediate, which has an inverted anomeric center relative to the substrate, undergoes general base-catalyzed nucleophilic attack by an incoming water molecule to yield a hemiacetal product having retained anomeric stereochemistry (Figure 1.2b, *upper pathway*) [9]. Both steps of the double displacement mechanism proceed through transition states having substantial oxocarbenium ion character [2, 11, 12].

B.3. Retaining mechanism using a substrate nucleophile.

It has been well documented in physical organic chemistry that a chemical reaction can be accelerated by the intramolecular participation of a functional group on the reactant [13]. The functional group acts as an intramolecular catalyst and provides anchimeric assistance to the catalytic mechanism of the reaction. Having a component of the catalytic machinery attached to the reactant such that it is not stereochemically nor structurally hindered from accessing the reactive center, not only increases the effective concentration of the catalyst at the site of chemical reactivity, but also it provides an entropic advantage by reducing at least one aspect of the reaction mechanism from what would typically be a bimolecular to a unimolecular event (reviewed in [14]).

Neighbouring C2-acetamido group participation has been documented to provide anchimeric assistance in the acid-catalyzed hydrolysis of *N*-acetylglucosaminides in solution [15], and a number of glycosidases have evolved to incorporate C2-acetamido group participation into their catalytic mechanisms. Enzymes that hydrolyze *N*-acetylhexosaminide containing substrates (GlcNAc or GalNAc) are generally referred to as hexosaminidases.

The configuration-retaining β-hexosaminidases from glycosidase families 18, 20 and 56 lack an apparent enzymic nucleophile needed to form the glycosyl-enzyme intermediate in the first step of the double-displacement mechanism shown in Figure 1.2b, *upper pathway* [16]. Instead, the C2-acetamido group of the substrate participates in the reaction and assumes the role of the missing nucleophile (Figure 1.2b, *lower pathway*) [9, 17-19].

Glycosidases from families 18, 20 and 56 belong to clan GH-K [7] and share a conserved $(\beta\alpha)_8$ catalytic domain (Figure 1.3). A number of equivalent catalytic residues within the active sites of these enzymes result in common structural features that distort the bound substrate so that the carbonyl oxygen atom of the C2-acetamido group can become appropriately positioned for nucleophilic attack at the anomeric center [18-20] (Figure 1.4). In keeping with the geometrical considerations of the general retaining mechanism, this substrate distortion creates an average distance of 5.4 Å between the C2-acetamido group oxygen atom of the substrate and the carboxyl oxygens of the enzymic general acid/base catalytic residue. In the first step of this 'substrate-assisted' mechanism, the enzymic carboxyl group promotes general acid catalyzed cleavage of the glycosidic bond, while a concerted nucleophilic attack by the C2-acetamido group oxygen results in an enzyme stabilized, oxazolinium ion intermediate [17, 20, 21]. After departure of the leaving group, the cyclized intermediate undergoes base-catalyzed nucleophilic attack by an incoming water molecule at the anomeric center to yield a hemiacetal product with retained stereochemistry. As for the 'normal' double displacement mechanism, both steps of the reaction are believed to pass through oxocarbenium ion-like transition states [20] (Figure 1.2b, *lower pathway*).

C. The G_{M2}-gangliosidoses: Tay-Sachs and Sandhoff disease

The G_{M2}-gangliosidoses are a group of three human autosomal recessive, lysomal storage disorders caused by the heritable deficiency of β-hexosaminidase activity. The lack of hexosaminidase activity results in a pathological accumulation of G_{M2}-ganglioside in the brain and peripheral nervous

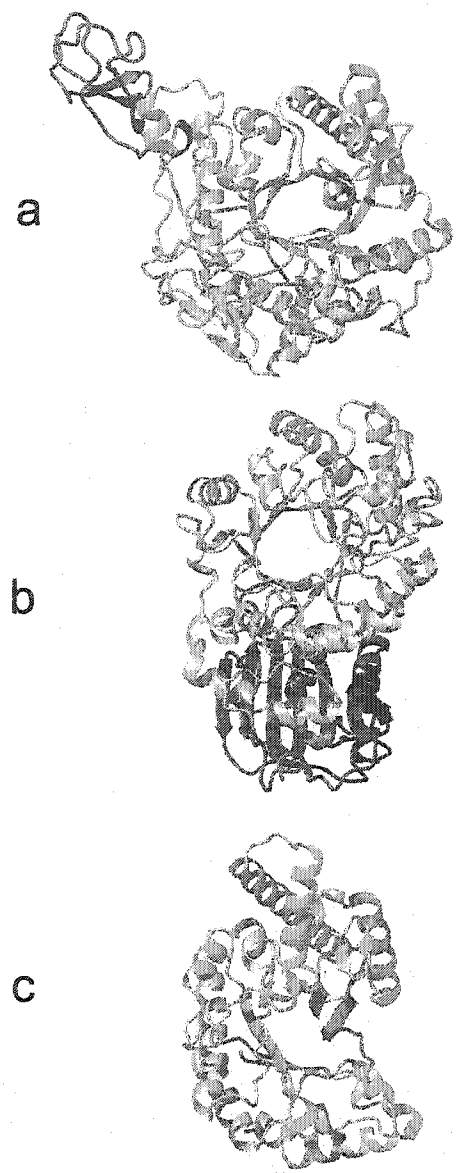


Figure 1.3. Ribbon diagrams of representative three-dimensional glycosidase structures from the sequence related families 18 (a), 20 (b) and 56 (c). Although there is little sequence identity shared amoung the three families, they all have conserved $(\beta/\alpha)_8$ barrel catalytic domains and belong to clan GH-K (extra domains not part of the barrel are shown in dark blue). a, Family 18 chitinase B from *Serratia marcescens* [107]. b, Family 20 β-hexosaminidase from *Streptomyces plicatus* [16]. c, Family 56 hyaluronidase from bee venom [19].

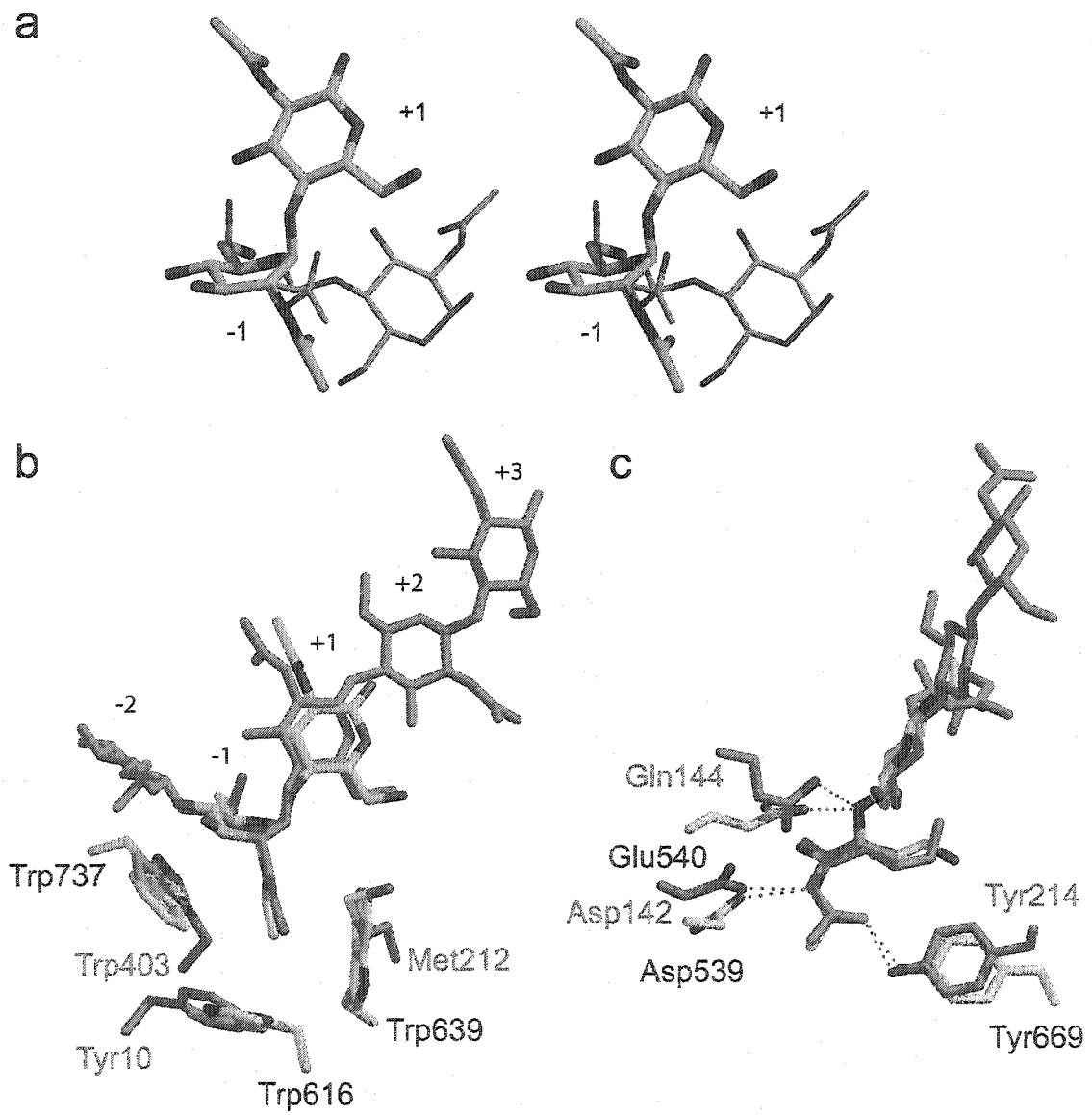


Figure 1.4. Crystallographic evidence for substrate distortion in the Michaelis complexes of family 18 and 20 glycosidases. a, Stereographic representation of a superposition of the small molecule crystal structure of unbound chitobiose (green carbon atoms) [123] and chitobiose bound to a family 20 chitobiase from *Serratia marcescens* as a Michaelis complex. As previously shown by Tews *et al.* [17], this superposition demonstrates that the non-reducing sugar of chitobiose (bound in the -1 sugar binding subsite) becomes distorted into a boat conformation, and the scissile bond and leaving group (bound in the +1 subsite) adopt a pseudo-axial position. This allows the carbonyl oxygen of the C2-acetamido group to rotate beneath the anomeric carbon into position for nucleophilic attack. b, Superposition of a crystallographic Michaelis complex of a family 18 *endo*-chitinase from *S. marcescens* (green) [107] and the family 20 chitobiase Michaelis complex from *panel a*. This overlay demonstrates that family 18 chitinases, which are *endo* acting and thus cleave bonds randomly within the polymer, distort the substrate in a manner similar to the *exo* acting glycosidases of family 20. c, Same superposition as in *panel b*, but rotated 90 degrees and showing polar as opposed to hydrophobic residues of the active site.

system [22]. The clinical phenotype of the most severe form of G_{M2}-gangliosidosis was first described in a publication by Warren Tay in 1881 [23] and followed shortly thereafter by a description of the morphological features of the disease in 1887 by Bernhard Sachs [24]. The most notable pathology observed by Sachs was the distended cytoplasm and enlarged dendrites of the neurons, and those afflicted with the disease typically died within their fourth or fifth year of life. Sachs named the disorder familial amaurotic idiocy [25] (it quickly became known as infantile Tay-Sachs disease [26]), and incredibly, 82 years passed before the link between G_{M2}-gangliosidosis and β-hexosaminidase activity was finally discovered in 1969 by Okada and O'Brien [27]. Since this crucial discovery, a great deal of research has been carried out on characterizing the biochemical function, and now the molecular structure, of human β-hexosaminidase. A large body of research exists and it has made human β-hexosaminidase the primary model for lysosomal glycosidases. This section describes briefly the biochemistry of human β-hexosaminidase, followed with a description of the G_{M2}-gangliosidoses, their classification and potential methods of treatment.

C.1. Human lysosomal β-hexosaminidase

C.1.1. Genes and isoenzyme products.

Humans express three configuration-retaining β-hexosaminidase isoenzymes (Hex) (EC 3.2.1.52) [22], and they belong to glycosidase family 20. The Hex isoenzymes remove terminal β-(1,4) linked *N*-acetyl-β-hexosaminides (βGlcNAc or βGalNAc) from the non-reducing end of glycoproteins, glycolipids, oligosaccharides and glycosaminoglycans (see review [8]). Two of the three isoenzymes, Hex A and Hex B, dominate the isoenzyme pool and are expressed in approximately equal amounts [28], while the third isoform, Hex S [29], is less abundant. All three isoforms are found primarily in the lysosome; Hex A is a heterodimer of subunits α (encoded by *HEXA*) and β (encoded by *HEXB*); Hex B is a homodimer of β-subunits, whereas Hex S is a homodimer of α-subunits (see

review [22]). *HEXA* is a 35 kb gene located on chomosome 15 [30-33]; *HEXB* is 45 kb and is located on chomosome 5 [34, 35]. There exists an approximate 60% sequence identity at both the nucleotide and amino acid level between *HEXA* and *HEXB*, and both genes contain coding regions of approximately 1600 bp [35]. Each gene is divided into 14 exons, with exons 2-14 having splice junctions at similar positions within the coding sequence [35]. The similarities in gene structure and sequence suggest that *HEXA* and *HEXB* arose from a common ancestral gene.

C.1.2. Isoenzyme biosynthesis and posttranslational processing.

Both the α- and β-subunit gene products have a complex biosynthetic pathway (Figure 1.5). The mRNA transcripts from *HEXA* and *HEXB* each code for an endoplasmic reticulum (ER) signal peptide that targets translation to the ER lumen [36, 37]. During translation, a 23 aa ER signal peptide is proteolytically removed from the newly synthesized α-polypeptide, and a 43 aa ER signal peptide is removed from the new β-polypeptide [38, 39]. These cleavages result in soluble pro-α- and pro-β-subunits of 506 aa and 513 aa, respectively. While traversing the ER, the pro-subunits become *N*-glycosylated [40, 41] at selected Asn-X-Ser/Thr triplets [42] and fold to a near native conformation so that they may dimerize with one another to form either semi-processed Hex A (αβ), Hex B (ββ) or Hex S (αα) isoenzymes [43]. Only the dimeric species of the isoenzymes can pass into the *Golgi* network for further processing [43]. Like other lysosomal enzymes, at least one *N*-linked high mannose oligosaccharide on each subunit of the semi-processed isoenyzmes will acquire a mannose-6-phosphate marker [44, 45] so that the pro-isoenzymes can be recognized by a mannose-6-phosphate receptor in the *trans* Golgi for redirection to the lysosome [46]. Correctly folded pro-α- and pro-β-subunits are required for lysosomal targeting because phosphorylation is believed to require a three-dimensional structural motif for recognition by UDP-*N*-acetylglucosamine-1-phosphotransferase [47], an enzyme that catalyzes the transfer of GlcNAc phosphate onto the high-mannose residues of select glycoproteins (reviewed in [48]).

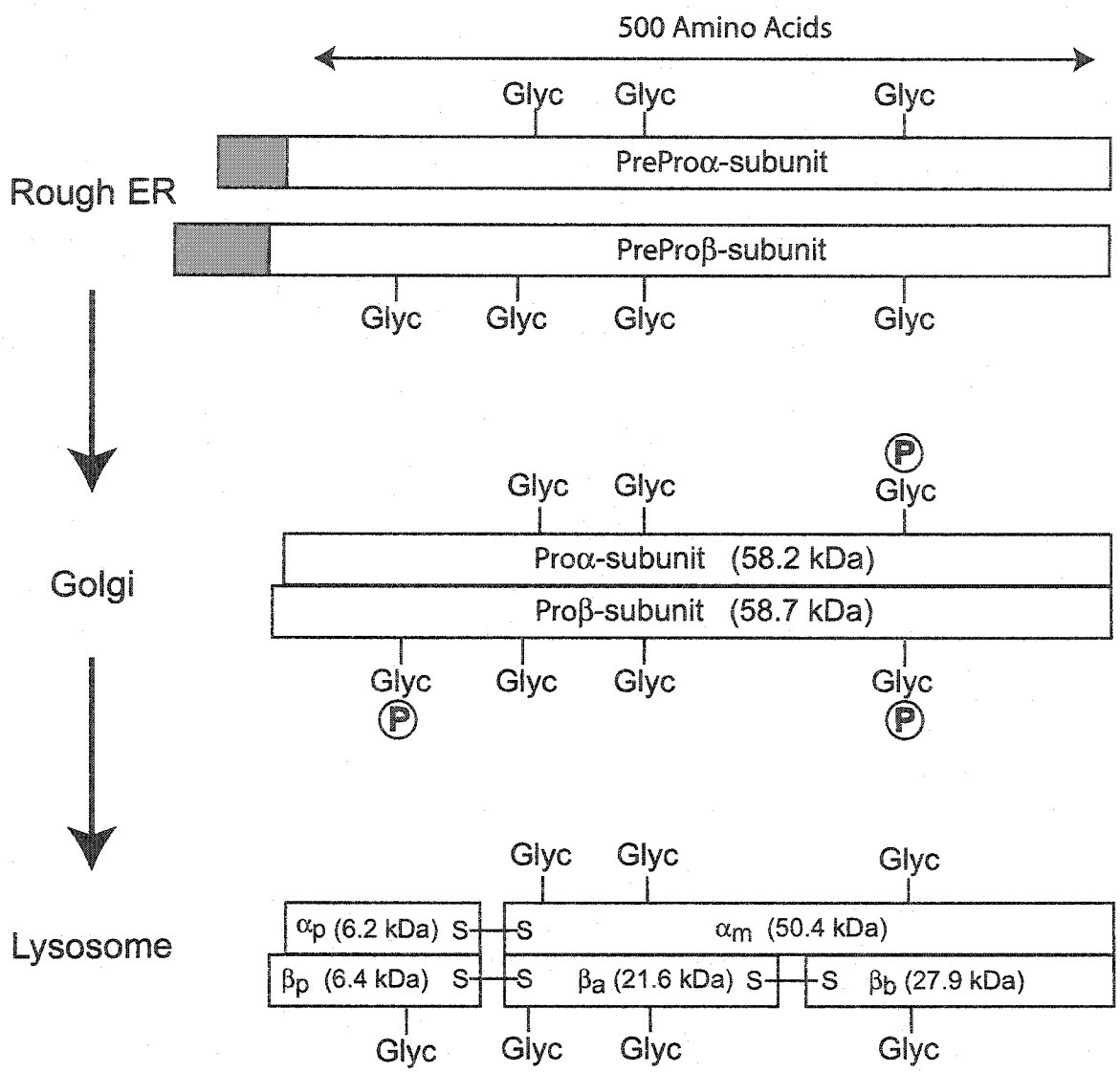


Figure 1.5. Posttranslational processing of Human Hexosaminidase. Both the α– and β–subunits are translated into the ER lumen after which the ER signal peptide (orange) is removed. While traversing the ER both subunits fold to near native conformations and become glycosylated (Glyc). Dimerization occurs in the ER to form either Hex A (αβ) (as shown above) or Hex B (ββ). Hex A and B then pass into the Golgi for further processing and become phosphorylated (mannose-6-phosphate) for redirection to the lysosome. Once in the lysosome, both subunits have further modifications done to their glycans and undergo limited proteolysis to form the mature Hex isoforms. Figure adapted from [43].

Upon entering the late endosome, or the lysosome itself, further glycosidic processing and specific proteolytic cleavages on the α- and β-subunits of each isoenzyme give rise to the mature proteins (reviewed in [43]). Limited proteolysis of the Hex B homodimer results in the removal of two surface loops from each pro-β-subunit, giving rise to a mature Hex B isoform comprised of two β-subunits each cut into three distinct polypeptide chains: β_p (residues 50-107), β_b (residues 122-311) and β_a (residues 316-556) [37, 39, 49, 50]. Limited proteolysis of the Hex A heterodimer (αβ) gives rise to a mature β-subunit as described above, and an α-subunit from which only one loop is removed, to yield a mature α-subunit composed of two polypeptide chains: α_p (residues 23-107) and α_m (residues 122-529) [38]. Interestingly, the proteinase cleavage sites are unusual, and the lysosomal proteinases responsible for catalyzing the removal of these loops from Hex isoenzymes have not been identified. Hex S is comprised of two of these mature α-subunits. The polypeptides of each subunit type are covalently joined together by disulphide bonds. Both the α- and β-subunit contain a functional active site; however, dimerization is required for the sites to become active, and dimer-stability is not equal among the isoforms (ββ >αβ>αα) (reviewed in [51]).

C.1.3. Human Hex substrate specificity.

Hex A is essential for degrading G_{M2}-ganglioside, a glycosphingolipid generated primarily in the brain and peripheral neurons [27] [28] (Figure 1.6). The enzyme initiates the degradation of G_{M2}-ganglioside by removing a terminal β(1→4) linked GalNAc residue from the ganglioside via the substrate-assisted catalytic mechanism described above (Figure 1.2b, *lower pathway*) [17]. The product, G_{M3}-ganglioside, is then further processed by the remaining lysosomal enzymes of the degratory pathway [52]. Although both the α- and β-subunit can remove *N*-acetyl-β-hexosaminides from neutral glycoproteins and oligosaccharides, only the α-subunit active site can efficiently accommodate negatively-charged substrates such as G_{M2}-ganglioside [53, 54] and GlcNAc-6-SO_4^- containing glycosaminoglycans such as keratan sulfate [55] and substrate analogues [56] (Figure 1.6).

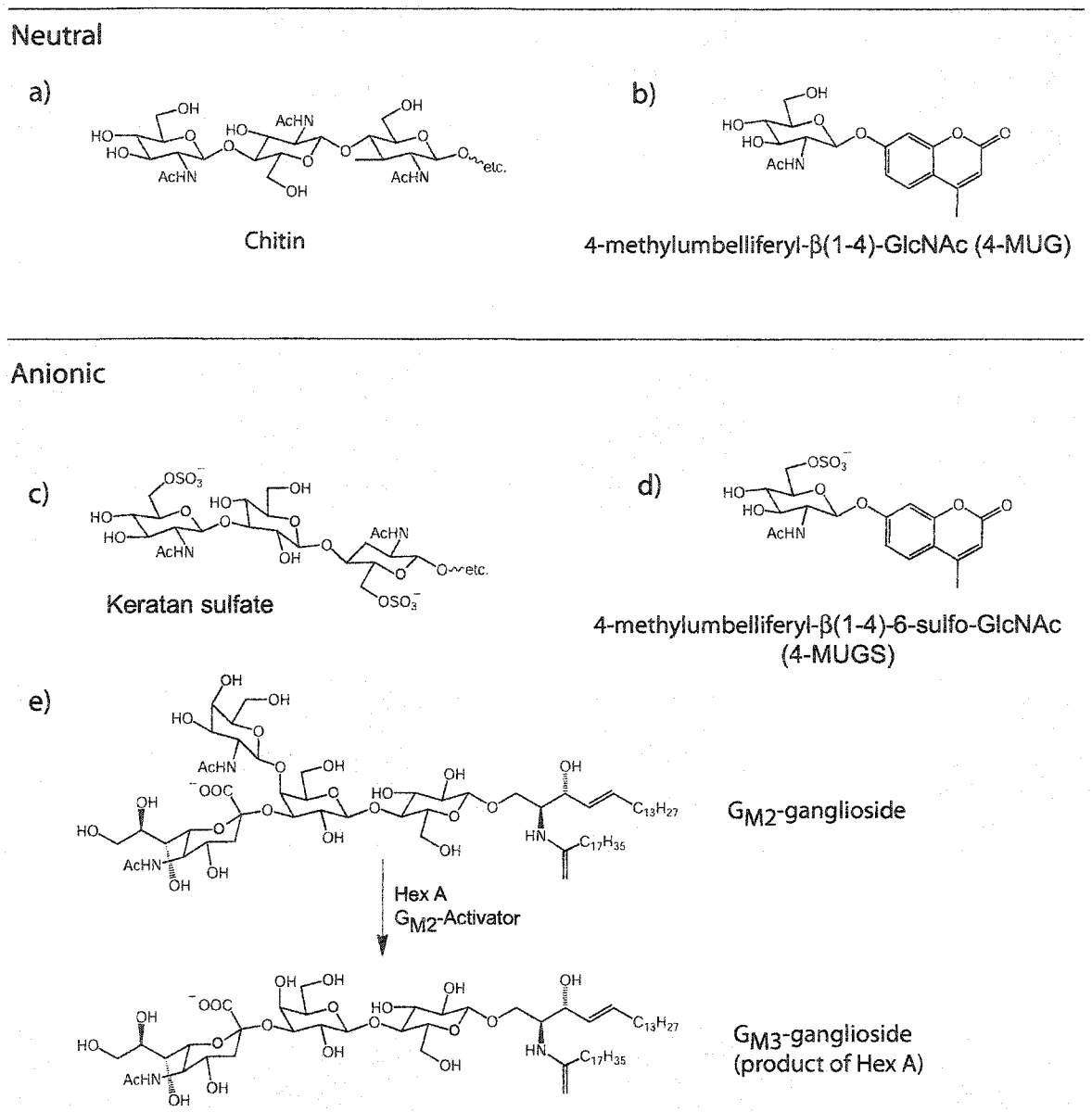


Figure 1.6. Natural (a,c,e) and artificial substrates (b,d) for human hexosaminidase. Substrates a & b are neutral and can be hydrolyzed by Hex A or B; whereas, only Hex A can hydrolyze substrates c, d, & e because of the negatively charged functional groups (SO_4^- or COO^-) these substrates contain. For substrate (c), the hydrolyzed bond is β(1-3).

The neutral substrate analogue 4-methylumbelliferyl-β-N-acetylglucosaminide (4-MUG) [57] is commonly used to measure β-hexosaminidase activity [58, 59]. Both Hex A and B have a pH optimum of 4.4; they hydrolyze 4-MUG with similar K_m values (~0.9 mM) and V_{max} values of 1.8 x 10^{-4} mol/min/mg and 4.4 x 10^{-4} mol/min/mg, respectively [60]. Conveniently, Hex A is heat labile whereas Hex B is not; thus, the activities of the two isoenzymes can be quickly differentiated by introducing a heat denaturation step during the analysis of a serum sample for Hex activity [61]. Direct analysis of Hex A activity is achieved by using the negatively charged substrate analogue 4-methylumbelliferyl-6-sulfo-β-N-acetylglucosaminide (4-MUGS) [56], which is preferentially hydrolyzed by Hex A with a K_m of 0.31 mM and V_{max} of 25 x 10^{-4} mol/min/mg as opposed to Hex B which hydrolyses this analogue much more slowly (K_m of 3.4 mM with a V_{max} of 1.2 x 10^{-6} mol/min/mg) [60].

C.1.4. The G_{M2}-activator protein.

In order for Hex A to degrade G_{M2}-ganglioside, it requires the assistance of an additional lysosomal protein: the G_{M2}-activator protein (activator) [62, 63]. The activator is encoded by the gene *GM2A* [64] and is located on chromosome 5 [65]. The activator carries out two important functions: 1) it extracts single G_{M2}-ganglioside molecules from the intralysosomal membrane or vesicle to form a 1:1 water soluble complex and 2) it docks specifically onto Hex A and presents the terminal GalNAc residue of G_{M2}-ganglioside to the α-subunit active site for removal [66]. Interestingly, the activator does not require the presence of G_{M2}-ganglioside in order to form a complex with Hex A, and the activator binding site on Hex A appears to be a combination of structural elements from both the α- and β-subunit [67-69]. Thus, even though Hex S is composed of two α-subunits and contains two functional active sites that can catalyze the removal of negatively charged residues from substrates such as keratin sulphate and G_{M2}-ganglioside, Hex S cannot interact sufficiently well with the activator protein to remove GalNAc from G_{M2}-ganglioside [67].

C.2. Classification of the G_{M2}-gangliosidoses.

The physiological importance of the Hex A / Activator protein machinery is illustrated by the fatal neurodegenerative disorders that result from the inheritable deficiency of these proteins (reviewed in [22]). Type 1 G_{M2}-gangliosidosis or variant B (Tay-Sachs disease) results from inborn errors of metabolism in *HEXA* (reviewed in [70]) causing an α-subunit defect or deficiency leading to a specific lack of Hex A activity, whereas the lysosomal Hex B activity is normal [27, 28, 71]. The degree of residual Hex A activity appears to be inversely proportional to the clinical severity of this disease [72]. Mutations resulting in little or no Hex A activity cause the rapid accumulation of G_{M2}-ganglioside in the brain and peripheral nervous system. Neurodegeneration begins during infancy, giving rise to infantile onset G_{M2}-gangliosisdosis or classic Tay-Sachs disease. Mutations having a less severe effect on Hex A catalytic activity correspondingly give rise to less severe disease phenotypes such as juvenile onset G_{M2}-gangliosidosis, adult onset or chronic G_{M2}-ganglosidosis, and Hex A pseudodeficiency (reviewed in [22]). The residual level of Hex A activity in patients with 'late-onset' forms of the disease appears to be sufficiently high so that years may pass before enough G_{M2}-ganglioside accumulates to cause detectable neurological symptoms.

Type 2 G_{M2}-gangliosidosis or variant O (Sandhoff disease) results from mutations in *HEXB* (reviewed in [70]) that cause a β-subunit defect or a deficiency that leads to a loss of both Hex A and Hex B activity. The phenotype of this disease is similar to infantile Tay-Sachs disease. The variant AB form of G_{M2}-gangliosidosis is caused by mutations in the *GM2A* gene resulting in a G_{M2}-activator deficiency but in the presence of normal levels of Hex A and Hex B [62].

Cases of Tay-Sachs and Sandhoff disease have been found in a diverse range of ethnic groups throughout the world, whereas the AB variant form of G_{M2}-gangliosidosis is relatively more rare. These are autosomal recessive disorders and the highest incidence of the disease (1/3600) occurs in the Ashkenazi Jewish population, with a carrier rate of approximately 1 in 35. The carrier rate in the general population is about 1 in 300 [22, 73]. Fortunately, the frequency of Tay-Sachs and Sandhoff disease has been reduced by more than 90% since the

introduction of enzyme based screening programs in the late 1970's [22, 74]. To date, there are 100 known mutations in *HEXA*, 25 known mutations in *HEXB* and 5 in *GM2A* that have been reported to cause G_{M2}-gangliosidosis. The URL http://data.mch.mcgill.ca/gm2-gangliosidoses/ points to a database of known *HEXA*, *HEXB* and *GM2A* mutations. These include missense mutations that cause amino acid substitutions, nonsense mutations causing deletions and insertions resulting in frame-shifts or chain terminations, and splice-site mutations that yield incorrectly processed mRNA [70]. Much biomedical research has been done in an attempt to explain the observed phenotypes based on the biochemistry of these known mutations [75].

C.3. Potential therapeutic approaches for G_{M2}-gangliosidosis.

Unfortunately, like many genetic diseases that result in loss-of-function, treatment protocols for the G_{M2}-gangliosidoses are scarce and remain experimental. Prevention of the disease through carrier screening is the only effective means available to manage these disorders. Enzyme replacement therapies through direct injection or bone marrow transplantation have been largely unsuccessful, and the promise of gene therapy methods to target wild type *HEXA* or *HEXB* to neuronal cells remains only a distant possibility [76].

However, research into a substrate deprivation approach appears to hold promise for the treatment for those patients having residual Hex activity and suffering from late-onset forms of G_{M2}-gangliosidosis. By specifically inhibiting the glycosyltransferase-catalyzed biosynthesis of glucosylceramide, a precursor of G_{M2}-ganglioside, it is argued that the rate of G_{M2}-ganglioside accumulation in Hex A compromised patients might be slowed significantly, so that pathological accumulations of the ganglioside could be avoided completely [77]. To test this hypothesis the glycosyltransferase inhibitor *N*-butyl-deoxynorjirimycin (Figure 1.7) was used to treat a mouse model of Tay-Sachs disease (*HEXA* gene knock-out [78]) [77]. The Tay-Sachs mouse best emulates late-onset forms of

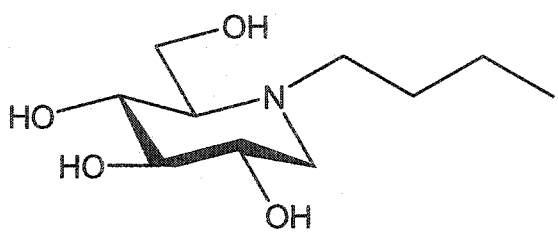


Figure 1.7. Schematic of *N*-butyl-deoxynojirimycin [77]
(*N*-(n-butal)-1,5-dideoxy-1,5-imino-D-glucitol)

G_{M2}-gangliosidosis because unlike humans, mice express a sialidase that can convert a significant proportion of the accumulating G_{M2}-ganglioside into G_{A2}-ganglioside (sialic acid group removed), that can then be converted to G_{M3}-ganglioside by Hex B and cleared from the lysosome [79]. Although Tay-Sachs mice accumulate a significant amount of G_{M2}-ganglioside relative to wild type mice, ganglioside stores never reach the level where neurological symptoms develop [78].

Treatment of Tay-Sachs mice with *N*-butal-deoxynorjirimycin is well tolerated [80]. It has been shown to cross the blood-brain barrier in quantities sufficient to result in an astonishing 50% reduction in G_{M2}-ganglisoside content in the brains of treated mice (relative to untreated controls) [77]. Thus, this small molecule therapy appears to prevent ganglioside storage by establishing an improved physiological equilibrium between G_{M2}-ganglioside synthesis and degradation in the absence of normal Hex A activity. *N*-butyl-deoxynorjirimycin is a polar molecule. It was estimated that only a small percentage of the compound actually passes thorough the blood-brain barrier; thus, the efficacy of the drug may be improved dramatically though structure activity relationship studies carried out to increase the drug's lipophilicity.

Recently, *N*-butal-deoxynorjirimycin treatment in combination with enzyme replacement therapy via bone marrow transplant (BMT) was carried out on a mouse model of Sandhoff disease (*HEXA* and *HEXB* knock-out [79]) to determine if such an approach could be used to treat infantile onset disease variants [81]. Sandhoff mice, like humans with classical infantile Tay-Sachs disease, display severe, progressive neurological degeneration starting about 3 months post partum [79]. *N*-butal-deoxynorjirimycin treatment alone increases the survival rate of Sandhoff mice by 40% [82], and BMT therapy extends life expectancy by up to 8 months [83]. When given in combination, the effect on survival rate was 13-25% more than additive, suggesting a synergistic enhancement of the individual treatments when given in combination [81]. However, the maximum level of brain β-hexosaminidase activity provided by BMT therapy was found only to be 2% to 5% of that of wild type. It was not

sufficient to establish an equilibrium between G_{M2}-ganglioside synthesis and degradation that was compatible with life, even in the presence of *N*-butyl-deoxynorjirimycin. Nonetheless, the synergistic effect observed during the combination therapy trials was encouraging. The authors suggest that if high enough levels of brain β-hexosaminidase activity can be reconstituted through BMT, combination therapy would most likely be the treatment of choice for infantile onset G_{M2}-gangliosidosis [81].

D. Understanding the structural basis of G_{M2}-gangliosidosis

Unfortunately, structural and functional studies on human Hex have been limited by two major factors: 1) difficulty in expressing sufficient amounts of recombinant Hex isoenzyme in mammalian expression systems required for detailed and accurate kinetic analysis, 2) very low residual activity of recombinant mutant Hex isoenzyme [84, 85]. Transient expression of wild type and mutant α-subunit cDNA in monkey COS-7 cells is the most popular means by which to obtain recombinant Hex A for characterizing the biochemical phenotype of natural and engineered mutations; however, COS-7 cells express a significant amount of endogenous Hex A and B which tends to dominate the low level expression of the recombinant Hex. Recently, this problem has been overcome by expressing an α-subunit containing a C-terminal poly-histidine affinity tag [86]. Using this affinity tag, adequate amounts of recombinant Hex A could be purified away from endogenous protein for accurate kinetic studies to be carried out; nonetheless, the amount of recovered protein is still too little to begin crystallization trials.

An additional problem with the COS-7 cell expression system is the large amount of Hex S isoenzyme that is produced over Hex A. In order to drive the isoenzyme equilibrium towards recombinant Hex A, human fetal Tay-Sachs disease neuroglial cells, which lack endogenous α-subunit expression, have been transfected with recombinant α-subunit cDNA [87]. Because these cells do not express endogenous α-subunit, all Hex A produced in these cells will contain a recombinant α-subunit; however, as for the COS-7 cells, the level of Hex A

depends on the amount of endogenous β-subunit expression and this results in low yields. Dependence on the amount of endogenous subunit expression has been overcome by a dual transfection of expression vectors, one expressing the α-subunit, and the other expressing the β-subunit [88]. This approach to overcome low expression levels in combination with an affinity tag system should substantially increase the yield of pure recombinant Hex from mammalian expression systems in the future.

The expense and labour of establishing and operating a large-scale mammalian expression system make it an economically unfeasible method for producing the quantities of recombinant protein required for X-ray crystallographic studies. Unfortunately, *Saccharomyces cerevisiae*, a yeast species routinely used to express large quantities of recombinant eukaryotic protein, fails to express functional human Hex [89]. The Hex B isoenzyme however, has been successfully expressed in the *Spodoptera frugiperda* insect cell lines Sf9 [90] and Sf21 [91]. Sf9 or Sf21 cells transfected with recombinant baculovirus containing β-subunit cDNA express active Hex B, and for the Sf9 cells at least, it has been claimed that the yields of recombinant protein from this system are sufficient for X-ray crystallographic studies [90]. Crystals of recombinant Hex B grown with protein isolated from Sf9 cells have not been reported in the literature.

Two potential problems with these insect cell expression systems are the unusual glycosylation pattern observed on the recombinant Hex B isolated from these cells (as compared to Hex B from mammalian expression systems), and the lack of proteolytic processing that occurs in the mammalian lysosome (see section ***C.1.2.***) [91]. Otherwise, Hex B isolated from these cell lines appears to be correctly folded, and the disulfide bonding pattern of Hex B isolated from Sf21 cells has recently been demonstrated to be identical to that of native Hex B purified from human placenta [91]. Hence, insect cell expression systems may be potentially very useful for the future crystallographic studies of disease causing human Hex variants that cannot be isolated from a natural source. Currently, human placenta is the only source of Hex that yields sufficient quantities of the

isoenzymes for X-ray crystallographic studies [92].

D.1. Identification of human Hex catalytic residues through natural mutations.

Prior to any information about the three-dimensional structure of family 20 glycosidases, the only data available regarding catalytically important residues in human Hex came from studies carried out on a unique subset of naturally occurring *HEXA* point mutations known as variant B1 mutations (reviewed in [22]). For most Tay-Sachs disease cases, *HEXA* mutations are so detrimental to the correct synthesis of the α-subunit that dimerization with the β-subunit rarely occurs and the mutated subunit is rapidly degraded in the ER. However, variant B1 mutations allow Hex A to reach maturity even though the isoenzyme is completely inactive toward the negatively-charged substrates 4-MUGS or G_{M2}-ganglioside. But, the variant B1 isoenzyme is active toward 4-MUG, indicating the presence of an active β-subunit. Thus, it was suggested that variant B1 mutations were affecting catalytic residues located within or very near the α-subunit active site, and do not affect subunit dimerization or trafficking to the lysosome.

The most common and well-characterized variant B1 mutations occur at codon α178: αArg178His [93-96], αArg178Cys [96] and αArg178Leu [97]. Each mutation causes infantile Tay-Sachs disease; however, substitution for a hydrophobic residue at this position (αArg178Leu) causes the greatest loss of Hex A activity of the three [97]. A notable experiment carried out to substantiate the variant B1 hypothesis involved introducing the variant B1 mutation αArg178His into the equivalent codon of the β-subunit (β211) [98]. Indeed, the βArg211His (made *in vitro*) resulted in mature lysosomal Hex B that was devoid of catalytic activity. The more conservative *in vitro* mutation βArg211Lys produced a Hex B variant with a K_m elevated 10-fold and a k_{cat} reduced 500-fold over wild type [86]. Thus, it appeared that the presence of a polar guanidinium group at this position was required for optimal catalytic activity of both the α- and β-subunit. This Arg residue is conserved in all family 20 glycosidases, and

crystallographic studies on bacterial family 20 hexosaminidase and human Hex B (chapter 5) now demonstrate clearly that this conserved Arg residue is intimately involved in substrate binding.

The only other naturally occurring variant B1 mutations that have been found are αAsp258His [99], αSer210Phe [100] and αTrp420Cys [101]. A recombinant Hex A variant containing the mutation αAsp258His was found to be catalytically active, and kinetic measurements demonstrated that although there was a decrease in catalytic rate and a shift in the pH optimum, the K_m was unaffected relative to wild-type Hex A [99]. From these kinetic data, it was postulated that Asp258 might be the general acid/base catalytic residue of the α-subunit; however, work presented here and elsewhere [17] clearly shows that this suggestion is incorrect. Reasons for the adverse affect on catalytic function of the later two variant B1 mutations could not be explained in the absence of structural data.

D.2. Identification of human Hex catalytic residues through affinity labeling.

In an attempt to identify catalytic residues directly in human Hex, Liessem *et al.* [102], performed a photoaffinity labeling experiment using a radio labeled substrate analogue composed of a GalNAc glycone thio-linked to an aglycone containing a diazirine functionality. The diazirine group undergoes photolysis at 350 nm to yield a highly reactive carbene that can react covalently with nearby nucleophilic amino acid side chains (Figure 1.8). Irradiation (350 nm) of Hex B in the presence of the analogue resulted in a 15% loss of enzymatic activity. The loss in activity could be reduced by 75% when the same experiment was performed in the presence of the β-hexosaminidase inhibitor 2-acetamido-2-deoxyD-glucono-1,5-lactone, indicating that the photoaffinity label was indeed reacting at the enzyme active site. Trypsinisation of the photoaffinity labeled enzyme yielded a single radiolabelled trypsin fragment to which the label was found to be associated with Glu355. Multiple sequence alignments revealed Glu355 to be 100% conserved between the α- and β-subunits of human and

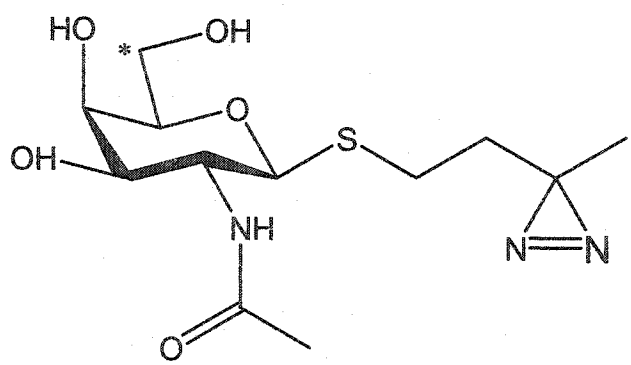


Figure 1.8. Schematic of [^{3}H]-1-ATB-GalNAc (3-azi-1-[(6-^{3}H]-2-acetamido-2-deoxy-1-β-D-galactopyranosyl)thio]-butane) [102]. The position of the tritium label [^{3}H] is indicated by the asterisk.

mouse hexosaminidase and a hexosaminidase from *Dictyostelium discoideum*.

Subsequent structural studies of related bacterial family 20 glycosidases described below have provided convincing evidence that Glu355 is the general acid/base residue in the catalytic mechanism of Hex B [17, 20]. Indeed, the determination of the three-dimensional structure of human Hex B by X-ray crystallography, which comprises part of the work presented in this thesis (chapter 5), demonstrates definitively that Glu355 is the general acid/base residue involved in the catalytic mechanism shown in Figure 1.2b, *lower pathway*.

D.3. Crystallographic studies of bacterial family 20 glycosidases.

X-ray crystallographic studies of bacterial family 20 glycosidases have advanced greatly our understanding of the structural and functional characteristics of the family 20 enzymes (reviewed in [124]). The first family 20 enzyme to have its three-dimensional structure determined was the 818 aa chitobiase from *Serratia marcescens* (*Sm*CHB) [17]. This was followed shortly thereafter by the determination of the crystal structure of the 506 aa family 20 β-hexosaminidase from *Streptomyces plicatus* (*Sp*HEX) [20]. Crystallographic studies of *Sp*HEX comprise a significant portion of the research described in this thesis and will discussed in detail in chapters 2-4. Both *Sm*CHB and *Sp*HEX contain similar $(\alpha/\beta)_8$ catalytic domains and multiple sequence alignments of the active site regions of these proteins with other family 20 glycosidases showed that the active site architectures of all family 20 glycosides are similar (Figure 1.9) [103]. However, *Sm*CHB and *Sp*HEX are monomeric, and the % identity of these enzymes with either subunit of human Hex is <30% in the active site region and virtually non-existent elsewhere. Hence, homology modeling studies of human Hex based on either *Sm*CHB or *Sp*HEX do not address questions concerning the mechanism of G_{M2} ganglioside hydrolysis by Hex A nor the protein:protein interactions involved in isoenzyme dimer formation. Nonetheless, while not definitive, the crystal structures of *Sm*CHB and *Sp*HEX, by virtue of the

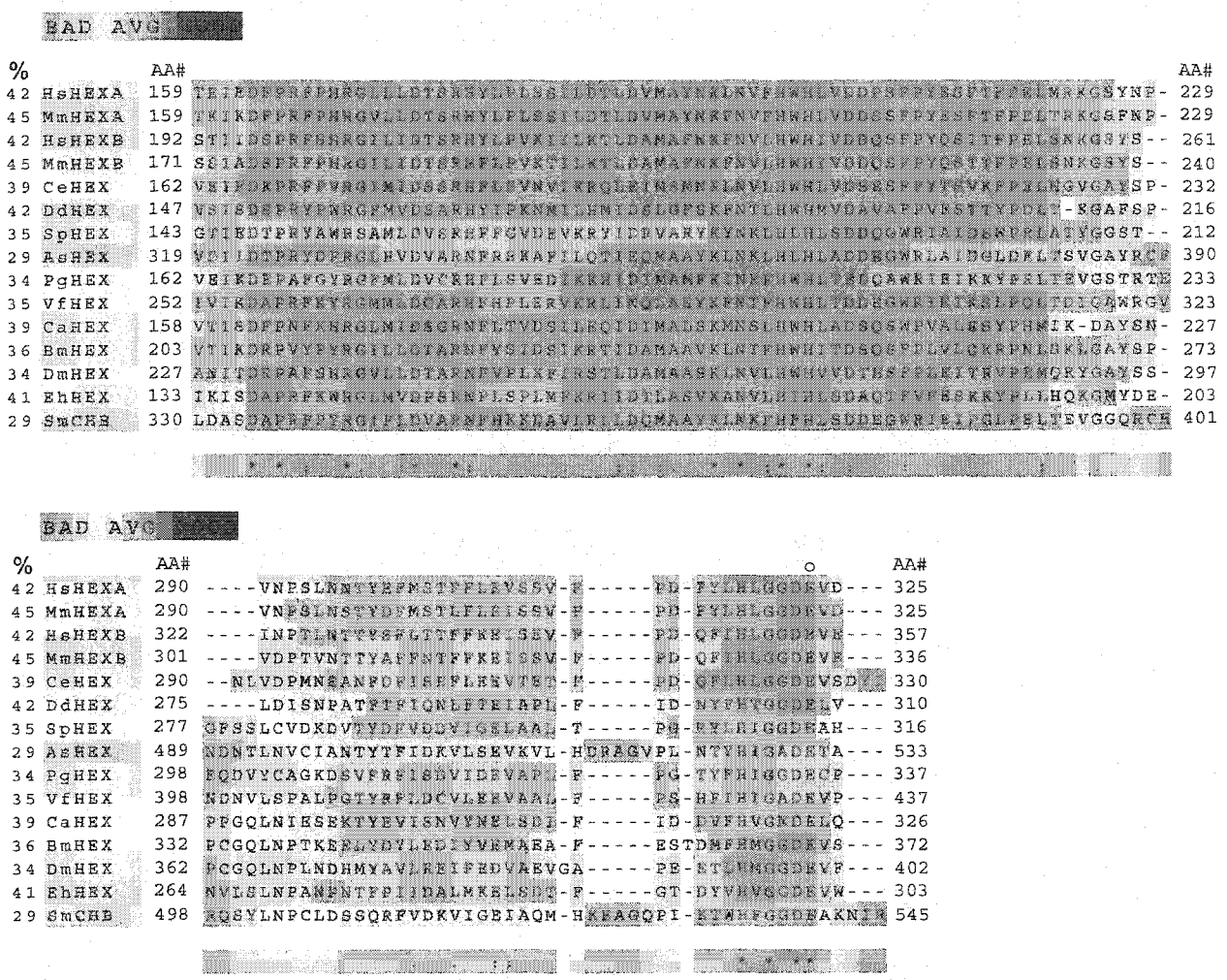


Figure 1.9. Multiple sequence alignment of two separate regions composing residues of the active site of representative family 20 hexosaminidases. The overall % identity for each sequence is indicated in front of each sequence name, and each sequence name is colored according to overall alignment score (bad, average or good). Consensus is shown underneath the alignment for identical (*), very similar (:), and similar(.) residues. The conserved general acid/base residue is indicated (o). Sequences were aligned with the program T-COFFEE (http://www.ch.embnet.org/software/TCoffee.html). The abbreviations used and database identifiers are as follows: HsHEXA, human Hex α–chain (Swiss-prot: P06865); MmHEXA, mouse Hex α-chain (Swiss-prot: P29416); HsHEXB, human Hex β-chain (Swiss-prot: P07686); MmHEXB, mouse β-chain (Swiss-prot: P20060); CeHEX, *Caenorhabditis elegans* Hex (Swiss-prot: Q22492); DdHEX, *Dictyostelium discoidem* Hex (Swiss-prot: P12723); SpHEX, *Sterptomyces plicatus* Hex (TrEMBL: O85361); AsHEX, *Alteromonas sp.* Hex (Swiss-prot: P49007); PgHEX, *Porphyromonas gingivalis* Hex (Swiss-prot: P49008); VfHEX, *Vibrio furnissii* Hex (Swiss-prot: P96155); CaHEX, *Candida albicans* Hex (Swiss-prot: P43077); BmHEX, *Bombyx mori* Hex (Swiss-prot: P49010); DmHEX, *Drosophila melanogaster* Hex (TrEMBL: Q9W3C4); EhHEX, *Entamoeba histolytica* Hex (Swiss-prot: P49009); SmCHB, *Serratia marcescens* Chitobiase (Swiss-prot: Q54468).

sequence similarity shared among glycosidases within the family, have been able to resolve many elusive questions regarding the active site architecture and catalytic mechanism of human Hex, and the family 20 glycosidases in general.

D.3.1. Substrate distortion.

According to the currently accepted –n to +n nomenclature for glycosidase sugar binding subsites [104] (where n is an integer and –n signifies sugar residues extending towards the non-reducing end and +n towards the reducing end of the bound oligosaccharide), the scissile glycosidic bond of the substrate spans subsites -1 and +1. Crystallographically determined structures of complexes between intact substrates and retaining β-glycosidases demonstrate the presence of significant distortions in the pyranoside ring of the sugar bound in the –1 subsite prior to bond cleavage (Figure 1.4) [17, 105-107]. Distortion in the pyranoside ring bound in the –1 subsite was first proposed for HEW lysozyme in the late 1960's [108], and most recently supported by a high-resolution structure determination of the complex of HEW lysozyme with NAM-NAG-NAM bound as a product [109]. However, although the product complex between HEW lysozyme and NAM-GlcNAc-NAM suggested a sofa conformation for the pyranose ring of the NAM residue bound in the –1 subsite, the true extent of substrate distortion was not revealed until structures of retaining β-glycosidases were determined in complex with intact substrates spanning the active site. These structural results indicated that substrate interactions at the +1 subsite contribute significantly to pyranoside ring distortion in the –1 subsite [17, 110].

The first three-dimensional structure of an intact substrate spanning the active site of a glycosidase was determined crystallographically and was a complex between *Sm*CHB and its natural substrate chitobiose [17]. Although it is unclear why chitobiose was not turned over in these chitobiase crystals, the non-reducing and reducing NAG residues of the intact chitobiose molecule were located in the –1 and +1 subsites, respectively. The pyranoside ring of the sugar bound in subsite –1 was distorted from the standard $^4C_1$ chair conformation into a skew-boat conformation such that the scissile bond and the leaving group were

held in a pseudo-axial orientation by the enzyme (Figure 1.4). Representative hexosaminidases of both *endo* and *exo*-acting enzymes from families 18 and 20, respectively, demonstrate this mode of substrate distortion in their respective Michaelis complexes (Figure 1.4) [17, 107]. The three-dimensional structure of a Michaelis complex for a family 56 glycosidase has not been determined thus far. The structure of a family 56 hyaluronidase from bee venom however, has been solved in complex with a hyaluronic acid tetramer bound as a product [19].

The substrate distortions seen in the Michaelis complexes for the family 18 and 20 glycosidases are nearly identical to those observed in the crystallographically determined Michaelis complexes of retaining cellulases from families 5 and 7 [105, 106]. It has been postulated that similar substrate distortion also occurs in the enzyme-substrate complexes of inverting β-glycosidases [111, 112]. Together, these findings suggest that a distorted pseudo-axial position for the glycosidic linkage and leaving group is a general mechanistic requirement for all retaining and inverting β-glycosidases. Indeed, distortion of the pyranoside ring in the –1 site from a $^4C_1$ chair towards a boat conformation helps to satisfy the requirements of stereoelectronic theory. A boat conformation ($^4B^1$) orients a lone pair of electrons on the ring oxygen (O5) antiperiplanar with respect to the scissile bond and leaving group (Figure 1.10). This conformational arrangement maximizes electron donation from O5 to the antibonding orbital at the electron deficient anomeric center and thus accommodates much of the positive charge developing at the anomeric center during the transition states [113, 114]. The pseudo-axial orientation of the leaving group also places the oxygen atom of the glycosidic bond in an appropriate position for lateral protonation by the enzymic general acid catalyst [110].

D.3.2. Substrate assisted catalysis revealed.

Stereochemical outcome studies on *Sm*CHB [115] and human β-hexosaminidase B [116] have demonstrated that this family operates via a

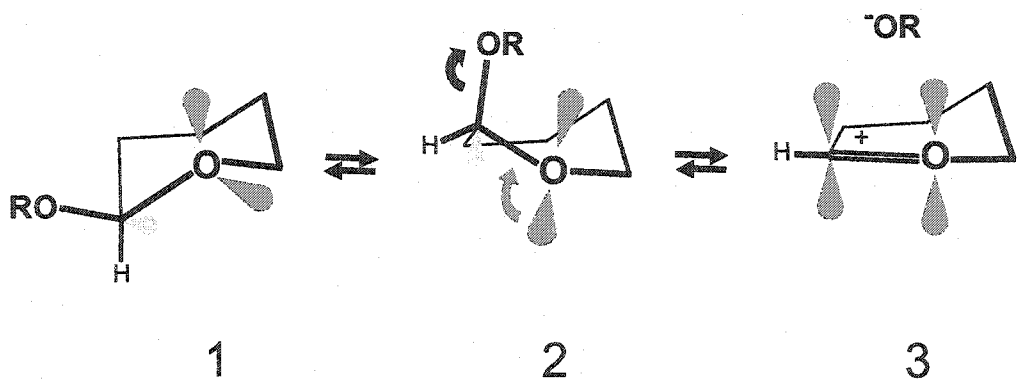


Figure 1.10. Schematic for the conformational itinerary of the substrate leading to the first transition state during β-glycosidase catalyzed bond hydrolysis. 1) Pyranose ring conformation ($^4C_1$) prior to being bound by the enzyme. 2) Pyranose ring conformation ($^4B_1$)as determined for the Michaelis complex [17]. 3) Sofa or half chair conformation of the first transition state structure. Note in (2) how the distortion rotates a lone pair of electrons from the ring oxygen antiperiplanar to the scissile bond and leaving group (R). This distortion also moves the hydrogen atom of the anomeric center out of the way of the waiting nucleophile, and exposes the antibonding orbital.

retaining mechanism. Unexpectedly however, the crystal structure of *Sm*CHB was found to lack a carboxylate group suitably disposed to stabilize the oxacarbenium ion transition state through the formation of a covalent glycosyl-enzyme intermediate [17]. Instead, the X-ray structural analysis of *Sm*CHB and further kinetic studies with inhibitors provided strong evidence for catalysis involving participation of the neighbouring C2-acetamido group on the substrate as described in section *B.3*, and illustrated in Figure 1.2b, *lower pathway* [17, 21, 115, 117].

Interestingly, it can be shown that the nucleophilic oxygen atom of Glu197 of the family 7 β-retaining endoglucanase I from *Fusarium oxysporum* (which uses the covalent catalytic mechanism described in Figure 1.2b, *upper pathway*) essentially overlaps the position of the C2-acetamido carbonyl oxygen atom of the chitobiose bound to the family 20 chitobiase from *Serratia marcescens* when the two structures are superimposed (Figure 1.11) [105]. Thus, anchimeric assistance by the neighboring C2-acetamido group of the substrate not only replaces the need for a second enzymic carboxyl group , but also the stereochemistry of the nucleophilic attack is conserved between retaining β-glycosidases that use either an enzymic or substrate-based nucleophile.

Based on the crystallographic complex of *Sm*CHB with chitobiose, it was proposed that nucleophilic attack by the carbonyl oxygen atom of the C2-acetamido group would result in the formation of an enzyme stabilized oxazolinium ion intermediate. Indeed, research described in chapter 2 of this thesis provides the crystallographic evidence necessary for the stabilization of an analogue of the oxazolinium ion intermediate in the active site of *Sp*HEX.

D.3.3. Identification of family 20 general acid/base and substrate binding residues. Analysis of the crystallographic complex between *Sm*CHB and chitobiose revealed numerous important catalytic residues. In particular, the carboxyl group of Glu540 was found to be within 2.5 Å of the glycosidic oxygen of

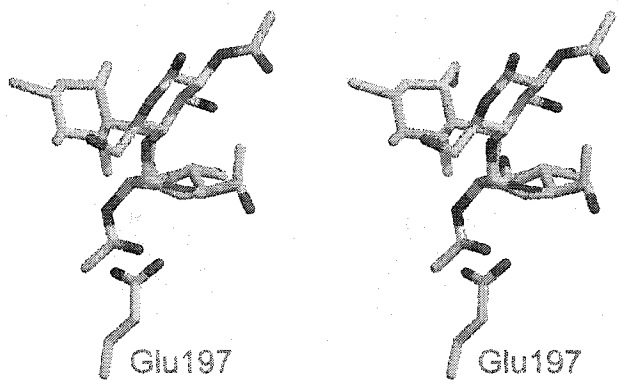


Figure 1.11. Superposition of the crystallographic Michaelis complex of a family 7 configuration-retaining β-endoglucanase I from *Fusarium oxysporum* **(green) [105] and the family 20 chitobise from** *Serratia marcescens* **(gray) [17].** As previously demonstrated by Sulzenbacher *et al.* [105], this superposition demonstrates that the non-hydrolyzable substrate analogue bound to endoglucanse I adopts the same distorted conformation as chitobiose bound to the family 20 *S. marcescens* chitobiase. Moreover, the carboxylate of the enzymic nucleophile of endoglucanse I superposes almost perfectly with the C2-acetamido group oxygen of the non-reducing sugar of chitobiose bound to the family 20 enzyme. This remarkable spatial equivalence clearly indicates that the carbonyl oxygen of the C2-acetamido group is suitably positioned to act in place of an enzymic nucleophile in the catalyic mechanism of family 20 glycosidases.

the bound substrate, and it was concluded that this residue was the general acid/base catalytic residue of the enzyme [17]. Indeed, the mutation Glu540Ala lead to an approximately 1000-fold decrease in enzymatic efficiency (K_{cat}/K_m), thereby providing support for the idea that Glu540 was the general acid/base catalytic residue [118]. Multiple sequence alignments and homology modeling studies showed that the α-subunit residue Glu323 of human Hex A is the general acid/base similar to *Sm*CHB Glu540 [87]; whereas, the related residue in the β-subunit is Glu355 [119] (Figure 1.9). Indeed, identifying Glu355 as an important catalytic residue in the β-subunit of human Hex corroborates the photoaffinity labeling experiments of Liessem *et al.* [102] (Section ***D.2.***).

Mutagenesis studies based on the above data verified Glu323 and Glu355 as likely candidates for the general acid/base catalytic residues for the α- and β-subunit, respectively. The mutation αGlu323Gln produced mature Hex A with an α-subunit having normal substrate binding characteristics, but a dramatically reduced activity towards the α-subunit specific substrate 4-MUGS [87]. Substitution of βGlu355 to Ala, Gln, Asp or Trp resulted in mature Hex A with no activity towards 4-MUG (inactive β-subunit), but with normal activity toward G_{M2}-ganglioside in the presence of the G_{M2} activator protein (active α-subunit) [119]. Thus, these mutations disrupted β-subunit activity without affecting the ability of the α-subunit to hydrolyze G_{M2}-ganglioside. No point mutations at codons α323 or β355 of human Hex have been found to occur naturally.

Prior to determining the crystal structure of *Sp*HEX (chapter 2), a homology model of the enzyme was built using *Sm*CHB as a template, and identified Glu314 as the general acid/base catalytic residue of this enzyme. Glu314 is analogous to Glu540 of *Sm*CHB and αGlu323 and βGlu355 of human Hex [103]. The *Sp*HEX variant Glu314Gln provided biochemical support for the role of this residue as the general acid/base in the catalytic mechanism of this enzyme: the mutation reduced both the V_{max} (296-fold) and K_m (7-fold) relative to wild-type enzyme [103]; however, this is only a modest reduction in V_{max}, and not what would be expected when removing the general acid/base residue from the

enzyme. The appreciable activity remaining after elimination of the proposed general acid/base residue from the active sites of *Sp*HEX and *Sm*CHB, may be due to the favorable stereoelectronic effects achieved by the enzyme via the substrate distortions (see section ***D.3.1.***). Substrate distortion alone may lower the activation energy sufficiently to drive the reaction forward at a measurable rate. Such residual activity was probably present in the general acid/base variants of human Hex A (αGlu323Gln) and B (βGlu355Gln); however, the residual activity may not have been measurable due to the technical difficulties associated with the expression and purification of these recombinant proteins (see introduction to section ***D***).

Structural analysis of *Sm*CHB revealed decisive insight into the catalytic role of αArg178 and αAsp258 of human Hex A, mutations of which give rise to the variant B1 phenotype (see section ***D.1***). The crystallographic complex of *Sm*CHB with chitobiose indicated that Arg349 was a crucial substrate binding residue; the guanidinium group of Arg349 formed a bi-dentate hydrogen bonding interaction with O3 and O4 of the non-reducing sugar of chitobiose [17]. Arg349 of *Sm*CHB is conserved in all known family 20 glycosidases. It is clearly related in sequence to αArg178 and βArg211 of the α- and β-subunits of human Hex, respectively, and to Arg162 of *Sp*HEX [103]. Comparative molecular models of the α- and β-subunits of human Hex [17] and *Sp*HEX [103] suggested that this residue carried out the same function in these enzymes. Indeed, emulating the variant B1 mutation αArg178His in *Sp*HEX (Arg162His) resulting in an *Sp*HEX variant with a K_m 40-fold greater than that of the wild-type enzyme. The V_{max} was reduced only 5-fold relative to wild-type enzyme [103]. These kinetic data collected from the Arg162His variant of *Sp*HEX provided the first biochemical evidence that this conserved Arg residue acts as an important substrate binding residue in the catalytic mechanism of family 20 glycosidases.

Based on the variant B1 phenotype of the human Hex A mutant αAsp258His, it was suggested that Asp258 may act as the general acid/base residue in the catalytic mechanism of this enzyme [99]. The analogous residue in *Sm*CHB (Asp448), although close to the active site, was found to be

inappropriately positioned to carry out such a role during catalysis. Mutation of the related residue in *Sp*HEX (Asp246Asn) [103] and human Hex B (Asp290Asn) [120] resulted in only small reductions in kinetic performance of these enzymes (*Sp*HEX: V_{max} decreased 1.2-fold and K_m increased 1.2-fold relative to wild-type; Hex B: kcat 70% of wild-type, K_m increased 3-fold relative to wild-type). Asp258 is conserved in all family 20 enzymes and is a second tier active site residue that acts to stabilize the orientation of another conserved Asp that is directly involved in orienting the substrate C2-acetamido group during catalysis [20, 118, 121]. Investigation into the biochemical and structural properties of this latter Asp residue in the catalytic mechanisms of *Sp*HEX comprises the research described in Chapter 4 of this thesis.

E. Research Overview

As illustrated above, by choosing to study bacterial family 20 glycosidases that are amenable to currently available prokaryotic recombinant protein expression systems, it is possible to produce and isolate enough bacterial glycosidase so that crystallographic studies and accurate kinetic analyses can be readily carried out. This approach to studying family 20 β-hexosaminidases has generated a wealth of data about these enzymes that has not been obtainable through studies on the human isoenzymes. Chapters 2 to 4 of this thesis continue along the theme of using bacterial glycosidases and describe crystallographic studies of a bacterial family 20 glycosidase (*Sp*HEX) that are aimed at providing a more comprehensive understanding of the catalytic mechanism of these enzymes. Moreover, chapter 5 provides direct insight into human Hex structure and function through the crystal structure determination of human Hex B.

Thus, the overall goal of this research thesis has been to enhance our knowledge of the structural and functional relationships inherent in family 20 glycosidases, and to use this information to understand the molecular basis of G_{M2}-gangliosidosis better. Specific research projects attempting to fulfill this goal are outlined below:

Chapter 2: This chapter describes the results of two X-ray crystallographic studies; the *de novo* crystal structure determination of native *Sp*HEX obtained by a multiple wavelength anomalous dispersion (MAD) phasing experiment, and a complex of this enzyme with the cyclic intermediate analogue NAG-thiazoline [20]. NAG-thiazoline is a relatively stable analogue of the proposed oxazolinium ion intermediate formed by family 20 glycosidases (Figure 1.2b, *lower pathway*). The crystallographic complex between this mechanistic inhibitor and *Sp*HEX provides decisive structural evidence for substrate-assisted catalysis and the formation of a covalent, cyclic intermediate in family 20 β-hexosaminidases.

Chapter 3: A crystallographic complex between *Sp*HEX and the transition state mimic GalNAc-isofagomine is presented [122]. 1-*N*-azasugar inhibitors of the isofagomine class are potent competitive inhibitors of configuration-retaining β-glycosidases. This potency results from the formation of a strong hydrogen-bonded electrostatic interaction between the protonated endocyclic nitrogen at the 'anomeric' center of the inhibitor and the catalytic nucleophile of the enzyme. However, family 20 enzymes lack an enzymic nucleophile. Nonetheless, GalNAc-isofagomine was found to be a potent inhibitor of *Sp*HEX, and the crystallographic complex between this enzyme and GalNAc-isofagomine reveals a novel binding mode for the inhibitor which could form the basis for a new class of pharmaceuticals [122].

Chapter 4: The role of a conserved aspartate residue (Asp313) in the active site of *Sp*HEX is discussed in chapter 4 [121]. Through the X-ray crystallographic and kinetic studies of two *Sp*HEX variants, Asp313Ala and Asp313Asn, it was discovered that Asp313 participates in catalysis by orienting the 2-acetamido group of the substrate and by stabilizing the transition state (in addition to stabilizing the positive charge that develops on the oxazoline ring nitrogen upon cyclization) (Figure 1.2b, *lower pathway*).

Chapter 5: This chapter describes the *de novo* crystal structure determination of human placental Hex B by the multiple isomorphous replacement (MIR) method. This is the first three-dimensional structure determination of a human Hex isoenzyme, and because the isoenzyme was purified from its natural source, post-translational modifications could be identified in the electron density maps and are included in the final molecular model. Furthermore, the mechanistic inhibitors GalNAc-isofagomine or NAG-thiazoline were soaked into Hex B crystals and the structures of these complexes determined by difference Fourier analysis. From these crystal structures of Hex B and the known X-ray structure of the G_{M2}-activator protein, a model of Hex A was built as well as the quaternary complex of Hex A in complex with the activator protein and G_{M2}-ganglioside. Together, the crystallographic and modeling data demonstrate how the α- and β-subunits dimerize to form either Hex A or Hex B, how these isoenzymes hydrolyze diverse substrates, and how many documented point mutations cause Sandhoff disease (β-subunit mutations) and Tay-Sachs disease (α-subunit mutations).

Chapter 6: Summary.

F. References

1. Sinnott, M.L., *Catalytic Mechanisms of Enzymic Glycosyl Transfer.* Chem Rev, 1990. **90**: p. 1171-1202.
2. McCarter, J.D. and S.G. Withers, *Mechanisms of enzymatic glycoside hydrolysis.* Curr Opin Struct Biol, 1994. **4**(6): p. 885-92.
3. Henrissat, B., *A classification of glycosyl hydrolases based on amino acid sequence similarities.* Biochem J, 1991. **280**(Pt 2): p. 309-16.
4. Henrissat, B. and A. Bairoch, *New families in the classification of glycosyl hydrolases based on amino acid sequence similarities.* Biochem J, 1993. **293**(Pt 3): p. 781-788.
5. Henrissat, B. and A. Bairoch, *Updating the sequence-based classification of glycosyl hydrolases.* Biochemical Journal, 1996. **316**: p. 695-696.
6. Henrissat, B., et al., *Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases [published erratum appears in Proc Natl Acad Sci U S A 1996 May 28;93(11):5674].* Proc Natl Acad Sci U S A, 1995. **92**(15): p. 7090-4.
7. Henrissat, B. and G. Davies, *Structural and sequence-based classification of glycoside hydrolases.* Curr Opin Struct Biol, 1997. **7**(5): p. 637-644.
8. Davies, G.J., M.L. Sinnott, and S.G. Withers, *Glycosyl Transfer*, in *Comprehensive Biological Catalysis*, M.L. Sinnott, Editor. 1998, Academic Press Ltd.: New York. p. 119-208.
9. Koshland, D.E., *Stereochemistry and the mechanism of enzymatic reactions.* Biol. Rev., 1953. **28**: p. 416-436.
10. Davies, G. and B. Henrissat, *Structures and mechanisms of glycosyl hydrolases.* Structure, 1995. **3**(9): p. 853-859.
11. McCarter, J.D., M.J. Adam, and S.G. Withers, *Binding-energy and catalysis - fluorinated and deoxygenated glycosides as mechanistic probes of Escherichia-coli (Lacz) beta-galactosidase.* Biochemical Journal, 1992. **286**: p. 721-727.

12. Namchuk, M.N. and S.G. Withers, *Mechanism of Agrobacterium beta-glucosidase: Kinetic analysis of the role of noncovalent enzyme/substrate interactions.* Biochemistry, 1995. **34**(49): p. 16194-16202.

13. Jencks, W.P., *Catalysis in chemistry and enzymology*. McGraw-Hill series in advanced chemistry. 1969, New York: McGraw-Hill, Inc. 644.

14. Fersht, A., *Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding*. 1998: W. H. Freeman and Co. 631.

15. Piszkiewicz, D. and T.C. Bruice, *Glycoside hydrolysis. II. Intramolecular carboxyl and acetamido group catalysis in beta-glycoside hydrolysis.* J Am Chem Soc, 1968. **90**(8): p. 2156-2163.

16. Mark, B.L. and M.N. James, *Anchimeric assistance in hexosaminidases.* Can. J. Chem., 2002. **80**(8): p. 1064-1074.

17. Tews, I., et al., *Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease.* Nat Struct Biol, 1996. **3**(7): p. 638-648.

18. Tews, I., et al., *Substrate-assisted catalysis unifies two families of chitinolytic enzymes.* J Am Chem Soc, 1997. **119**: p. 7954-7959.

19. Markovic-Housley, Z., et al., *Crystal structure of hyaluronidase, a major allergen of bee venom.* Structure Fold Des, 2000. **8**(10): p. 1025-1035.

20. Mark, B.L., et al., *Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase.* Journal of Biological Chemistry, 2001. **276**(13): p. 10330-10337.

21. Knapp, S., et al., *NAG-thiazoline, an N-Acetyl-beta-hexosaminidase inhibitor that implicates acetamido participation.* J Am Chem Soc, 1996. **118**: p. 6804-6805.

22. Gravel, R.A., et al., *The G_{M2} gangliosidoses*, in *The metabolic and molecular bases of inherited disease*, C.R. Scriver, et al., Editors. 1995, McGraw-Hill: New York. p. 2839-2879.

23. Tay, W., *Symmetrical changes in the region of the yellow spot in each eye of an infant.* Trans. Ophthalmol. Soc. U.K., 1881. **1**: p. 155-157.

24. Sachs, B., *On arrested cerebral development with special reference to its cortical pathology.* J. Nerv. Ment. Dis., 1887. **14**: p. 541-553.

25. Sachs, B., *A family form of idiocy, generally fatal, associated with early blindness (amaurotic family idiocy).* NY State J. Med., 1896. **63**: p. 697-703.

26. Sandhoff, K., *The G_{M2}-gangliosidoses and the elucidation of the β-hexosaminidase system.* Adv Genet, 2001. **44**: p. 67-91.

27. Okada, S. and J.S. O'Brien, *Tay-Sachs disease: generalized absence of a beta-D-N- acetylhexosaminidase component.* Science, 1969. **165**(894): p. 698-700.

28. Robinson, D. and J.L. Stirling, *N-Acetyl-beta-glucosaminidases in human spleen.* Biochem J, 1968. **107**(3): p. 321-327.

29. Ropers, H.H. and U. Schwantes, *On the molecular basis of Sabdhoff's disease.* Humangenetik, 1973. **20**: p. 167-170.

30. Myerowitz, R., et al., *Human beta-hexosaminidase alpha chain: coding sequence and homology with the beta chain.* Proc Natl Acad Sci U S A, 1985. **82**(23): p. 7830-7834.

31. Proia, R.L. and E. Soravia, *Organization of the gene encoding the human beta-hexosaminidase alpha- chain.* J Biol Chem, 1987. **262**(12): p. 5677-5681.

32. Nakai, H., et al., *Assignment of beta-hexosaminidase A alpha-subunit to human chromosomal region 15q23----q24.* Cytogenet Cell Genet, 1991. **56**(3-4): p. 164.

33. Korneluk, R.G., et al., *Isolation of cDNA clones coding for the alpha-subunit of human beta- hexosaminidase. Extensive homology between the alpha- and beta-subunits and studies on Tay-Sachs disease.* J Biol Chem, 1986. **261**(18): p. 8407-8413.

34. O'Dowd, B.F., et al., *Isolation of cDNA clones coding for the beta subunit of human beta- hexosaminidase.* Proc Natl Acad Sci U S A, 1985. **82**(4): p. 1184-1188.

35. Proia, R.L., *Gene encoding the human beta-hexosaminidase beta chain: extensive homology of intron placement in the alpha- and beta-chain genes.* Proc Natl Acad Sci U S A, 1988. **85**(6): p. 1883-1887.

36. Neote, K., et al., *Translation initiation in the HEXB gene encoding the β-subunit of human β-hexosaminidase.* J. Biol. Chem., 1990. **265**: p. 20799-20806.

37. Quon, D.V.K., et al., *Proteolytic processing of the β-subunit of the lysosomal enzyme, β-hexosaminidase, in normal human fibroblasts.* J. Biol. Chem., 1989. **264**: p. 3380-3384.

38. Little, L.E., et al., *Proteolytic processing of the α chain of the lysosomal enzyme β-hexosaminidase, in normal human fibroblasts.* J. Biol. Chem., 1988. **263**: p. 4288-4292.

39. Stirling, J., et al., *Localization of the Pro-Sequence within the Total Deduced Primary Structure of Human β-Hexosaminidase B.* FEBS. LETT., 1988. **231**: p. 47-50.

40. Weitz, G. and R.L. Proia, *Analysis of the glycosylation and phosphorylation of the alpha-subunit of the lysosomal enzyme, beta-hexosaminidase A, by site-directed mutagenesis.* J Biol Chem, 1992. **267**(14): p. 10039-10044.

41. Sonderfeld-Fresko, S. and R.L. Proia, *Analysis of the Glycosylation and Phosphorylation of the lysosomal enzyme, β-hexosaminidase B, by site-directed mutagenesis.* J. Biol. Chem., 1989. **264**: p. 7692-7697.

42. Kornfeld, R. and S. Kornfeld, *Assembly of asparagine-linked oligosaccharides.* Ann. Rev. Biochem., 1985. **54**: p. 631-664.

43. Mahuran, D.J., *Beta-hexosaminidase: biosynthesis and processing of the normal enzyme, and identification of mutations causing Jewish Tay-Sachs disease.* Clin Biochem, 1995. **28**(2): p. 101-106.

44. Bach, G., R. Bargel, and M. Cantz, *I-cell Disease: Deficiency of extracellular hydrolase phosphorylation.* Biochem. Biophys. Res. Commun., 1979. **91**: p. 976-981.

45. Kornfeld, S., *Trafficking of lysosomal enzymes in normal and disease states.* J. Clin. Invest., 1986. **77**: p. 1-6.

46. Griffiths, G., et al., *The mannose 6-phosphate receptor and the biogenesis of the lysosomes.* Cell, 1988. **52**: p. 329-341.

47. Lang, L., et al., *Lysosomal enzyme phosphorylation. Recognition of a protein-dependent determinant allows specific phosphorylation of oligosaccharides present on lysosomal enzymes.* J Biol Chem, 1984. **259**(23): p. 14663-14671.

48. Kornfeld, S. and I. Mellman, *The biogenesis of lysosomes.* Annu Rev Cell Biol, 1989. **5**: p. 483-525.

49. Mahuran, D.J., et al., *Proteolytic processing of human pro-β hexosaminidase: Identification of the internal site of hydrolysis that produces the nonidentical β_a and β_b polypeptides in the mature β-subunit.* J. Biol. Chem., 1988. **263**: p. 4612-4618.

50. Mahuran, D.J., *Characterization of human placental β-hexosaminidase I_2: Proteolytic processing intermediates of hexosaminidase A.* J. Biol. Chem., 1990. **265**: p. 6794-6799.

51. Mahuran, D.J., *Biochemical consequences of mutations causing the GM2 Gangliosidoses.* Biochim. Biophys. Acta., 1999. **1455**: p. 105-138.

52. Schuette, C.G., et al., *Sphingolipid activator proteins: proteins with complex functions in lipid degradation and skin biogenesis.* Glycobiology, 2001. **11**(6): p. 81R-90R.

53. Srivastava, S.K. and E. Beutler, *Hexosaminidase-A and hexosaminidase-B: studies in Tay-Sachs' and Sandhoff's disease.* Nature, 1973. **241**(5390): p. 463.

54. Beutler, E. and W. Kuhl, *Subunit structure of human hexosaminidase verified: interconvertibility of hexosaminidase isozymes.* Nature, 1975. **258**(5532): p. 262-264.

55. Kresse, H., et al., *Liberation of N-acetylglucosamine-6-sulfate by human beta-N- acetylhexosaminidase A.* J Biol Chem, 1981. **256**(24): p. 12926-12932.

56. Bayleran, J., P. Hechtman, and W. Saray, *Synthesis of 4-methylumbelliferyl-beta-D-N-acetylglucosamine-6-sulfate and its use in classification of GM2 gangliosidosis genotypes.* Clin Chim Acta, 1984. **143**(2): p. 73-89.

57. Navon, R., B. Padeh, and A. Adam, *Apparent deficiency of hexosaminidase A in healthy members of a family with Tay-Sachs disease.* Am J Hum Genet, 1973. **25**(3): p. 287-293.

58. Tallman, J.F., et al., *Ganglioside catabolism in hexosaminidase A-deficient adults.* Nature, 1974. **252**(5480): p. 254-255.

59. Wenger, D.A., S. Okada, and J.S. O'Brien, *Studies on the substrate specificity of hexosaminidase A and B from liver.* Arch Biochem Biophys, 1972. **153**(1): p. 116-129.

60. Kytzia, H.J. and K. Sandhoff, *Evidence for two different active sites on human beta-hexosaminidase A. Interaction of GM2 activator protein with beta-hexosaminidase A.* J Biol Chem, 1985. **260**(12): p. 7568-7572.

61. Kaback, M.M., *Thermal fractionation of serum hexosaminidases: applications to heterozygote detection and diagnosis of Tay-Sachs disease.* Methods in Enzymology, 1972. **28**: p. 862-867.

62. Conzelmann, E. and K. Sandhoff, *AB variant of infantile GM2 gangliosidosis: deficiency of a factor necessary for stimulation of hexosaminidase A-catalyzed degradation of ganglioside GM2 and glycolipid GA2.* Proc Natl Acad Sci U S A, 1978. **75**(8): p. 3979-3983.

63. Conzelmann, E. and K. Sandhoff, *Purification and characterization of an activator protein for the degradation of glycolipids GM2 and GA2 by hexosaminidase A.* Hoppe Seylers Z Physiol Chem, 1979. **360**(12): p. 1837-1849.

64. Heng, H.H., et al., *Refined mapping of the GM2 activator protein (GM2A) locus to 5q31.3- q33.1, distal to the spinal muscular atrophy locus.* Genomics, 1993. **18**(2): p. 429-431.

65. Burg, J., et al., *Mapping of the gene coding for the human GM2 activator protein to chromosome 5.* Ann Hum Genet, 1985. **49**(Pt 1): p. 41-45.

66. Sandhoff, K., K. Harzer, and W. Fürst, *Sphingolipid activator proteins*, in *The Metabolic Basis of Inherited Disease*, C.R. Scriver, et al., Editors. 1995, McGraw-Hill: New York. p. 2427-2441.

67. Yadao, F., P. Hechtman, and F. Kaplan, *Formation of a ternary complex between GM2 activator protein, GM2 ganglioside and hexosaminidase A.* Biochim Biophys Acta, 1997. **1340**(1): p. 45-52.

68. Tse, R., et al., *Identification of Functional Domains within the α and β Subunits of β-Hexosaminidase A Through the Expression of α-β Fusion Proteins.* Biochemistry, 1996. **35**: p. 10894-10903.

69. Pennybacker, M., et al., *Identification of domains in human beta-hexosaminidase that determine substrate specificity.* J Biol Chem, 1996. **271**(29): p. 17377-17382.

70. Kolodny, E.H., *Molecular genetics of the beta-hexosaminidase isoenzymes: an introduction.* Adv Genet, 2001. **44**: p. 101-126.

71. Sandhoff, K., *Variation of beta-N-acetylhexosaminidase-pattern in Tay-Sachs disease.* FEBS Lett, 1969. **4**(4): p. 351-354.

72. Conzelmann, E. and K. Sandhoff, *Partial enzyme deficiencies: residual activities and the development of neurological disorders.* Dev Neurosci, 1983. **6**(1): p. 58-71.

73. Kaback, M.M., T.J. Nathan, and S. Greenwald, *Tay-Sachs disease: Heterozygote screening and prenatal diagnosis - U.S. experience and world perspective*, in *Tay-Sachs disease: Screening and prevention*, M.M. Kaback, D.L. Rimoin, and J.S. O'Brien, Editors. 1977, Alan R. Liss Inc.: New York.

74. O'Brien, J.S., et al., *Tay-sachs disease. Detection of heterozygotes and homozygotes by serum hexosaminidase assay.* N Engl J Med, 1970. **283**(1): p. 15-20.

75. Triggs-Raine, B., D.J. Mahuran, and R.A. Gravel, *Naturally Occurring Mutations in G_{M2} Gangliosidosis: A Compendium*, in *Tay-Sachs Disease*, R.J. Desnick and M.M. Kaback, Editors. 2001, Academic Press: San Diego. p. 199-224.

76. Rattazzi, M.C. and K. Dobrenis, *Treatment of GM2 gangliosidosis: past experiences, implications, and future prospects.* Adv Genet, 2001. **44**: p. 317-339.

77. Platt, F.M., et al., *Prevention of lysosomal storage in Tay-Sachs mice treated with N- butyldeoxynojirimycin.* Science, 1997. **276**(5311): p. 428-431.

78. Yamanaka, S., et al., *Targeted disruption of the Hexa gene results in mice with biochemical and pathologic features of Tay-Sachs disease.* Proc Natl Acad Sci U S A, 1994. **91**(21): p. 9975-9979.

79. Sango, K., et al., *Mouse models of Tay-Sachs and Sandhoff diseases differ in neurologic phenotype and ganglioside metabolism.* Nat Genet, 1995. **11**(2): p. 170-176.

80. Platt, F.M., et al., *Extensive glycosphingolipid depletion in the liver and lymphoid organs of mice treated with N-butyldeoxynojirimycin.* J Biol Chem, 1997. **272**(31): p. 19365-19372.

81. Jeyakumar, M., et al., *Enhanced survival in Sandhoff disease mice receiving a combination of substrate deprivation therapy and bone marrow transplantation.* Blood, 2001. **97**(1): p. 327-329.

82. Jeyakumar, M., et al., *Delayed symptom onset and increased life expectancy in Sandhoff disease mice treated with N-butyldeoxynojirimycin.* Proc Natl Acad Sci U S A, 1999. **96**(11): p. 6388-6393.

83. Norflus, F., et al., *Bone marrow transplantation prolongs life span and ameliorates neurologic manifestations in Sandhoff disease mice.* J Clin Invest, 1998. **101**(9): p. 1881-1888.

84. Cao, Z.M., et al., *Benign HEXA mutations, C739T(R247W) and C745T(R249W), cause beta-hexosaminidase A pseudodeficiency by reducing the alpha-subunit protein levels.* J. Biol. Chem., 1997. **272**(23): p. 14975-14982.

85. Brown, C.A. and D.J. Mahuran, *β-hexosaminidase isozymes from cells co-transfected with α and β cDNA constructs: Analysis of α subunit missense*

mutation associated with the adult form of Tay-Sachs disease. Am. J. Hum. Genet., 1993. **53**: p. 497-508.

86. Hou, Y., et al., *The role of beta-Arg211 in the active site of human beta-hexosaminidase B.* Biochemistry, 2000. **39**: p. 6219-6227.
87. Fernandes, M.J.G., et al., *Identification of candidate active site residues in lysosomal beta-hexosaminidase A.* J. Biol. Chem., 1997. **272**(2): p. 814-820.
88. Hou, Y., R. Tse, and D.J. Mahuran, *The Direct Determination of the Substrate Specificity of the α-Active site in Heterodimeric β-Hexosaminidase A.* Biochemistry, 1996. **35**: p. 3963-3969.
89. Prezant, T.R., *Expression of human lysosomal beta-hexosaminidase in yeast vacuoles.* Biochem Biophys Res Commun, 1990. **170**(1): p. 383-90.
90. Boose, J.A., et al., *Synthesis of a human lysosomal enzyme, beta-hexosaminidase B, using the baculovirus expression system.* Protein Expr Purif, 1990. **1**(2): p. 111-120.
91. Schuette, C.G., J. Weisgerber, and K. Sandhoff, *Complete analysis of the glycosylation and disulfide bond pattern of human beta-hexosaminidase B by MALDI-MS.* Glycobiology, 2001. **11**(7): p. 549-56.
92. Mahuran, D.J. and J.A. Lowden, *The subunit and polypeptide structure of hexosaminidase from human placenta.* Can. J. Biochem., 1980. **58**: p. 287-294.
93. dos Santos, M.R., et al., *GM2-gangliosidosis B1 variant: analysis of beta-hexosaminidase alpha gene mutations in 11 patients from a defined region in Portugal.* Am J Hum Genet, 1991. **49**(4): p. 886-890.
94. Goebel, H.H., et al., *B1 variant of GM2 gangliosidosis in a 12-year-old patient.* Pediatr Res, 1989. **25**(1): p. 89-93.
95. Tanaka, A., K. Ohno, and K. Suzuki, *G_{M2}-Gangliosidosis B1 variant: A wide geographic and ethnic distribution of the specific β-hexosaminidase α chain mutation originally identified in a Puerto Rican patient.* Biochem. Biophys. Res. Comm., 1988. **156**(2): p. 1015-1019.

96. Tanaka, A., et al., *GM2-gangliosidosis B1 variant: analysis of beta-hexosaminidase alpha gene abnormalities in seven patients [published erratum appears in Am J Hum Genet 1991 Jan;48(1):176].* Am J Hum Genet, 1990. **46**(2): p. 329-339.

97. Triggs-Raine, B.L., et al., *Sequence of DNA flanking the exons of the HEXA gene, and identification of mutations in Tay-Sachs disease.* Am J Hum Genet, 1991. **49**(5): p. 1041-1054.

98. Brown, C.A., et al., *Introduction of the α subunit mutation associated with the B1 variant of Tay-Sachs disease into the β subunit produces a β-hexosaminidase B without catalytic activity.* J. Biol. Chem., 1989. **264**: p. 21705-21710.

99. Bayleran, J., et al., *Tay-Sachs disease with hexosaminidase A: characterization of the defective enzyme in two patients.* Am J Hum Genet, 1987. **41**(4): p. 532-548.

100. Akli, S., et al., *Seven novel Tay-Sachs mutations detected by chemical mismatch cleavage of PCR-amplified cDNA fragments.* Genomics, 1991. **11**(1): p. 124-134.

101. Tanaka, A., H.H. Punnett, and K. Suzuki, *A new point mutation in the beta-hexosaminidase alpha subunit gene responsible for infantile Tay-Sachs disease in a non-Jewish Caucasian patient (a Kpn mutant).* Am J Hum Genet, 1990. **47**(3): p. 568-574.

102. Liessem, B., et al., *Photoaffinity labeling of human lysosomal beta-hexosaminidase B. Identification of Glu-355 at the substrate binding site.* J Biol Chem, 1995. **270**(40): p. 23693-23699.

103. Mark, B.L., et al., *Structural and functional characterization of Streptomyces plicatus beta-N-acetylhexosaminidase by comparative molecular modeling and site-directed mutagenesis.* J Biol Chem, 1998. **273**(31): p. 19618-19624.

104. Davies, G.J., K.S. Wilson, and B. Henrissat, *Nomenclature for sugar-binding subsites in glycosyl hydrolases [letter].* Biochem J, 1997. **321**(Pt 2): p. 557-559.

105. Sulzenbacher, G., et al., *Structure of the Fusarium oxysporum endoglucanase I with a nonhydrolyzable substrate analogue: Substrate distortion gives rise to the preferred axial orientation for the leaving group.* Biochemistry, 1996. **35**(48): p. 15280-15287.

106. Davies, G.J., et al., *Snapshots along an enzymatic reaction coordinate: analysis of a retaining beta-glycoside hydrolase.* Biochemistry, 1998. **37**(34): p. 11707-11713.

107. van Aalten, D.M., et al., *Structural insights into the catalytic mechanism of a family 18 exo- chitinase.* Proc Natl Acad Sci U S A, 2001. **98**(16): p. 8979-8984.

108. Phillips, D.C., Sci. Amer., 1966. **215**: p. 78-90.

109. Strynadka, N.C. and M.N. James, *Lysozyme revisited: crystallographic evidence for distortion of an N-acetylmuramic acid residue bound in site D.* J Mol Biol, 1991. **220**(2): p. 401-424.

110. Zechel, D.L. and S.G. Withers, *Glycosidase mechanisms: Anatomy of a finely tuned catalyst.* Acc Chem Res, 2000. **33**(1): p. 11-18.

111. Varrot, A., M. Schulein, and G.J. Davies, *Structural changes of the active site tunnel of Humicola insolens cellobiohydrolase, Cel6A, upon oligosaccharide binding.* Biochemistry, 1999. **38**(28): p. 8884-8891.

112. Teeri, T.T., et al., *Trichoderma reesei cellobiohydrolases: why so efficient on crystalline cellulose?* Biochemical Society Transactions, 1998. **26**(2): p. 173-178.

113. Deslongchamps, P., *Stereoelectronic Effects in Organic Chemistry*. 1983, New York: Pergamon Press.

114. Kirby, A.J., *Stereoelectronic Effects on Acetyl Hydrolysis*. Accounts of Chemical Research, 1984. **17**: p. 305-311.

115. Drouillard, S., et al., *Serratia marcescens chitobiase is a retaining glycosidase utilizing substrate acetamido group participation.* Biochem J, 1997. **328**(Pt 3): p. 945-949.

116. Lai, E.C.K. and S.G. Withers, *Stereochemistry and kinetics of the hydration of 2-acetamido-D-glucal by beta-N-acetylhexosaminidases.* Biochemistry, 1994. **33**(49): p. 14743-14749.

117. Legler, G. and R. Bollhagen, *(+/-)-6-Acetamido-1,2-Anhydro-6-Deoxy-Myo-Inositol: a tight-binding inhibitor and pseudosubstrate for N-Acetyl-Beta-Glucosaminidases.* Carbohydrate Research, 1992. **233**: p. 113-123.

118. Prag, G., et al., *Structures of chitobiase mutants complexed with the substrate Di-N-acetyl-d-glucosamine: the catalytic role of the conserved acidic pair, aspartate 539 and glutamate 540.* J Mol Biol, 2000. **300**(3): p. 611-617.

119. Pennybacker, M., et al., *Evidence for the involvement of Glu-355 in the catalytic action of human beta-hexosaminidase B.* J Biol Chem, 1997. **272**(12): p. 8002-8006.

120. Hou, Y., et al., *Characterization of Glu and Asp residues in the active site of human β-hexosaminidase B.* Biochemistry, 2001. **40**: p. 2201-2209.

121. Williams, S.J., et al., *Aspartate 313 in the Sterptomyces plicatus hexosaminidase plays a critical role in substrate assisted catalysis by orienting the 2-acetamido group and stabilizing the transition state.* J. Biol. Chem., 2002. **277**(42): p. 40055-65.

122. Mark, B.L., et al., *Biochemical and structural assessment of the 1-N-azasugar GalNAc-isofagomine as a potent family 20 β-N-acetylhexosaminidase inhibitor.* J. Biol. Chem., 2001. **276**(45): p. 42131-42137.

123. Mo, F., Jensen, L.H., Acta Crystallogr B, 1978. **34**: p. 1562.

124. Mark, B.L. and James, M.N.G., *Anchimeric assistance in hexosaminidases.* Can. J. Chem., 2002. **80**(8): p. 1064-1074.

Chapter 2

Crystallographic evidence for substrate-assisted catalysis in a bacterial β-hexosaminidase

A. Introduction

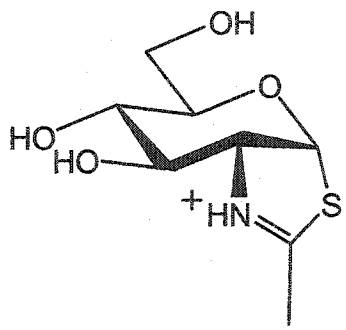
There are three intermediates along the reaction coordinate of family 20 glycosidases (Figure 1.2b, *lower pathway*) that can possibly be captured and studied crystallographically: the enzyme-substrate complex (Michaelis complex), the enzyme stabilized intermediate, and the product complex. The structure of the Michaelis complex for a family 20 glycosidase has already been determined crystallographically for *Sm*CHB (bound to chitobiose) [1]. As described in Chapter 1, section **D.3**, the complex revealed two crucial features about the catalytic mechanism of family 20 glycosidases: 1) The pyranoside conformation of the terminal GlcNAc to be removed (bound in the in the –1 site), is distorted from a relaxed $^4C_1$ chair into a strained $^4B^1$ boat conformation, resulting in the scissile bond and leaving group being moved into a pseudo-axial position with respect to the distorted terminal sugar, 2) *Sm*CHB lacks a suitably positioned enzymic nucleophile that would be required to form a glycosyl-enzyme intermediate. According to stereoelectronic theory [2, 3], the substrate distortion described above, reduces the activation barrier of the first step of the double displacement reaction by bringing atoms C5, O5, C1 and C2 of the terminal GlcNAc close to a co-planar conformation. This conformation allows for effective overlap of a non-bonding lone pair of electrons on O5 with the antibonding orbital at the electron-deficient anomeric center of the developing oxacarbenium ion transition state. In the absence of a suitably positioned enzymic nucleophile, the active site of SmCHB is instead designed to rotate and hold the C2-acetamido group of the terminal GlcNAc in such a way that the carbonyl oxygen atom of this functional group can participate in the reaction, and effectively substitute for the missing enzymic nucelophile.

The intramolecular assistance provided by the C2-acetamido group results in the formation of an enzyme stabilized oxazolinium ion intermediate (Figure

1.2b, *lower pathway*), which is the next step of the reaction pathway that can be studied crystallographically. However, the oxazoline ring of the bound intermediate is hydrolytically unstable, and certainly does not exist long enough to determine its molecular structure by X-ray crystallography. Fortunately, a relatively stable analogue of the cyclic intermediate, *N*-acetylglucosamine-thiazoline (NAG-thiazoline), has been synthesized and was shown to be a potent competitive inhibitor of jack bean β-hexosaminidase (K_i = 280 nM) (Figure 2.1a) [4] . It is also an excellent competitive inhibitor of both *Sp*HEX and human Hex B. Proof that this analogue is a good mimic of the oxazoline intermediate was demonstrated when a precursor substrate of NAG-thiazoline, 4-methylumbelliferyl-2-deoxy-2-*thio*acetamido-β-glucoside (Figure 2.1b), was shown to be converted into NAG-thiazoline by Jack bean β-hexosaminidase [4]. Conversion of the precursor by the enzyme correlated with a time dependent loss of activity, and provided biochemical evidence for enzyme catalyzed anchimeric assistance and the formation of the covalent, cyclized intermediate analogue NAG-thiazoline.

This chapter describes the determination of the three-dimensional crystal structure of *Sp*HEX alone and in complex with NAG-thiazoline, and appears in large part as a published paper in the *Journal of Biological Chemistry* [5]. *Sp*HEX (55 kDa) is a highly active and stable family 20 glycoside hydrolase that functions over a broad pH range [6]. Co-crystallization of *Sp*HEX with NAG-thiazoline, and the subsequent crystallographic analysis of the complex provides decisive structural evidence for a substrate-assisted catalytic mechanism involving C2-acetamido group participation and the formation of a covalent, cyclic intermediate. Comparison of this complex with the crystallographic Michaelis complex of *Sm*CHB bound to chitobiose [1] reveals an interesting conformational itinerary of the substrate that occurs along the reaction coordinate of family 20 glycosidases.

a



b

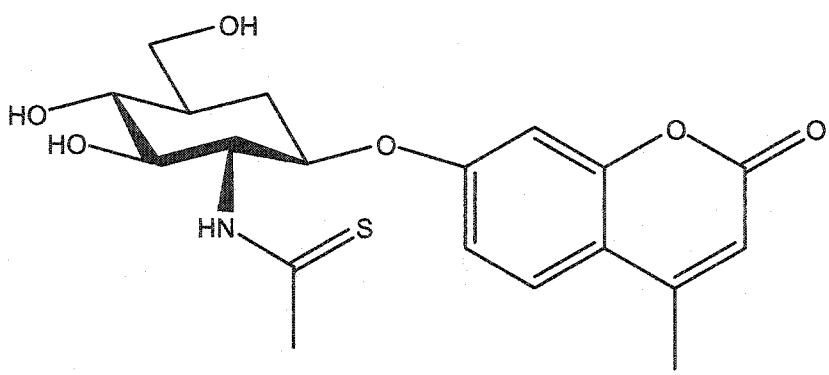


Figure 2.1. Chemical structures of the cyclic intermediate analogue NAG-thiazoline (a) and the precursor substrate of NAG-thiazoline, 4-methylumbelliferyl-2-deoxy-2-thioacetamido -β-glucoside [4]. Compound ***b*** can be enzymatically converted into compound **a** by a family 20 hexosaminidase.

B. Experimental Procedures

B.1 SpHEX expression and purification

Escherichia coli strain JM109 was used for plasmid amplification and plasmid purification was carried out using Qiagen purification systems. Restriction enzymes and Vent DNA polymerase were from New England Biolabs. T4 DNA ligase was from Boehringer Mannheim. All cloning procedures are described in [7]. *Sp*HEX is a 506 amino acid protein having a predicted molecular weight of 55010 Da (GenBank accession number: AF063001). It was expressed as a recombinant, N-terminal 7xHis tagged fusion protein. Briefly, the plasmid psHEX-1.8 [6] contained the *Sp*HEX open reading frame (ORF). The first 100 bp of the 5'-end of the *Sp*HEX ORF was amplified by the polymerase chain reaction (PCR) using the sense primer (5'-GGAATTCCATATGCATCATCATCATCATCATCACACCGGCGCCGCCCCG-GGAAG-3') and the antisense primer (5'-TGGCGCGCCGCCGGGGT-CGACCGAGGCGGG-3'). This PCR product was restriction digested with *Asc*I and *Nde*I for ligation into the final expression plasmid. To obtain the remaining 1.7 kbp fragment of the *Sp*HEX ORF, a further aliquot of psHEX-1.8 was restriction digested with *Asc*I and *BamH*I. The 100 bp (*Nde*I/*Asc*I) and 1.7 kbp (*Asc*I/*BamH*I) fragments were then ligated into the T7 expression plasmid pET-3a (Novagen) that had been linearized by digestion with *Nde*I and *BamH*I. The ligation product resulted in the expression plasmid p3AHEX-1.8 whose sequence was verified prior to use in fusion protein expression.

The 7xHis-*Sp*HEX fusion protein was expressed in *E. coli* strain BL21 (DE3). Transformed cells were grown at 37° C to an OD_{600} ~ 0.5 and then induced with 0.4 mM IPTG for 3 hrs at 25°C. Cells were pelleted by centrifugation, resuspended in a lysis buffer (20 mM Tris-Cl, pH 8.0, 300 mM NaCl, 10 mM imidazole, 10 mM β-mercaptoethanol) and lysed by French press. After centrifugation at 20000 x g for 1 hour, the supernatant was loaded onto a Ni-NTA superflow (Qiagen) column pre-equilibrated with lysis buffer (Figure 2.2a). Once loaded, the column was washed with the lysis buffer supplemented with 80 mM imidazole (pH 8.0). The fusion protein was eluted from the column

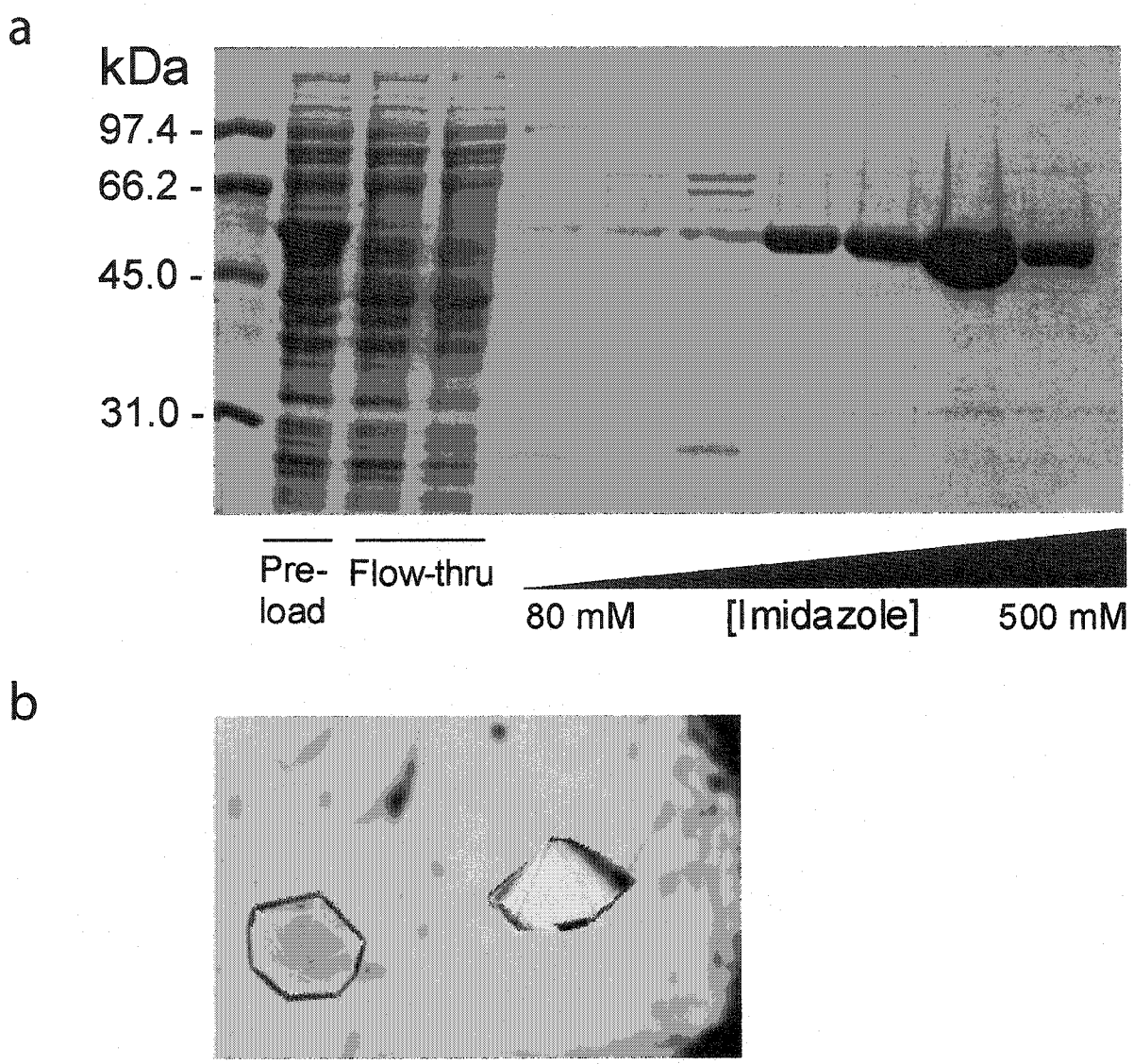


Figure 2.2. Purification and crystallization of recombinant 7xHis-*Sp*HEX
Panel a, SDS-PAGE gel of protein fraction collected during the Ni-NTA affinity purification procedure. *Panel b*, photograph of the hexagonal bipyramidal crystals of recombinant 7xHix-SpHEX grown by the hanging drop method, using mother liquor consisting of 2.2 M ammonium sulphate, 100 mM tri-sodium citrate, pH 6.0 and 20-25% glycerol.

using lysis buffer supplemented with 250 mM imidazole (pH 8.0) and precipitated with 55% ammonium sulfate for storage at 4°C. Aliquots of the precipitated protein were routinely resuspended and dialyzed twice against 50 mM tri-sodium citrate, pH 6.0, 300 mM NaCl and 0.5 mM DTT and then concentrated to approximately 10 mg/ml with a Millipore concentrator. Approximately 40-60 mg of pure fusion protein was routinely obtained per liter of culture. Electrospray ionization mass spectrometric analysis, using a VG Quattro triple quadrupole mass spectrometer (VG Biotech, Altringham, UK), determined the mass of the purified fusion protein to be 56054 Da, in good agreement with the theoretical mass of 56049 Da.

Seleno-Met substituted 7xHis-*Sp*HEX was expressed in *E. coli* strain BL21 (DE3) pLys S using the method of [8]. Transformed cells were grown at 37 °C in M9 minimal media until mid-log phase growth was reached. The culture was then supplemented with 0.5 mM Lys, 0.8 mM Thr, 0.6 mM Phe, 0.8 mM Leu, 0.8 mM Ile and 0.8 mM Val to inhibit endogenous Met biosynthesis. After a 30 min. incubation, the culture was further supplemented with 0.25 mM seleno-Met and induced with 0.5 mM IPTG for 10 hours. Seleno-Met substituted 7xHis-*Sp*HEX was purified in the same manner as native 7xHis-*Sp*HEX except that the protein was dialyzed against 3 mM DTT before concentrating to avoid selenium oxidation. Electrospray mass spectrometric analysis verified that all 6 Met residues in the 512 amino acid *Sp*HEX protein had been substituted with seleno-Met. All purified fusion protein was visualized for purity by SDS-PAGE (Figure 2.2).

B.2. Crystallization and data collection

Both native and seleno-Met substituted 7xHis-*Sp*HEX crystallized in the hexagonal space group $P6_122$ within two weeks by vapor diffusion at room temperature (Figure2.2b). The mother liquor consisted of 2.2 M ammonium sulphate, 100 mM tri-sodium citrate, pH 6.0 and 20-25% glycerol. Hanging drops were set up by mixing an aliquot of *Sp*HEX (concentrated to 10 mg/ml) with an equal amount of the mother liquor. Crystals of 7xHis-*Sp*HEX in complex with

NAG-thiazoline were obtained by co-crystallization of the native fusion protein (from which DTT had been removed by dialysis) with 2-5 mM NAG-thiazoline. Diffraction data for a MAD-phasing experiment were collected at the Advanced Photon Source, BioCARS sector beamline BM-14-C and BM-14-D on native and seleno-Met substituted 7xHis-SpHEX crystals flash cooled to 100K, respectively (Table 2.1). Diffraction data from crystals of the complex between 7xHis-*Sp*HEX and NAG-thiazoline were collected at SSRL, beamline 9-2 (Table 2.1). All diffraction data were processed using DENZO and SCALEPACK [9].

B.3. Structure determination and refinement

A solution to the crystal structure of the protein was obtained by a MAD-phasing experiment performed on seleno-Met-substituted protein crystals [10]. A combination of data derived from the MAD-phasing experiment at beamline BM-14-D with data collected from native *Sp*HEX crystals at beamline 14-BM-C allowed for the determination of the three dimensional structure of *S. plicatus* β-hexosaminidase to 2.2 Å resolution. Although the *Sp*HEX crystals diffracted to slightly higher resolution than 2.2 Å, data collection was restricted to this resolution in order to avoid excessive data rejection due to spot overlap. The program SOLVE [11] was used for local scaling of the data and to calculate the anomalous and dispersive differences needed to find Se sites and to determine phase probability distributions. Patterson maps, calculated from the anomalous and dispersive differences, allowed us to find clearly five of the six Se atoms present in the *Sp*HEX structure. The missing Se atom was part of the initiation Met whose position could not be determined due to disorder of the first 14 residues of the 7xHis-tagged N-terminus.

Electron density maps, generated using structure factor phases obtained from the MAD-phasing experiment (initial figure of merit 0.8), were improved only slightly by solvent flattening using DM (Figure 2.3) [12]. Map boundaries were extended beyond the CCP4 asymmetric unit using EXTEND [13] and skeletonized using MAPMAN [14]. A molecular model of the enzyme was built from the skeletonized map using O [15]. Residues 8 through 512 were readily fit

Table 2.1 Crystallographic statistics for *Sp*HEX

	SpHEX		Se-Met-SpHEX		SpHEX:NGT
Crystal Information					
Space group	$P6_122$		$P6_122$		$P6_122$
Unit cell dimensions (Å)	a = b = 133.1, c = 176.8		a = b = 132.9, c = 177.1		a = b = 133.6, c = 174.2
Data Collection (values in parentheses refer to the high-resolution shell)					
Data set	Monochomatic	Edge	Peak	Remote	Monochomatic
Detector	ADSC Q4	ADSC Q4	ADSC Q4	ADSC Q4	ADSC Q4
Wavelength (Å)	1.00	0.9797	0.9795	0.9496	1.03
Resolution (Å)	40 - 2.20	30 - 2.20	30 - 2.20	30 - 2.20	40 - 2.10
High-resolution (Å)	2.26 - 2.20	2.28 - 2.20	2.28 - 2.20	2.28 - 2.20	2.16 - 2.10
Total observations	416806	432049	432434	431524	361546
Unique reflections	47455 (3854)	47697 (4680)	47690 (4675)	47648 (4667)	52266 (4059)
$< I / \sigma >$	56.1 (21.6)	34.9 (19.5)	35.9 (21.1)	39.2 (26.5)	26.4 (6.64)
Completeness[a] (%)	99.8 (99.3)	100.0	100.0	100.0	96.7 (92.0)
R_{sym}[b]	0.031 (0.067)	0.061 (0.126)	0.056 (0.124)	0.044 (0.091)	0.084 (0.263)
R_{anom}[c]		0.053 (0.072)	0.048 (0.068)	0.036 (0.051)	
Refinement					
Resolution (Å)	40 - 2.2				40 - 2.1
R_{work}[d]	0.18				0.20
R_{free}[d]	0.21				0.22
Number of atoms					
protein	3864				3864
NGT					14
water	382				270
Average B ($Å^2$)	17.2				28.7
Average NGT B ($Å^2$)					20.42
RMSD from ideal geometry					
Bond lengths (Å)	0.0052				0.0052
bond angles (°)	1.29				1.31
Ramachandran plot					
%-most favoured[e]	89.4				88.9
%-additionally allowed[e]	10.6				11.1

[a]Calculated by treating Bijvoet pairs as equivalent.
[b]$R_{sym} = \Sigma_h \ \Sigma_i \ (|I_i(\mathbf{h}) - \langle I(\mathbf{h})\rangle|) \ / \ \Sigma_h \ \Sigma_i \ I_i(\mathbf{h})$, where $I_i(\mathbf{h})$ is the ith intensity measurement and $\langle I(\mathbf{h})\rangle$ is the weighted mean of all measurements of $I(\mathbf{h})$.
[c]$R_{anom} = \Sigma_h \ (|I^+(\mathbf{h}) - I^-(\mathbf{h})|) \ / \ \Sigma_h \ (I^+(\mathbf{h}) + I^-(\mathbf{h}))$, where $I^+(\mathbf{h})$ and $I^-(\mathbf{h})$ are the Bijvoet pairs of $I(\mathbf{h})$.
[d]R_{work} and $R_{free} = \Sigma_h \ ||F(\mathbf{h})_{obs}| - |F(\mathbf{h})_{calc}|| \ / \ |F(\mathbf{h})_{obs}|$ for reflections in the working and test sets (10% of all data) respectively.
[e]Regions defined by PROCHECK (37)
NGT = NAG-thiazoline

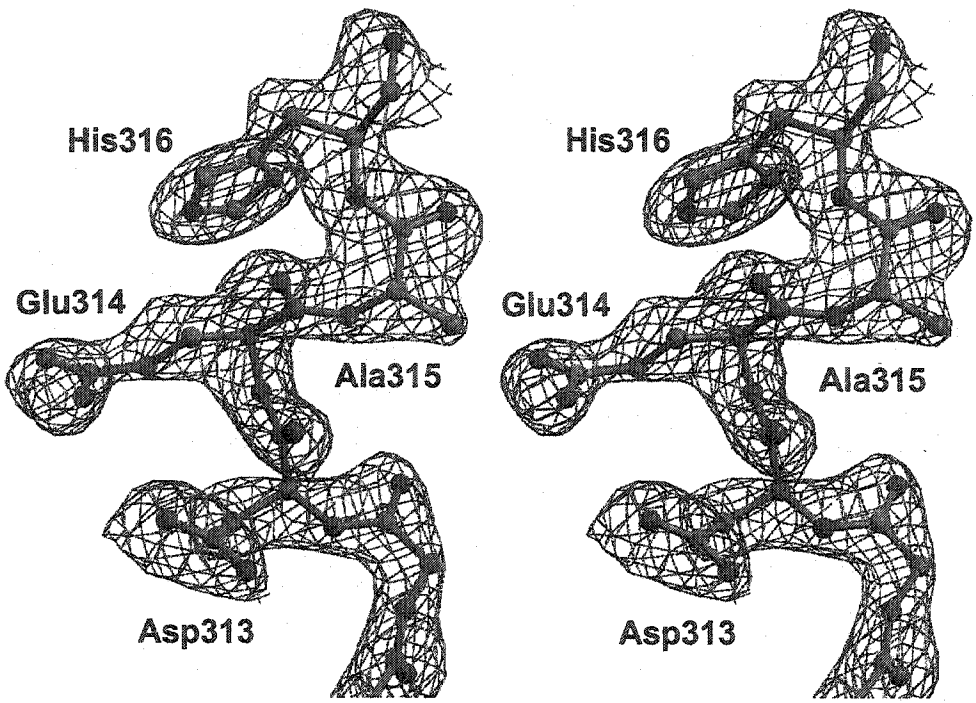


Figure 2.3. Stereo view showing the experimental electron density for a region of the *Sp*HEX active site. The map is contoured at 1.4σ and has a d_{min} of 2.2 Angstroms. The map was computed using $m|F_O|\exp(\alpha_{MAD/DM})$ coefficients. The refined model of SpHEX is drawn as sticks with carbon atoms in *purple*, nitrogen in *blue* and oxygen in *red*.

into the density as one continuous chain. Coordinates for the small molecules glycerol and SO_4^- were obtained through the HIC-UP world wide web site and their geometries were optimized by X-PLOR [16] prior to use in model building.

The molecular model of native *Sp*HEX was refined using a maximum likelihood target function during both simulated annealing and conjugate gradient minimization as implemented in CNS [17]. Prior to refinement, 10% of the diffraction data were randomly flagged for cross validation using the free *R*-factor. After each round of refinement, the model was manually inspected with O using $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ maps. The final refinement statistics for the model reflect the high quality data (Table 2.1).

B.4. NAG-thiazoline complex

An $|F_o|-|F_c|$ map, used to visualize NAG-thiazoline in the active site, was obtained using structure factor phases calculated from the native *Sp*HEX model that had been positioned into the unit cell of the NAG-thiazoline complex using rigid body refinement followed by conjugate gradient minimization. Solvent molecules were removed from the model before placing it into the new cell and were relocated during later rounds of refinement. Any waters found in the active site were deleted until the NAG-thiazoline had been modeled into the electron density ascribed to it. The initial geometry of the NAG-thiazoline model was based on the X-ray crystal structure of *N*-acetylgalactosamine-thiazoline (GalNAc-thiazoline) (coordinates provided by D. Vocadlo, unpublished data). Refinement of the NAG-thiazoline complex was carried out using CNS as described for the native *Sp*HEX model above. The final refinement statistics are presented in Table 2.1.

B.5. Coordinates

The coordinates and structure factors have been deposited into the Protein Data Bank (PDB): native *Sp*HEX PDB code 1HP4; *Sp*HEX·NAG-thiazoline complex PDB code 1HP5.

C. Results and Discussion

C.1. Structure of β-hexosaminidase from S. plicatus

Excellent crystallographic data (Table 2.1) produced easily interpretable electron density maps into which a model of *Sp*HEX was built (Figure 2.3). The enzyme is a kidney shaped, two domain protein having overall dimensions of ~68 x 58 x 56 Å (Figure 2.4). The two domains of *Sp*HEX have a similar fold to domains II (residues 214-335) and III (residues 336-818) of *Sm*CHB (Figure 2.5); however, significant deviations between the two structures exist. The most striking structural difference between *Sp*HEX and *Sm*CHB is the absence in *Sp*HEX of two of the four domains that compose *Sm*CHB (Figure 2.5). This results in a solvent-exposed active site at the carboxy-terminal end of the $(\beta/\alpha)_8$-barrel forming domain II. Such a solvent-exposed active site appears to explain why β-hexosaminidases, such as human β-hexosaminidase A, can accommodate large glycoconjugates like G_{M2}-ganglioside (Figure 1.6).

Domain I of *Sp*HEX is composed of residues 1 through 151. As in *Sm*CHB, this domain has an α/β topology consisting of a solvent exposed, seven-stranded anti-parallel β-sheet that buries two, roughly parallel, α-helices (Figure 2.4). Similar topologies have been found in matrix metalloproteinases [1] and collagenases. The amino acid sequence identity between *Sp*HEX and *Sm*CHB is lowest throughout this domain. A structure-based alignment using SwissPDBviewer [18] indicated only a 16.1% amino acid identity. Nonetheless, the fold is well conserved, with 87 C^{α} atoms of the two homologous domains having a rms difference of only 1.34 Å. A multiple sequence alignment of all family 20 glycoside hydrolases indicates that domain I is conserved throughout the entire family. Such conservation suggests a functional requirement for this domain by family 20 glycoside hydrolases; ironically however, its function remains unknown.

Domain II of *Sp*HEX is composed of residues 151 through 512 and is folded into a $(\beta/\alpha)_8$-barrel with the active site of the enzyme residing at the C-termini of the 8 β-strands of the barrel. This domain is homologous to domain III

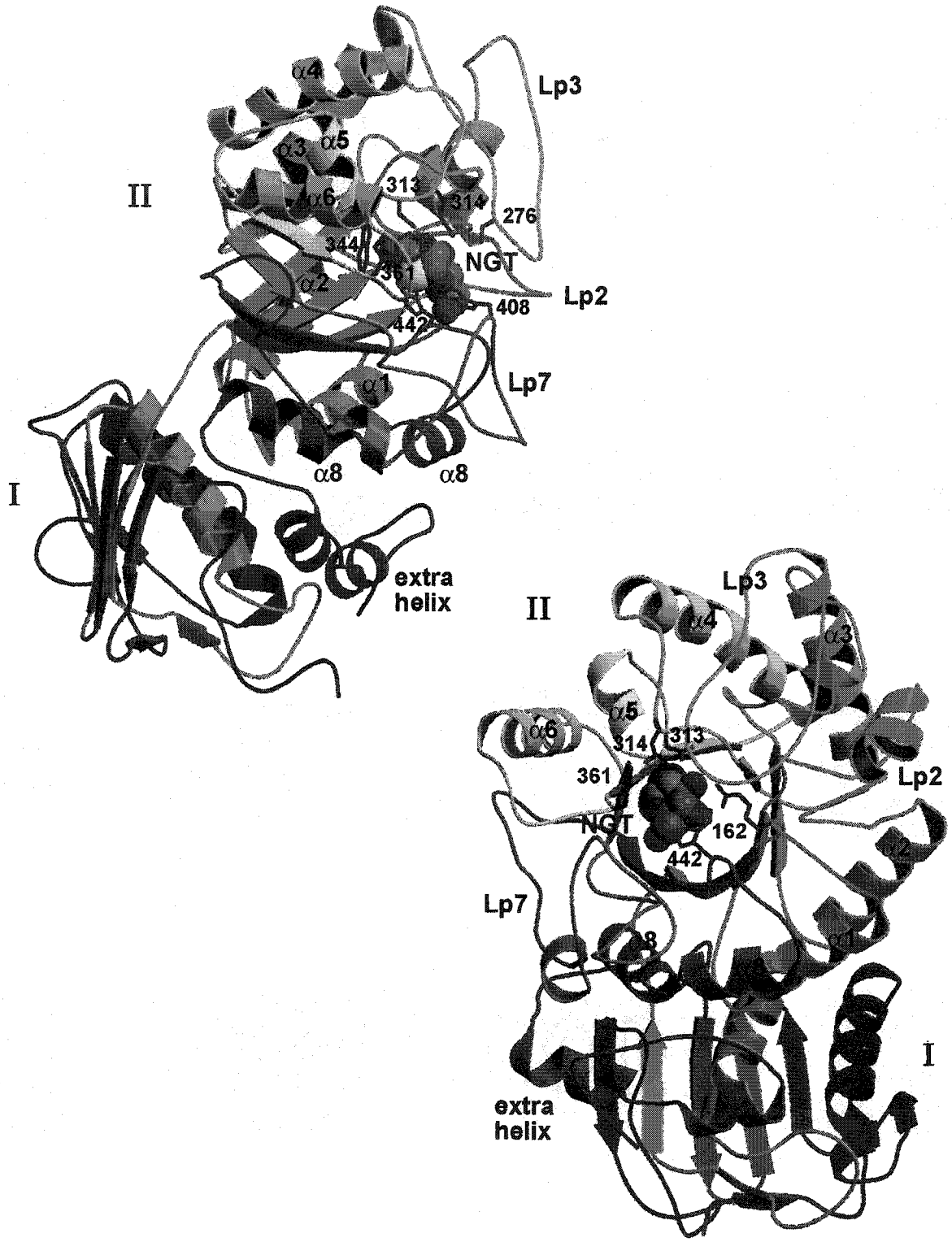


Figure 2.4. Ribbon diagram of SpHEX in complex with NAG-thiazoline (NGT). Colouring proceeds from blue (N-terminus) to red (C-terminus). *Top*, Side view of the complex. The α-helices of the (α/β)8 barrel composing domain II are labeled α1 to α8 (helix α7 is replaced by an extended loop, Lp7). Residues of the active site are shown in magenta and coloured. NGT is shown as a space filling model. Loops comprising the sugar binding subsite +1 are labeled Lp2, Lp3 and Lp7. *Bottom*, Front view of the complex.

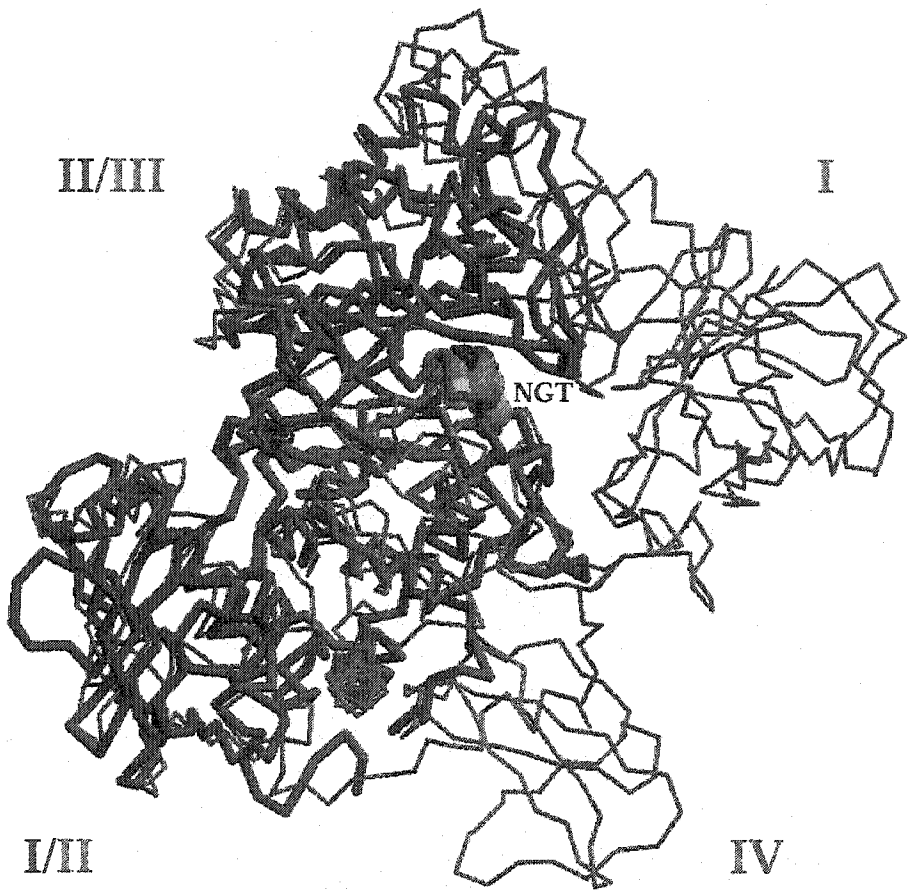


Figure 2.5. Superposition of Cα traces for *Sp*HEX (red) and *Sm*CHB (blue) using the program O [15]. Domain numbers for SpHEX are in red and the corresponding domain numbers for *Sm*CHB in blue. Domains II and III of *Sm*CHB are homologous to domains I and II of *Sp*HEX. NAG-thiazoline (NGT) bound within the *Sp*HEX active site is shown as a space filling model. The overall rms difference for the superposition was found to be 1.67 Angstroms for 389 Cα atoms.

in *Sm*CHB and a structure-based sequence alignment demonstrated there to be a 29.5% sequence identity between the two domains, where 236 of the C^{α} atoms had a rms difference of 1.30 Å. What may indeed be a common feature of this $(\beta/\alpha)_8$-barrel domain in family 20 glycoside hydrolases is the conspicuous absence of regular helices at positions α5 and α7 in the $(\beta/\alpha)_8$-barrel. In both *Sp*HEX and *Sm*CHB helix α5 consists of only a single turn of a 3/10 helix, whereas helix α7 is completely absent and is instead replaced by an extended loop. Overall, this domain in *Sp*HEX contains shorter surface loops and is much more compact than its homologous counterpart, domain III, in *Sm*CHB. Multiple sequence alignments of family 20 glycoside hydrolases suggest that such a compact $(\beta/\alpha)_8$-barrel domain may be a common feature among many family 20 β-hexosaminidases, including the human isoforms [1, 6].

Unlike the basic $(\beta/\alpha)_8$-barrel motif, domain II of *Sp*HEX contains three major loop structures that extend from the C-termini of three of the 8 β-strands of the barrel. First, loop Lp7 replaces helix α7 as described above (Figure 2.4). Second, a 36 amino acid loop, Lp2, extends from the C-terminus of strand β2 and contains a short helical segment that packs against, and stabilizes, the third major loop Lp3. Lp3 is a 41 amino acid loop that extends from the C-terminus of strand β3 and contains a helical segment that is complimentary to, and packs against, the helical segment found in Lp2 (Figure 2.4). There is only one disulphide bond in *Sp*HEX (Cys263-Cys282) and its presence close to the base of Lp3, may help to stabilize the conformation of this loop. Lp3 and Lp7 act in concert to form the hydrophobic faces of sugar binding site +1 described below (Figure 2.6). There are two homologous loops in *Sm*CHB; however, they are longer and perform an additional function by interacting with a domain not present in *Sp*HEX (*Sm*CHB domain I) (Figure 2.5). Finally, an extra helix continues on from helix α8 of the $(\beta/\alpha)_8$-barrel to complete the C-terminus of *Sp*HEX. This extra helix stabilizes domains I and II with respect to each other (Figure 2.4). It is interesting to observe that the relative orientation of domains I

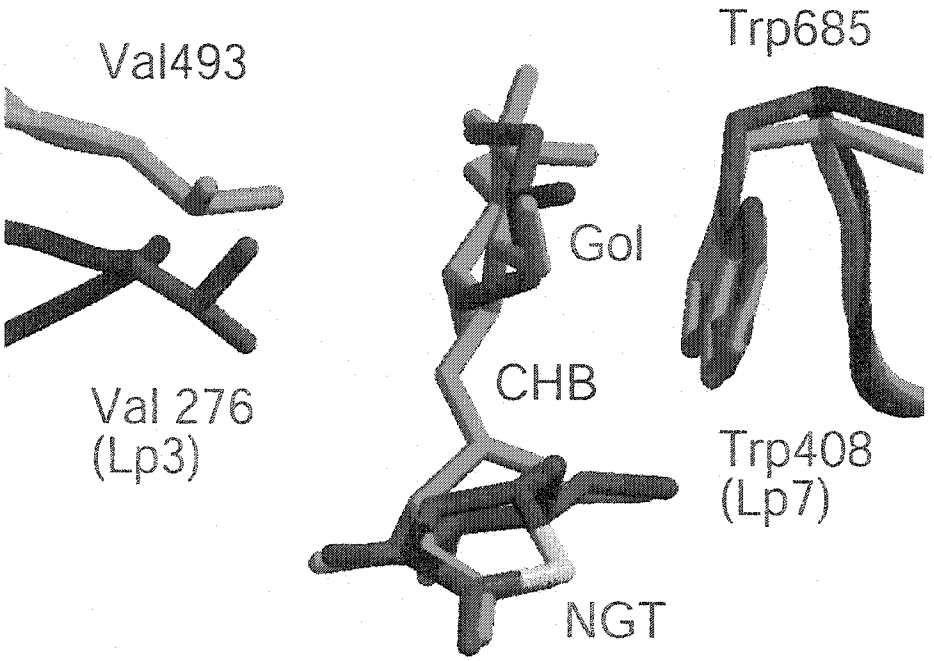


Figure 2.6. Superposition between *Sp*HEX (*blue*) and *Sm*CHB (*green*). The relative positions of residues forming the hydrophobic faces of sugar binding subsite +1, and the position of chitobiose (CHB) (as bound to *Sm*CHB [1]) relative to NAG-thiazoline (NGT) and glycerol (Gol) bound to *Sp*HEX subsite -1 and +1 is shown. Superposition was carried out in O [15] using *Sp*HEX active site residues Trp344, 361 and 442 and *Sm*CHB residues Trp616, 639 and 737. Val276 (*Sp*HEX) and Val493 (*Sm*CHB) are on loop Lp3 whereas Trp408 (*Sp*HEX) and Trp685 (*Sm*CHB) are on loop Lp7 and compose the hydrophobic sides of the +1 subsite. The non-reducing sugar of chitobiose superposes onto NAG-thiazoline bound to subsite -1. The reducing sugar of chitobiose, bound to subsite +1 and rotated 90° relative to the non-reducing sugar, superposes onto the glycerol bound to subsite +1 in *Sp*HEX.

and II of *Sp*HEX is the same as the homologous domains II and III in *Sm*CHB (Figure 2.5).

C.2. The complex with NAG-thiazoline: mechanistic implications

According to the X-ray structure of *Sp*HEX and that of the *Sm*CHB/chitobiose complex [1], family 20 glycoside hydrolases do not appear to contain a side chain in a position suitable to act as a catalytic nucleophile that would stabilize developing oxacarbenium ion character. Instead, it has been observed that, in the conformation bound by the enzyme, the C2 acetamido oxygen of the non-reducing sugar in subsite –1 is held within 3 Å of its C1 anomeric carbon. When in this position, it is believed that the acetamido oxygen can act as a nucleophile and attack the anomeric center to form a cyclic NAG-oxazolinium ion intermediate [1].

Figure 2.7 shows NAG-thiazoline bound in the *Sp*HEX active site and the quality of the electron density into which it was modeled. Excluding O4 and O6 due to differences in C4 chirality and enzyme packing effects respectively, the remaining atoms in NAG-thiazoline had an rms difference of only 0.071 Å compared to the equivalent atoms in the small molecule structure of GalNAc-thiazoline. NAG-thiazoline was bound in the -1 subsite of *Sp*HEX and adopts a conformation that is close to a $^4C_1$ chair, although the data do not exclude small distortions towards a sofa or skew boat conformation. There are no significant changes in the *Sp*HEX structure upon binding NAG-thiazoline except for a slight opening of the active site pocket. Figure 2.8 clearly shows Trp residues 344, 361, 442 of the -1 subsite of *Sp*HEX and the homologous residues in *Sm*CHB (Trp 616, Trp 639, and Trp 737) forming a tight hydrophobic pocket into which the non-reducing GlcNAc residue binds. This pocket appears to force the C2-acetamido oxygen atom into close proximity with the anomeric carbon and the tight packing between the acetamido group and the enzyme helps to ensure a precise alignment of the acetamido oxygen with the anomeric carbon.

Numerous hydrogen bonding interactions lock NAG-thiazoline into the active site of *Sp*HEX and disperse the positive charge distributed into the

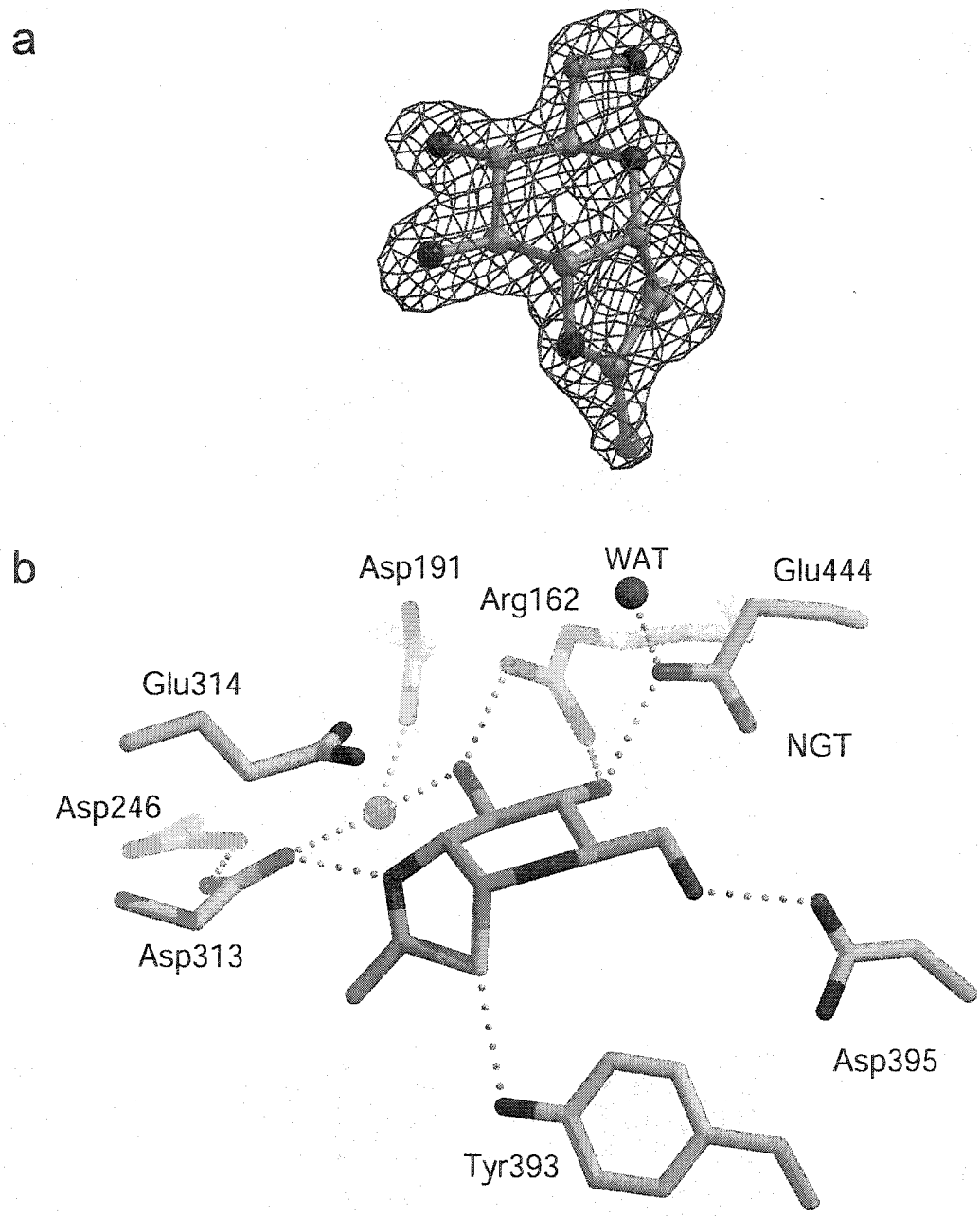


Figure 2.7. NAG-thiazoline bound to *Sp*HEX at 2.1 Angstrom resolution. a, Electron density for NAG-thiazoline (contoured at 2.8 σ). The refined model is drawn as sticks with carbon atoms in *grey*, nitrogen atoms in *blue*, oxygen atoms in *red* and the sulphur atom in *yellow*. The map was calculated as a $|F_O|$-$|F_C|$ simulated annealing omit map as implemented in CNS [16]. b, The *Sp*HEX active site architecture showing hydrogen bonding interactions with NAG-thiazoline (NGT). NAG-thiazoline is in a $^4C_1$ conformation. Asp313 and Tyr393 are primarily responsible for stabilizing the oxazoline ring during catalysis whereas Arg 162, Asp191, Glu444, Asp395 and Trp 408 lock the pyranose ring of the non-reducing sugar into place within the active site. WAT indicates the conserved incoming water molecule proposed by Tews *et al.* [1].

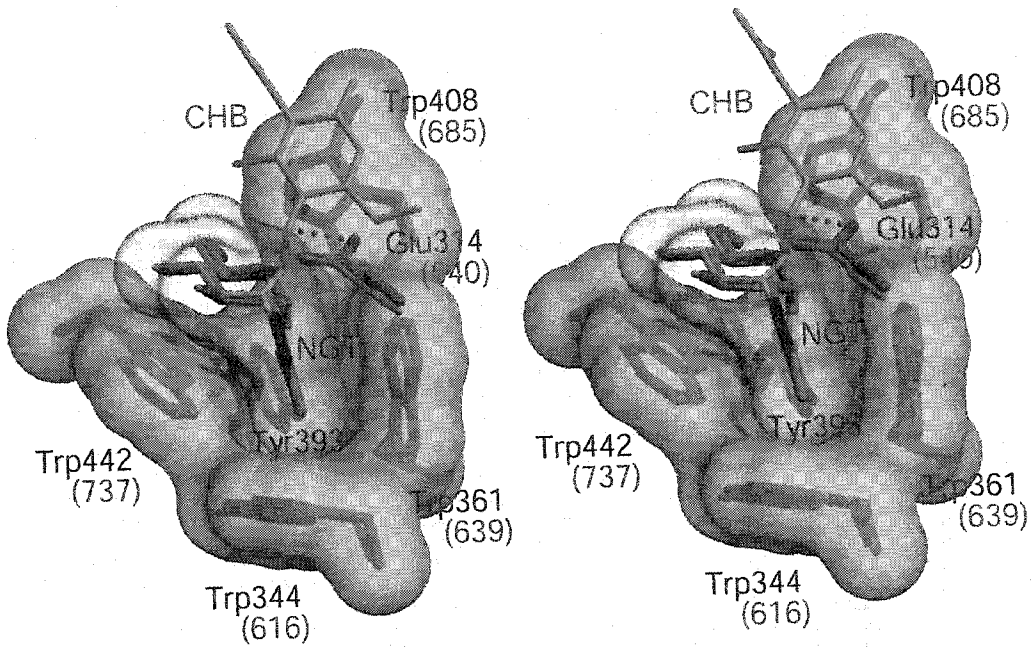


Figure 2.8. Superposition of hydrophobic residues forming the sugar binding subsite -1 of *Sp*HEX (*blue*) and *Sm*CHB (*green*). Superposition was carried out in O [15] using *Sp*HEX active site resides Trp344, 361 and 442 and *Sm*CHB residues Trp616, 639 and 737. Semi-transparent, solvent accessible surfaces have been drawn around the hydrophobic residues using GRASP [36]. The relative positions of NAG-thiazoline (NGT), chitobiose (CHB) (as bound to *Sm*CHB [ref]) and the general acid/base from each enzyme is shown. The proposed general acid/base residue of *Sp*HEX (Glu314) and *Sm*CHB (Glu540) superpose and Glu 314 of *Sp*HEX makes a hydrogen bond to the glycosidic oxygen of the chitobiose model.

thiazoline ring upon cyclization (Figure 2.7). These include at least one hydrogen bond to every hydroxyl group on the pyranose ring. However, no hydrogen bonds to the ring oxygen O5 are evident; indeed, a hydrogen bond to O5 would be counter-catalytic since it would decrease the extent of lone pair donation by O5 to the antibonding orbital of the scissile bond [3, 19].

NAG-thiazoline is held in place particularly strongly by Arg162, which forms hydrogen bonds to both O3 and O4 of the inhibitor. As disscused in Chapter 1, section **D.3.3**, the mutation Arg162His results in a 40-fold increase in K_m relative to wild type *Sp*HEX and a 5-fold decrease in V_{max} when assayed using 4-MUG [6], and confirms that this residue is involved in substrate binding. The analogous mutation in the α-subunit of human HexA (Arg178His) is associated with the B1 variant form of Tay-Sachs disease in which the enzyme appears to be normally folded and processed but lacks sufficient enzymatic activity, thus resulting in Tay-Sachs disease [20, 21] (see Chapter 1, section **D.1**). Such bidentate hydrogen bonding from an Arg side chain to two vicinal hydroxyl groups on the substrate is also seen between the non-reducing sugar of chitobiose and Arg349 of *Sm*CHB [1].

Two particularly important hydrogen-bonding interactions are formed with the thiazoline ring of NAG-thiazoline when it binds to *Sp*HEX. First, the OH of Tyr393 donates a hydrogen bond to the sulphur atom of the thiazoline ring. In the substrate complex such a hydrogen bond would orient the carbonyl oxygen into position for nucleophilic attack on the anomeric carbon C1. A similar role is envisioned for Tyr669 of *Sm*CHB [1]. Second, upon formation of the cyclic intermediate, the nitrogen atom N2 develops a positive charge and *Sp*HEX appears to stabilize this positive charge by delocalizing it through a hydrogen-bonding network between Asp313, Asp246 and the main-chain NH group of Met247. This is seen in the two short hydrogen bonds of 2.5 and 2.4 Å from the nitrogen N2 of the thiazoline ring and the carboxylate oxygens of Asp313 and Asp 246, respectively (Figure 2.7). These short hydrogen bond distances indicate that the carboxylate of Asp313 is likely deprotonated and possesses a

delocalized negative charge during catalysis. The role of Asp313 in the catalytic mechanism of *Sp*HEX is investigated further in Chapter 4.

The other key residue in the active site of retaining glycosidases is the acid/base catalyst, which adopts a dual role, functioning as a general acid to protonate the departing aglycone in the first step, then as a general base to deprotonate the incoming water in the second step. As discussed in Chapter 1, the structure of the complex of *Sm*CHB with chitobiose revealed a 2.9 Å hydrogen bond between the glycosidic oxygen of chitobiose and Glu540, leading to the assignment of Glu540 in *Sm*CHB as the acid catalyst [1]. Comparative molecular modeling combined with site directed mutagenesis and kinetic studies of *Sp*HEX and human β-hexosaminidase subunits α and β have shown Glu314, Glu323 and Glu355 to be homologous to *Sm*CHB Glu540, respectively [1, 6, 22, 23]. Superposition of the crystal structures of *Sp*HEX and *Sm*CHB confirms that Glu314 of *Sp*HEX is indeed positioned within the active site such that it too would make a hydrogen bond to the glycosidic oxygen of the superimposed chitobiose model (Figure 2.8).

The second and final step in the double displacement mechanism is the hydrolysis of the intermediate by general base-catalyzed attack of water at the anomeric center C1, resulting in overall retention of the anomeric configuration. Figures 2.6 and 2.9 show the position of a glycerol molecule bound in the +1 subsite. This glycerol superimposes onto half of the pyranose ring of chitobiose and suggests that subsite +1 in *Sp*HEX causes the sugar in this subsite to be twisted approximately 90° relative to the sugar bound in subsite –1 (Figure 2.6). Furthermore, one of the hydroxyl groups of this glycerol is within 3.4 Å of the anomeric C1 of NAG-thiazoline and forms a hydrogen-bonding interaction (2.7 Å) with the carboxylate of the general acid/base Glu314. It is postulated that this hydroxyl group occupies the position that an incoming water molecule would take to nucleophilically attack C1, thereby hydrolyzing the oxazolinium ion intermediate, with release of β-N-acetylglucosamine. Abstraction of the proton from water by Glu314 is assisted by a hydrogen-bonding network formed between its carboxylate group, the imidazole nitrogens of His 250, the

carboxylate of Asp 191 and the main-chain NH group of Asp 192 (Figure 2.9). The active site water molecule seen in the *Sm*CHB structure and proposed to be the reactant species [1] is indeed conserved in the *Sp*HEX structure and is indicated in Figures 2.7 and 2.9 as WAT. However, this water molecule is buried within the active site of both structures, and it seems more plausible that the incoming water enters directly from the bulk solvent after departure of the aglycone rather than occupying this site first.

There is convincing crystallographic evidence for the formation of an oxazolinium ion intermediate on the catalytic pathway of family 18 glycosidases. The potent family 18 chitinase inhibitor, allosamidin, is a natural pseudotrisaccharide with an aminocyclitol derivative at its 'non-reducing' end that mimics the stereochemical characteristics of the oxazolinium ion intermediate [24]. The structure of this inhibitor has been determined in complex with the chitinase/lysozyme hevamine [25], and Chitinase B (ChiB) from *S. marcescens* [26]. A superposition of the *Sp*HEX NAG-thiazoline and hevamine allosamidin complexes demonstrates that the oxazoline-like structure of allosamidin not only occupies the same location as the thiazoline ring of NAG-thiazoline, but also it forms the same hydrogen-bonding interactions with equivalent Asp and Tyr residues (Figure 2.10). A similar observation is found when the *S. marcescens* ChiB allosamidin and NAG-thiazoline β-hexosaminidase complexes are superimposed. Although the structure of a complex between a family 56 hyaluronidase and a cyclic intermedate analogue has not been determined, it has recently been demonstrated that a GlcA-β(1→3)-NAG oxazoline derivative can be used as an 'activated' substrate monomer for the artificial synthesis of hyaluronic acid by either bovine or ovine testicular hyaluronidase [27].

Finally, a well ordered water molecule can be observed approximately 3.3 to 3.4 Å from C1 of the allosamizoline ring when in complex with either hevamine or ChiB from *S. marcescens* [26]. This water molecule is hydrogen-bonded to the respective general acid/base residue of each enzyme (ChiB Glu 144, hevamine Glu 127) and is in roughly the same position as the glycosidic oxygen atom prior to cleavage [26]. If an oxazolinium ion intermediate were to be bound

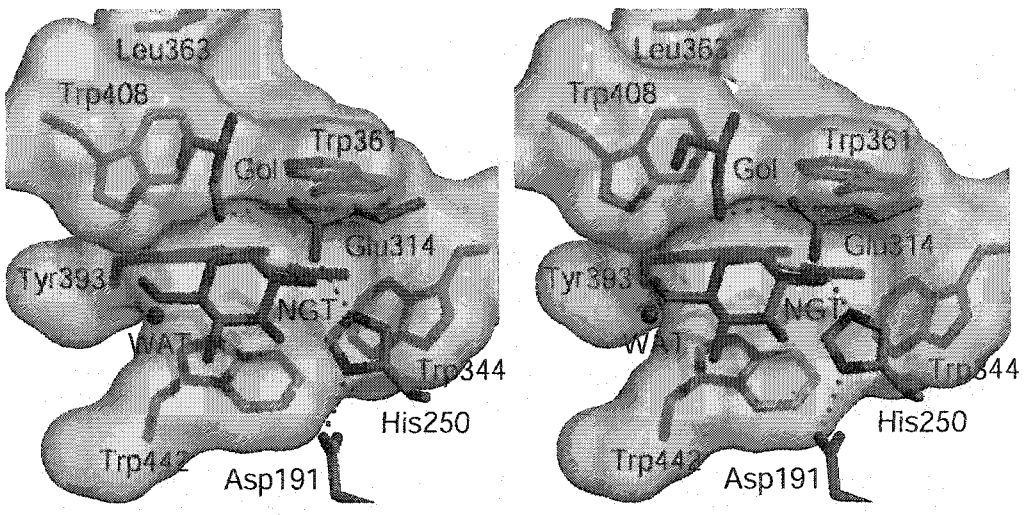


Figure 2.9. NAG-thiazoline (NGT) and glycerol (Gol) bound to sugar binding subsites -1 and +1 of *Sp*HEX respectively. Semi-transparent surfaces have been drawn around hydrophobic residues using GRASP [36]. The catalytic triad (Glu314, His250 and Asp191) has been drawn along with its hydrogen-bonding network. The glycerol hydroxyl group hydrogen bonding to the carboxylate of Glu314 is believed to occupy the position that an incoming water molecule would take to nucleophilically attack C1. WAT indicates the conserved incoming water molecule proposed by Tews *et al* [1].

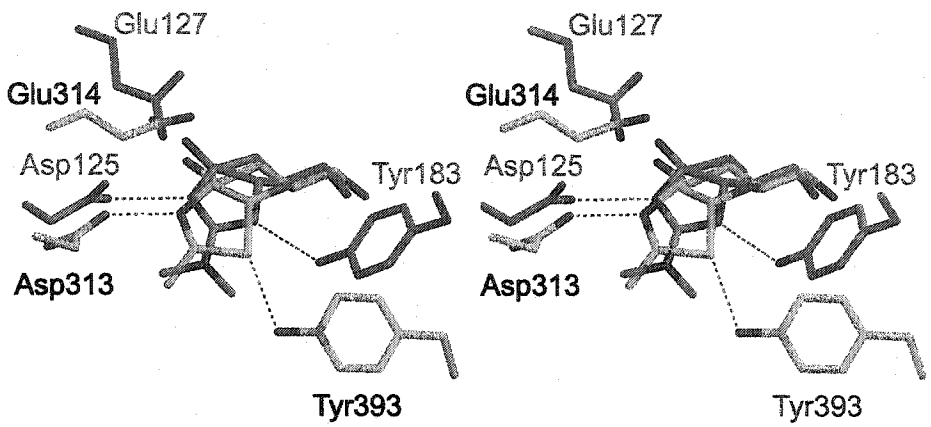


Figure 2.10. Superposition of the *Sp*HEX·NAG-thiazoline complex (grey and purple carbon atoms) and family 18 hevamine·allosamidin complex [25] (green carbon atoms). The oxazoline-like moiety of the naturally occurring chitinase inhibitor allosamidin and the thiazoline ring of NAG-thiazoline occupy the same relative position within the active site of each enzyme. The striking similarity of the active site architecture and hydrogen-bonding interactions responsible for binding the oxazoline-like moieties of allosamidin to hevamine and NAG-thiazoline to *Sp*HEX, provides convincing evidence for the formation of a cyclic intermediate in the reaction pathways of both family 18 and 20 glycosidases.

in the position of the allosamizoline ring in these complexes, this water molecule would be perfectly poised for base-catalyzed nucleophilic attack at the anomeric center of the intermediate, completing the catalytic cycle and producing a product with retained anomeric stereochemistry [26].

C.3. Conclusions

A combination of the results from this study, in which the structure of a complex with an intermediate analogue is presented, with those from a previous study of the structure of the substrate (chitobiose) complex with *Sm*CHB [1] allows interesting insights into the reaction mechanism, and particularly into the substrate conformational changes that occur along the reaction coordinate. Taken together, the crystal structures of the Michaelis complex from the family 20 chitobiase, and the intermediate analogue bound to the family 20 β-hexosaminidase, reveal that the anomeric carbon atom (C1) experiences the greatest nuclear motion along the reaction coordinate (Figure 2.11). Initially, atom C1 is above the plane of the sugar ring in a $^{4}B^{1}$ boat conformation and then, as the covalent bond to the glycosidic oxygen breaks, the anomeric carbon atom (C1) moves approximately 0.9 Å to a position below the plane of the sugar ring in a $^{4}C_{1}$ conformation (Figure 2.11), where it forms a new covalent bond with the C2-acetamido group oxygen atom, resulting in an oxazolinium ion intermediate (Fig. 1.2b, *lower pathway*). The nuclear motion proposed for the anchimeric assistance mechanism of family 20 glycosidases is in agreement with the recently proposed theory of electrophilic migration of the anomeric center along the reaction coordinate of HEW lysozyme and retaining β-glycosidases in general [28].

The pyranoside ring conformation of the cyclic intermediate analogue NAG-thiazoline bound to *Sp*HEX is strikingly similar to that observed for the recently determined glycosyl-enzyme intermediate structure of HEW lysozyme (Figure 2.12) [28] and other retaining β-glycosidases that use an enzymic nucleophile [29-31]. The 'textbook' mechanism of HEW lysozyme originally proposed by Phillips *et al.* [32], suggested a dissociative S_N1-type reaction

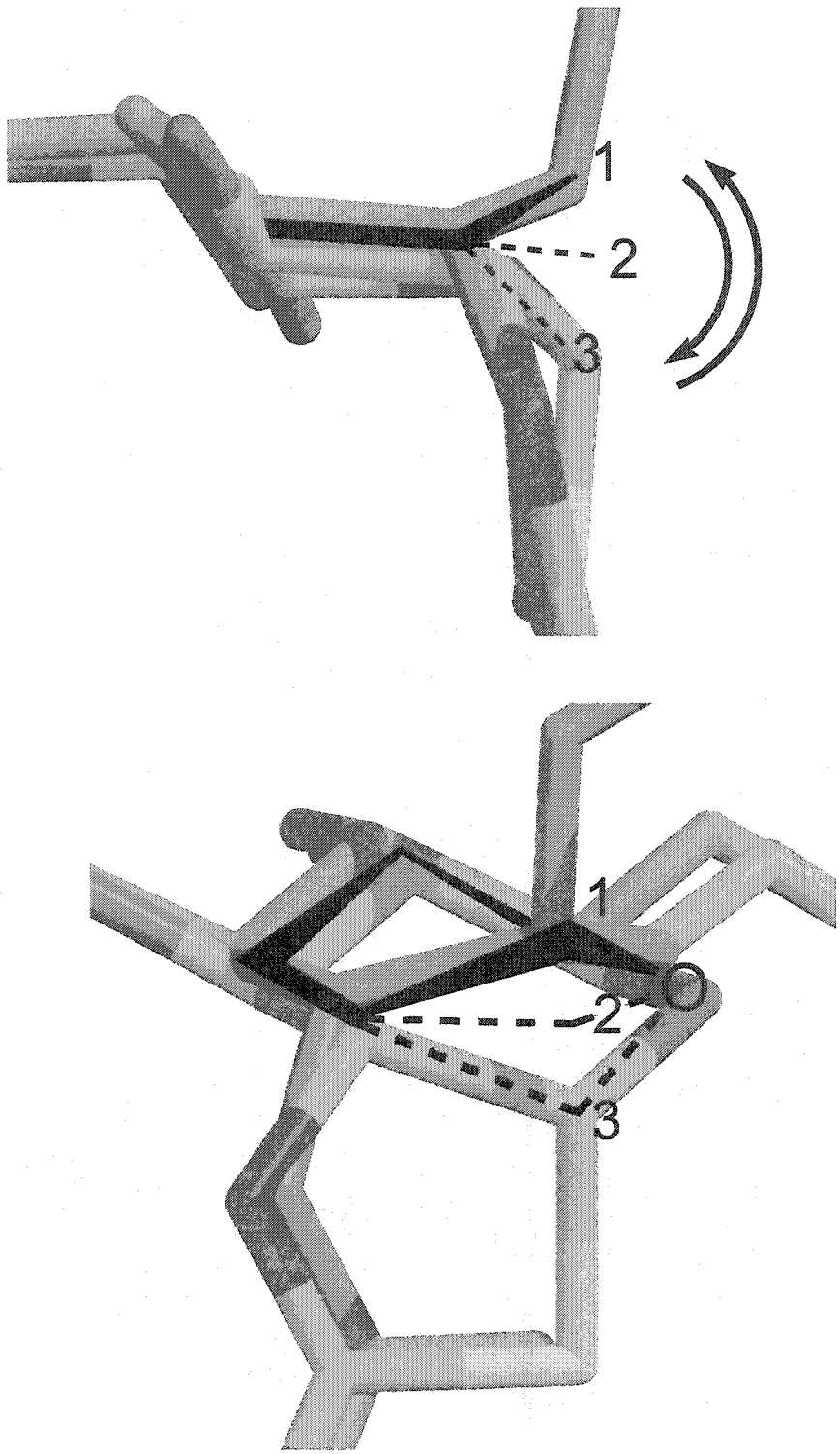


Figure 2.11. Cartoon indicating anomeric carbon movement during catalysis. As the reaction proceeds, the C1 atom scribes an arc from its initial position (**1**) as it breaks a covalent bond to the glycosidic oxygen to form a new bond with the acetamido oxygen (**3**). Approximately halfway along this arc is the transition state where C1, C2, C5 and O5 are co-planar (**2**). The Michaelis complex is chitobiose (*green carbon atoms*) bound to *Sm*CHB, and the intermediate is NAG-thiazoline bound to *Sp*HEX (*purple carbon atoms*).

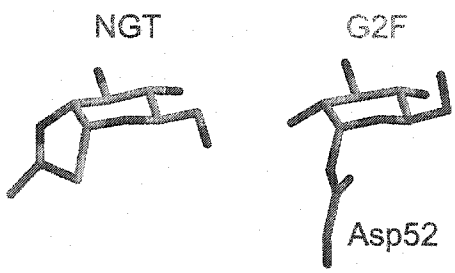


Figure 2.12. Comparison of the enzyme bound conformation of NAG-thiazoline (NGT) and the covalent NAG-(1,4)-2-deoxy-2-fluoro-β-D-glucopyranosy-HEW lysozyme intermediate structure [28] (only the sugar bound in the -1 subsite is shown (G2F)). The comparison clearly demonstrates that the sugar ring bound in the -1 subsite of both the SpHEX and HEW lysozyme intermediate complexes, adopts a conformation best described as a $^{4}C_{1}$ chair regardless of whether the nucleophile is provided by the substrate or by the enzyme.

mechanism involving a long-lived oxocarbenium ion intermediate that was stabilized by the enzyme through electrostatic interactions, primarily with the carboxylate of Asp 52. However, Vocadlo *et al.* [28] have recently trapped and solved the crystal structure of a glycosyl-enzyme intermediate form of HEW lysozyme, demonstrating that this enzyme proceeds via a covalent intermediate during catalysis and uses instead the S_N2-type double-displacement mechanism shown in Figure 1.2, *upper pathway*. This finding now places HEW lysozyme in the general class of retaining glycosidases and strengthens the general configuration-retaining mechanism that includes substrate distortion, electrophilic migration of the anomeric carbon atom during catalysis, and the formation of a covalent intermediate.

Thus, the crystal structures of the *Sm*CHB Michaelis complex and *Sp*HEX intermediate complex reveal that as the bound substrate proceeds along the reaction coordinate to yield the enzyme-bound product, the greatest nuclear motion of heavy atoms occurs at C1, as shown in Figure 2.11. As the reaction proceeds, the C1 atom scribes an arc from its initial position (**1**) as it breaks a covalent bond to the glycosidic oxygen to form a new bond with the acetamido oxygen (**3**). Approximately halfway along this arc is the transition state where C1, C2, C5 and O5 are coplanar (**2**). During hydrolysis of the intermediate, C1 traces the reverse path as it breaks the bond with the oxazolinium ion ring oxygen, proceeds through the transition state, and forms a covalent bond with an activated water molecule positioned above the β-face of the bound glycone. Thus, the motion of C1 through the catalytic cycle can be described as a "wagging" back and forth from positions above and below the plane of the sugar ring. Essentially no motion of protein atoms is required during the catalytic cycle. This overall behavior of *Sp*HEX as a catalyst is consistent with both the antiperiplanar lone pair hypothesis and the principle of least nuclear motion, and appears to be general for retaining glycosyl hydrolases [28, 30, 33].

In light of the recent structural and biochemical findings for family 18 and family 20 glycosidases and for HEW lysozyme, an alternative mechanism recently proposed on the basis of structural studies of a family 18 Chitinase A

(ChiA) from *Serratia marcescens* seems extremely unlikely [34]. A series of S. *marcescens* ChiA structures containing various active site mutations were solved in complex with octa- or hexa-*N*-acetylglucosamine. Based primarily on the electron density ascribed to a hexa-*N*-acetylglucosamine molecule bound in the active site of a ChiA mutant (Tyr390Phe, ChiA Tyr390 is equivalent to the Tyr 393 residue shown in Figures 2.7b), it was concluded that the C2-acetamido group of the sugar bound in the –1 subsite of the enzyme was inappropriately positioned to form an oxazolinium ion intermediate [34]. Instead, it was suggested that the carbonyl oxygen atom of the acetamido group assisted in stabilizing a long-lived oxocarbenium ion intermediate through electrostatic interactions in an S_N1-type mechanism. However, not only is there now a substantial amount of evidence refuting a glycosidase mechanism involving a long-lived oxocarbenium ion intermediate [28, 35], the electron density for the sugar bound in the –1 subsite of this complex is ambiguous. Thus, the mechanistic conclusions based on the structure of this complex should be verified, and C2-acetamido group participation that results in the formation of a cyclized oxazolinium ion intermediate in the catalytic mechanism of *S. marcescens* ChiA should not be ruled out based on these findings alone.

D. References

1. Tews, I., et al., *Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease.* Nat Struct Biol, 1996. **3**(7): p. 638-48.
2. Deslongchamps, P., *Stereoelectronic Effects in Organic Chemistry.* 1983, New York: Pergamon Press.
3. Kirby, A.J., *Stereoelectronic Effects on Acetyl Hydrolysis.* Accounts of Chemical Research, 1984. **17**: p. 305-311.
4. Knapp, S., et al., *NAG-thiazoline, an N-Acetyl-beta-hexosaminidase inhibitor that implicates acetamido participation.* J Am Chem Soc, 1996. **118**: p. 6804-6805.
5. Mark, B.L., et al., *Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase.* Journal of Biological Chemistry, 2001. **276**(13): p. 10330-10337.
6. Mark, B.L., et al., *Structural and functional characterization of Streptomyces plicatus beta-N-acetylhexosaminidase by comparative molecular modeling and site-directed mutagenesis.* J Biol Chem, 1998. **273**(31): p. 19618-24.
7. Sambrook, J., E.F. Fritsch, and T. Maniatis, *Molecular Cloning: A Laboratory Manual.* 2 ed. 1989, New York: Cold Spring Harbor. 9.31-10.17.
8. Van Duyne, G.D., et al., *Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin.* J Mol Biol, 1993. **229**(1): p. 105-24.
9. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Methods Enzymol, 1997. **276**: p. 307-326.
10. Hendrickson, W.A., *Determination of macromolecular structures from anomalous diffraction of synchrotron radiation.* Science, 1991. **254**(5028): p. 51-8.
11. Terwilliger, T.C. and J. Berendzen, *Automated MAD and MIR structure solution.* Acta Crystallogr D, 1999. **55**(Pt 4): p. 849-61.

12. Cowtan, K., *Error estimation and bias correction in phase-improvement calculations.* Acta Crystallogr D, 1999. **55**(Pt 9): p. 1555-67.

13. Collaborative Computational Project Number 4, *The CCP4 Suite: Programs for Protein Crystallography.* Acta Crystallogr D, 1994. **50**: p. 760-763.

14. Kleywegt, G.J. and T.A. Jones, *xdlMAPMAN and xdlDATAMAN - Programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets.* Acta Crystallogr D, 1996. **52**: p. 826-828.

15. Jones, T.A., et al., *Improved methods for the building of protein models in electron density maps and the location of errors in these models.* Acta Crystallogr A, 1991. **A47**(Pt 2): p. 110-119.

16. Brunger, A.T., *XPLOR: a system for X-ray Crystallography and NMR.* 1993, New Haven, CT: Yale University Press.

17. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallogr D, 1998. **54**(Pt 5): p. 905-21.

18. Guex, N. and M.C. Peitsch, *SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling.* Electrophoresis, 1997. **18**(15): p. 2714-2723.

19. Sidhu, G., et al., *Sugar ring distortion in the glycosyl-enzyme intermediate of a family G/11 xylanase.* Biochemistry, 1999. **38**(17): p. 5346-54.

20. Tanaka, A., et al., *GM2-gangliosidosis B1 variant: analysis of beta-hexosaminidase alpha gene abnormalities in seven patients [published erratum appears in Am J Hum Genet 1991 Jan;48(1):176].* Am J Hum Genet, 1990. **46**(2): p. 329-39.

21. Brown, C.A., et al., *Introduction of the α subunit mutation associated with the B1 variant of Tay-Sachs disease into the β subunit produces a β-hexosaminidase B without catalytic activity.* J. Biol. Chem., 1989. **264**: p. 21705-21710.

22. Fernandes, M.J.G., et al., *Identification of candidate active site residues in lysosomal beta-hexosaminidase A.* J. Biol. Chem., 1997. **272**(2): p. 814-820.

23. Pennybacker, M., et al., *Evidence for the involvement of Glu-355 in the catalytic action of human beta-hexosaminidase B.* J Biol Chem, 1997. **272**(12): p. 8002-6.

24. Sakuda, S., et al., *The structure of allosamidin, a novel insect chitinase inhibitor, produced by Streptomyces sp.* Tetrahedron, 1986. **27**(22): p. 2475-2478.

25. Terwisscha van Scheltinga, A.C., et al., *Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and X-ray structure of a complex with allosamidin: evidence for substrate assisted catalysis.* Biochemistry, 1995. **34**(48): p. 15619-23.

26. van Aalten, D.M., et al., *Structural insights into the catalytic mechanism of a family 18 exo- chitinase.* Proc Natl Acad Sci U S A, 2001. **98**(16): p. 8979-84.

27. Kobayashi, S., et al., *Enzymatic polymerization to artificial hyaluronan: a novel method to synthesize a glycosaminoglycan using a transition state analogue monomer.* J Am Chem Soc, 2001. **123**(47): p. 11825-6.

28. Vocadlo, D.J., et al., *Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate.* Nature, 2001. **412**(6849): p. 835-8.

29. Notenboom, V., et al., *Insights into transition state stabilization of the beta-1,4- glycosidase Cex by covalent intermediate accumulation in active site mutants.* Nature Structural Biology, 1998. **5**(9): p. 812-818.

30. Davies, G.J., et al., *Snapshots along an enzymatic reaction coordinate: analysis of a retaining beta-glycoside hydrolase.* Biochemistry, 1998. **37**(34): p. 11707-13.

31. White, A., et al., *Crystallographic observation of a covalent catalytic intermediate in a beta-glycosidase [see comments].* Nat Struct Biol, 1996. **3**(2): p. 149-54.

32. Phillips, D.C., Sci. Amer., 1966. **215**: p. 78-90.

33. Zechel, D.L. and S.G. Withers, *Glycosidase mechanisms: Anatomy of a finely tuned catalyst.* Acc Chem Res, 2000. **33**(1): p. 11-18.

34. Papanikolau, Y., et al., *High resolution structural analyses of mutant chitinase A complexes with substrates provide new insight into the mechanism of catalysis.* Biochemistry, 2001. **40**(38): p. 11338-43.

35. Davies, G.J., M.L. Sinnott, and S.G. Withers, *Glycosyl Transfer*, in *Comprehensive Biological Catalysis*, M.L. Sinnott, Editor. 1998, Academic Press Ltd.: New York. p. 119-208.

36. Nicholls, A., K.A. Sharp, and B. Honig, *Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons.* Proteins, 1991. **11**(4): p. 281-96.

37. Laskowski, R.A., et al., *PROCHECK: a program to check the stereochemical quality of protein structure coordinates.* Journal of Applied Crystallography, 1993. **26**: p. 283-291.

Chapter 3

Biochemical and stuctural assesment of GalNAc-isofagomine as a potent family 20 β-hexosaminidase Inhibitor

A. Introduction

Natural and synthetic glycosidase inhibitors are useful biological tools for helping to understand the catalytic mechanisms by which these ubiquitous enzymes process their natural substrates. Exploring the relationships between inhibitor structure, enzyme kinetics, and the molecular interactions that occur between an inhibitor and its enzyme target, not only provides insight into the catalytic mechanisms of these enzymes, but also gives an opportunity for knowledge-based design of potent and highly specific therapeutic agents [1]. Indeed, glycosidases have been implicated in numerous carbohydrate-mediated processes related to disease and much effort has been devoted to controlling such glycosidase activity selectively. Prominent examples include the use of sialidase inhibitors such as Zanamivir (Relenza), Oseltamivir (Tamiflu), and the experimental compound BCX-1812 for the treatment of influenza [2, 3] and the use of intestinal α-glucosidase inhibitors (e.g. acarbose and Miglitol) for the treatment of non-insulin-dependent diabetes mellitus [4, 5]. The recent successes of sugar based therapeutics has encouraged the continuing effort to develop a more complete understanding of the catalytic mechanisms of glycosidases through the synthesis and subsequent kinetic and structural analysis of novel inhibitors.

Transition state structures occurring along the reaction coordinates of both inverting and retaining glycosidases are known to have substantial oxocarbenium ion character [6-8]. The reversible azasugar inhibitors of the deoxynojirimycin and isofagomine classes are believed to mimic electrostatic charge distributions found within oxocarbenium ion transition state structures [1, 9, 10]. Assumed to be protonated and positively charged when bound in the enzyme active site, the endocyclic nitrogen atoms of azasugar inhibitors interact favorably with

catalytically important enzyme carboxylates [11]. Interestingly, depending on whether the nitrogen is located at the position corresponding to the endocyclic oxygen, O5, as in the deoxynojirimycin class, or is located at the anomeric center, as for the isofagomine inhibitors, there appears to be distinct selectivity for α and β-retaining glycosidases, respectively [9]. It was suggested that this selectivity resulted from the position of the enzyme nucleophile within the active sites of these enzymes.

Indeed, it was demonstrated that in cyclodextrin glycosyltransferase, a retaining α-glycosyltransferase/α-glycosidase, the nucleophile carboxylate is positioned such that its carboxyl oxygen atoms form a *syn* interaction with the anomeric center and endocyclic oxygen atom of the substrate [12]. It is conceivable that this interaction would favor binding of the deoxynojirimycin class of azasugar inhibitors because the protonated nitrogen (which replaces the endocyclic oxygen) could donate a stabilizing hydrogen bond to the carbonyl oxygen of the enzyme nucleophile [1, 10]. For retaining β-glycosidases however, the carboxyl oxygens of the nucleophile interact with the substrate from the opposite side and form instead a *syn* interaction with the anomeric center and 2-hydroxyl substituent of the substrate [12].

Having no interactions with the endocyclic oxygen, the enzyme nucleophile of retaining β-glycosidases will preferentially interact with the isofagomine class of azasugars whose protonated nitrogen is located at the anomeric center and is poised to donate a hydrogen bond to the enzyme nucleophile [9]. In support of this proposal is the recent crystallographic complex of a potent xylobiose-derived isofagomine inhibitor bound to a retaining family 10 xylanase (Figure 3.2). It showed a strong hydrogen-bonding interaction between the presumably protonated nitrogen of the inhibitor and the enzyme nucleophile [11]. Recent results, however, with the noeuromycin class of isofagomines, which contain a 2-hydroxyl adjacent to the ring nitrogen (N1), cast some doubts on this conjecture. This class of inhibitor bound more tightly than the isofagomines, and with approximately equal affinities to both α- and β-glycosidases [13].

Research described in this chapter was carried out to determine if the isofagomine class of azasugars could inhibit a β-retaining glycosidase that uses anchimeric assistance and lacks an apparent enzymic nucleophile. This research was very much a collaborative effort. The 1-*N*-azasugar inhibitor (2R,3R,4S,5R)-2-acetamido-3,4-dihydroxy-5-hydroxymethyl-piperidinium chloride (GalNAc-isofagomine · HCl) (Figure 3.1) was synthesized by Dalian Zhao and Spencer Knapp of Rutgers University, and assayed for its ability to competitively inhibit *Sp*HEX by David Vocadlo in the laboratory of Stephen Withers, University of British Columbia. The crystallographic structure of the complex of GalNAc-isofagomine bound to *Sp*HEX was determined by the author of this thesis, and reveals a novel binding mode for the inhibitor that provides insight into its unexpected potentency towards this family 20 glycosidase. The research presented here also appears as a published paper in the *Journal of Biological Chemistry* [14].

B. Experimental Procedures

B.1. Synthesis of (2R,3R,4S,5R)-2-acetamido-3,4-dihydroxy-5-hydroxymethyl-piperidinium chloride (GalNAc-isofagomine· HCl)

The synthesis of GalNac-isofagomine (Figure 3.1) was accomplished by Dalian Zhao and Spencer Knapp of Rutgers University as described in [14].

B.2. Protein expression and purification

*Sp*HEX was overexpressed in *Escherichia coli* strain BL21 (DE3) as an N-terminal His_7-tagged fusion protein and affinity purified using nickel-nitrilotriacetic acid superflow resin (Qiagen) as described in Chapter 2, section **A.1** [15]. Purified *Sp*HEX protein was stored at 4°C in elution buffer (20 mM Tris-Cl, pH 8.0, 300 mM NaCl, 10 mM β-mercaptoethanol, 250 mM imidazole pH 8.0) and was stable for several weeks.

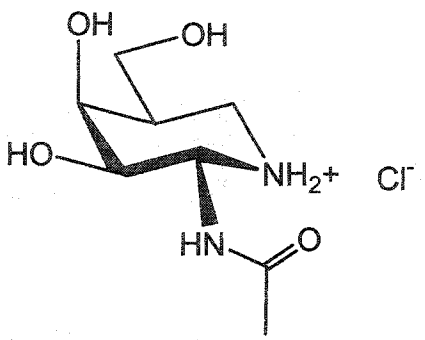


Figure 3.1. Chemical schematic of GalNAc-isofagomine · HCl (2R,3R,4S,5R)-2-acetamido-3,4-dihydroxy-5-hydroxymethyl-piperidinium chloride). The nitrogen atom (N1) that replaces the anomeric carbon atom is believed to be protonated and positively charged when bound in the enzyme active site. Compound synthesized by Dalian Zhao and Spencer Knapp as described in [14].

B.3. Enzyme kinetics

Enzyme kinetics, measured by David Vocadlo in the laboratory of Stephen Withers at the University of British Columbia, were carried out as follows: All inhibition constants for GalNac-isofagomine were measured at 25°C using *p*-nitrophenyl-2-acetamido-2-deoxy-β-D-glucopyranoside as substrate in a reaction buffer composed of 15 mM sodium phosphate / 15 mM sodium citrate buffer, 100 mM sodium chloride, pH 4.00 and 0.1% bovine serum albumin. Measurements were initiated by addition of *Sp*HEX previously dialyzed against the reaction buffer. Rates were determined by following the increase in absorbance at 360 nm arising from the release of *p*-nitrophenolate in a continuous assay. K_i values were determined by a direct fit of the experimental data using the program GraFit [16].

B.4. Crystallization and data collection

Prior to use in crystallization trials, *Sp*HEX was dialyzed against 50 mM trisodium citrate, pH 5.0 and 300 mM NaCl and concentrated to approximately 10 mg/ml with a Millipore concentrator. Hexagonal bipyramidal crystals ($P6_122$) of the *Sp*HEX-GalNAc-isofagomine complex were obtained by vapor diffusion at room temperature: the concentrated enzyme was incubated with 2 mM GalNAc-isofagomine for 30 min and then mixed in a 1:1 ratio with mother liquor (2.1 M ammonium sulphate, 100 mM trisodium citrate, pH 6.0 and 20% glycerol) to create several 8 μl hanging drops from which diffraction quality crystals were obtained within 2 weeks. High-resolution diffraction data were collected at the Stanford Synchrotron Radiation Laboratory (SSRL) beamline 9-1 from a single crystal of the complex flash-cooled to 100 K within a cryostatic N_2(g) stream. Immediately prior to data collection, the N_2(g) stream was diverted away from the flash-cooled crystal for approximately 3 seconds to reanneal the crystal, thereby reducing its mosaic spread. All diffraction data were processed using DENZO and SCALEPACK [17].

B.5. Structure determination and refinement

The 1.75 Å resolution $2|F_o|-|F_c|$, α_c and $|F_o|-|F_c|$, α_c electron density maps (where α_c are the calculated phases and $|F_o|$ and $|F_c|$ are the measured and calculated structure factor, respectively), used to visualize GalNAc-isofagomine in the *Sp*HEX active site, were obtained using structure factor phases calculated from a model of native *Sp*HEX (PDB code: 1HP4) [15]. Using a maximum likelihood target function implemented in the CNS program [18], the molecular model of native *Sp*HEX was positioned into the unit cell of the GalNAc-isofagomine complex using rigid body refinement followed by several rounds of conjugate gradient minimization. Solvent molecules were removed from the native enzyme model prior to placing it into the unit cell of the GalNAc-isofagomine complex and were relocated during later rounds of refinement. The initial $|F_o|-|F_c|$, α_c map, computed prior to solvent modeling (R_{work} = 25.9%, R_{free} = 28.3%), unambiguously defined the conformation of the bound GalNAc-isofagomine, and a model of the inhibitor was manually fit into the electron density using O [19]. Topology and parameter files for the GalNAc-isofagomine model were generated using Xplo2D [20] and subsequent rounds of solvent modeling and positional refinement using conjugate gradient minimization with a maximum likelihood target function were carried out within the CNS program [18] until *R*-factor convergence was reached. The final refinement statistics are presented in Table 3.1.

B.6. Coordinates

Coordinates and structure factors have been deposited into the Protein Data Bank (code: 1JAK).

C. Results and Discussion

The potency of isofagomine inhibitors for retaining β-glycosidases has been attributed to the formation of a strong electrostatic interaction between a protonated endocyclic nitrogen at the 'anomeric' center of the inhibitor and the catalytic nucleophile of the enzyme [11]. Because family 20 glycosidases lack the

Table 3.1 Crystallographic statistics

Crystal Information & data collection		*Refinement*	
Space group:	$P6_122$	Resolution (Å):	100 - 1.75
Unit cell:	a=b=132.8, c=177.0	R_{work}[b]:	17.6%
Detector:	Mar Research 345	R_{free}[b]:	19.2%
Wavelength (Å):	0.979	Number of atoms	
Resolution (Å):	100 - 1.75	• protein:	3864
Total observations:	1013193	• IFG:	14
Unique reflections:	92286 (4366)	• water:	533
$< I / \sigma >$	36.2 (10.9)	Average B ($Å^2$):	14.2
Completeness (%):	99.6 (95.7)	RMSD from ideal geometry	
R_{sym}[a]	0.043 (0.15)	• bond lengths (Å):	0.0047
		• bond angles (°):	1.34
(values in parentheses refer to the high-resolution shell: 1.78 - 1.75 Å)		Ramachandran plot	
		• %-most favoured[c]:	91.1%
		• %-additionally allowed[c]:	8.9%

[a]$R_{sym} = \Sigma_{\mathbf{h}} \Sigma_i (|I_i(\mathbf{h}) - \langle I(\mathbf{h})\rangle|) / \Sigma_{\mathbf{h}} \Sigma_i I_i(\mathbf{h})$, where $I_i(\mathbf{h})$ is the ith intensity measurement and $\langle I(\mathbf{h})\rangle$ is the weighted mean of all measurements of $I(\mathbf{h})$.

[b]R_{work} and $R_{free} = \Sigma_{\mathbf{h}} ||F(\mathbf{h})_{obs}| - |F(\mathbf{h})_{calc}|| / |F(\mathbf{h})_{obs}|$ for reflections in the working and test sets (10% of all data) respectively.

[c]Regions defined by PROCHECK [26]

IFG = GalNAc-isofagomine

enzymic nucleophile required to form this strong enzyme-inhibitor interaction, it was believed that isofagomines would be poor inhibitors of family 20 glycosidases. Contrary to this belief however, GalNAc-isofagomine (Figure 3.1) was found to act as a potent competitive inhibitor of the family 20 glycosidase *Sp*HEX, exhibiting a K_i of 2.7 μM when using pNPGlcNAc as substrate (data provided by D. Vocaldo and S. Withers) [14].

Insight into the mechanism by which GalNAc-isofagomine inhibited *Sp*HEX has been provided by the X-ray crystal structure of the enzyme-inhibitor complex (Figure 3.3a). Crystals of the complex diffracted to 1.75Å resolution, and the structure refined to an R_{work} of 17.6 % and R_{free} of 19.2% (Table 3.1). The excellent crystallographic data allowed for the calculation of an easily interpretable $|F_o|-|F_c|$, α_c electron density map for GalNAc-isofagomine (Figure 3.3). The inhibitor was located in the -1 subsite of the *Sp*HEX active site pocket. The azasugar ring of the inhibitor adopts a conformation approaching that of a half-chair, in which the endocyclic nitrogen N1 at the 'anomeric' center is displaced 0.4 Å from a least squares plane formed by atoms C2, C3, C5 and C9 (atom C9 replaces the atom O5 found in pyranose rings) and atom C4 is displaced 0.7 Å on the opposite side of this plane. Torsion angle measurements within the azasugar ring also demonstrate the flattened nature of the ring about N1 as compared to the X-ray structure of *N*-acetyl-galactosamine [21] (Table 3.2).

The flattened ring conformation of GalNAc-isofagomine differs from the relaxed $^4C_1$ chair conformation observed for the proximal azasugar ring of xylobiose-derived isofagomine bound to the family 10 glycosidase Cex from *Cellulomonas fimi* (Figure 3.2) [11], indicating that in *Sp*HEX, the enzyme bound conformation of GalNAc-isofagomine better approximates the planar conformation believed to occur for atoms C1, C2, O5 and C5 during the oxacarbenium ion transition state of the natural substrate. The oxacarbenium ion transition state is thought to be a half-chair or skew-boat with atoms C1, C2, O5 and C5 forming a plane [22]. As disscused in Chapter 1, section **D.3.1**, the co-planarity of these atoms during the transition state is required for effective

Table 3.2. Endocyclic and *N*-acetyl group torsion angles for GalNAc-isofagomine and *N*-acetyl-α-galactosamine [21]

	GalNAc-isofagomine	*N*-acetyl-α-galactosamine
Endocyclic*		
(C9/O5)-(N1/C1)-C2-C3	30.0°	56.6°
(N1/C1)-C2-C3-C4	-38.4°	-55.8°
C2-C3-C4-C5	52.0°	55.8°
C3-C4-C5-(C9/O5)	-58.6°	-57.5°
C4-C5-(C9/O5)-(N1/C1)	51.3°	60.2°
C5-(C9/O5)-(N1/C1)-C2	-36.3°	-59.3°
N-Acetyl group*		
(N1/C1)-C2-N2-C7	-51.5°	82.8°
C3-C2-N2-C7	76.0°	-154.9°
C2-N2-C7-O7	2.5°	-0.4°
C2-N2-C7-C8	-178.3°	179.0°

*Atoms names in brackets indicate atoms types for GalNAc-isofagomine and N-acetyl-α-galactosamine respectively.

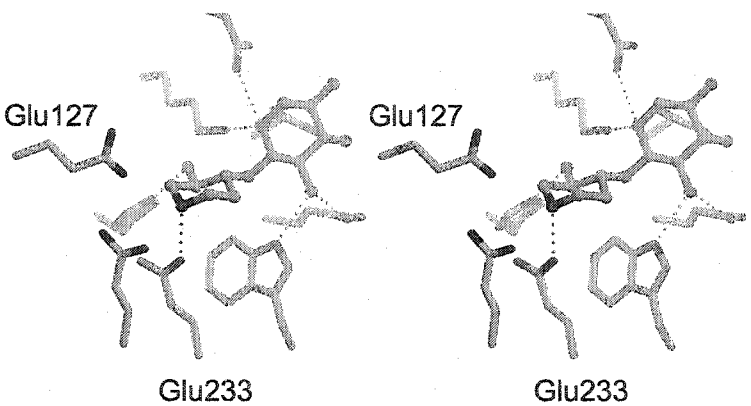


Figure 3.2. Stereo image of a xylobiose-derived isofagmine inhibitor bound in the active site of a family 10 xylanse Cex from *Cellulomonas fimi* **(K_i = 130 nM) [11].** The endocyclic nitrogen of the azasugar bound in the -1 sugar binding subsite forms a strong electrostatic interaction with the enzyme nucleophile Glu233. The general acid/base residue, Glu127, may form a weak hydrogen bonding interaction (3.2 A) with the bound azasugar.

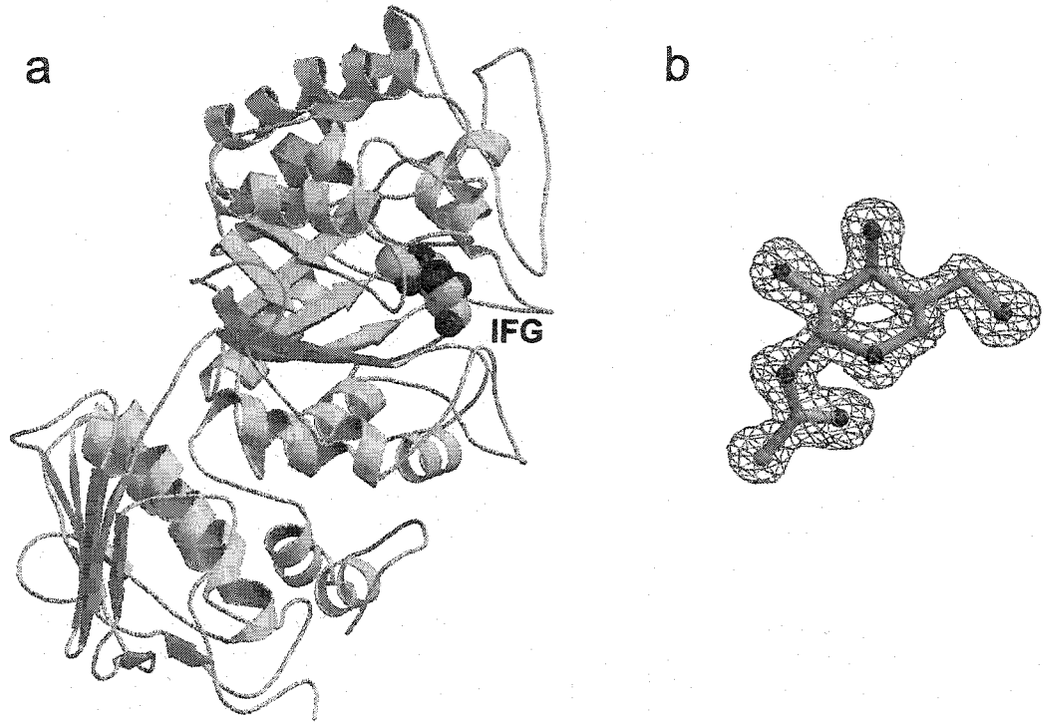


Figure 3.3. Ribbon diagram of *Sp*HEX in complex with GalNAc-isofagomine. a, GalNAc-isofagomine (IFG) bound in the -1 sugar binding subsite of *Sp*HEX. IFG is depicted as space filling CPK model with carbon atoms in *grey*, nitrogen atom in *blue* and oxygen atoms in *red* . b, Electron density for IFG. The refined model is drawn as sticks with carbon atoms in *gray*, nitrogen atoms in *blue* and oxygen atoms in *red*. The electron density was calculated as an |Fo|-|Fc| simulated annealing omit map as implemented in CNS [18] and contoured at 3.0 σ.

overlap of the nonbonding lone pair electrons of O5 and the antibonding orbital at the electron deficient anomeric center of the oxocarbenium ion (however, the conformation of GalNAC-isofagomine is not truly planar about atoms C1, C2, C9 and C5). Taken together with the observations for xylobiose-derived isofagomine bound to Cex [11], it appears that although the charge distribution in the isofagomine class of azasugars may mimic the oxocarbenium ion transition state, the enzyme bound conformation of the azasugar ring better reflects that of the pyranose ring of the enzyme bound intermediate that occurs during reactions catalyzed by β-retaining glycosidases [15, 23].

The *Sp*HEX active site contains a tight hydrophobic pocket composed of Trp344, Trp361, Trp442, Tyr393 and Asp313 that is responsible for orienting the C2-acetamido group in position for intramolecular nucleophilic attack (see Chapter 2, section **B.2)** [15]. Relative to the X-ray crystal structure of unbound *N*-acetyl-galactosamine [21], which has a C1-C2-N2-C7 torsion angle of 82.8°, the C2-acetamido group of GalNAc-isofagomine was found to be rotated 134° about the C2-N2 bond when bound to *Sp*HEX, resulting in a final N1-C2-N2-C7 torsion angle of –51.52° (Table 3.2). This rotation about the C2-N2 bond allows for a 2.7 Å intramolecular hydrogen bond between the carbonyl oxygen atom of the 2-acetamido group of the inhibitor and the axial proton of N1 at its 'anomeric' center (Figure 3.4). This intramolecular hydrogen bond is strikingly similar to the 2.6 Å hydrogen bond observed between N1 of xylobiose-derived isofagomine and the enzymic nucleophile of *Cex* (Glu233) to which it is bound (Figure 3.2) [11]. This comparison again substantiates the hypothesis that the carbonyl oxygen atom of the 2-acetamido group participates in catalysis and acts in place of an enzymic nucleophile during reactions catalyzed by family 20 glycosidases (18).

The conformational changes observed for GalNAc-isofagomine upon binding *Sp*HEX are consistent with the conformational changes observed for the 2-acetamido group of the non-reducing sugar of chitobiose upon binding to the family 20 chitobiase from *Seratia marcessens* (*Sm*CHB). The small molecule crystal structure of chitobiose reveals the carbonyl oxygen atom of the C2-

acetamido group of the non-reducing sugar to be +anticlinal to C1, with the torsion angle C1-C2-N2-C7 being 100.5° [24]. However, the crystallographically observed Michaelis complex between chitobiose and the *Sm*CHB shows the 2-acetamido group of the nonreducing GlcNAc residue (bound in the -1 subsite) to be rotated 156° about its C2-N2 bond with a final C1-C2-N2-C7 torsion angle of –56.27° (Figure 1.4). This bond rotation, in combination with a distortion of the pyranose ring from a $^4C_1$ chair towards a boat conformation, brings the carbonyl oxygen of the 2-acetamido group of chitobiose to within 3.0 Å of the anomeric center, positioning it for attack at C1 and formation of a cyclic oxazolinium ion intermediate (Figure 1.2, *lower pathway*) [25].

The notable difference between enzyme bound GalNAc-isofagomine and chitobiose is the shorter intramolecular distance between N1 and the carbonyl oxygen atom of the C2-acetamido group of GalNAc-isofagomine (2.7 Å) as compared to the 3.0 Å distance observed between the analogous atoms in chitobiose bound to *Sm*CHB. If the conformation of GalNAc-isofagomine bound to *Sp*HEX is considered equivalent to the pyranose ring conformation of the enzyme bound intermediate during a normal catalytic cycle, then the difference in distance between the nucleophile and anomeric centers for GalNAc-isofagomine and chitobiose can be attributed to conformational changes that occur in the pyranose ring as the reaction proceeds from the enzyme-substrate complex to the enzyme bound intermediate: the pyranose ring changes from a boat, as seen in the Michaelis complex between chitobiose and *Sm*CHB, to a chair conformation as observed for the cyclic thiazolinum ion intermediate analogue, NAG-thiazoline [15], and GalNAc-isofagomine bound to *Sp*HEX.

During a normal catalytic cycle *Sp*HEX donates a hydrogen bond from the side chain hydroxyl group of Tyr393 to the 2-acetamido carbonyl oxygen, helping to align the carbonyl oxygen with the anomeric center of the substrate [15]. Cyclization to form the oxazolinium ion intermediate results in a formal positive charge developing on the protonated 2-acetamido nitrogen (N2). *Sp*HEX stabilizes this positive charge by accepting a hydrogen bond from N2 to the carboxylate of Asp313 [15]. Both of these important hydrogen-bonding

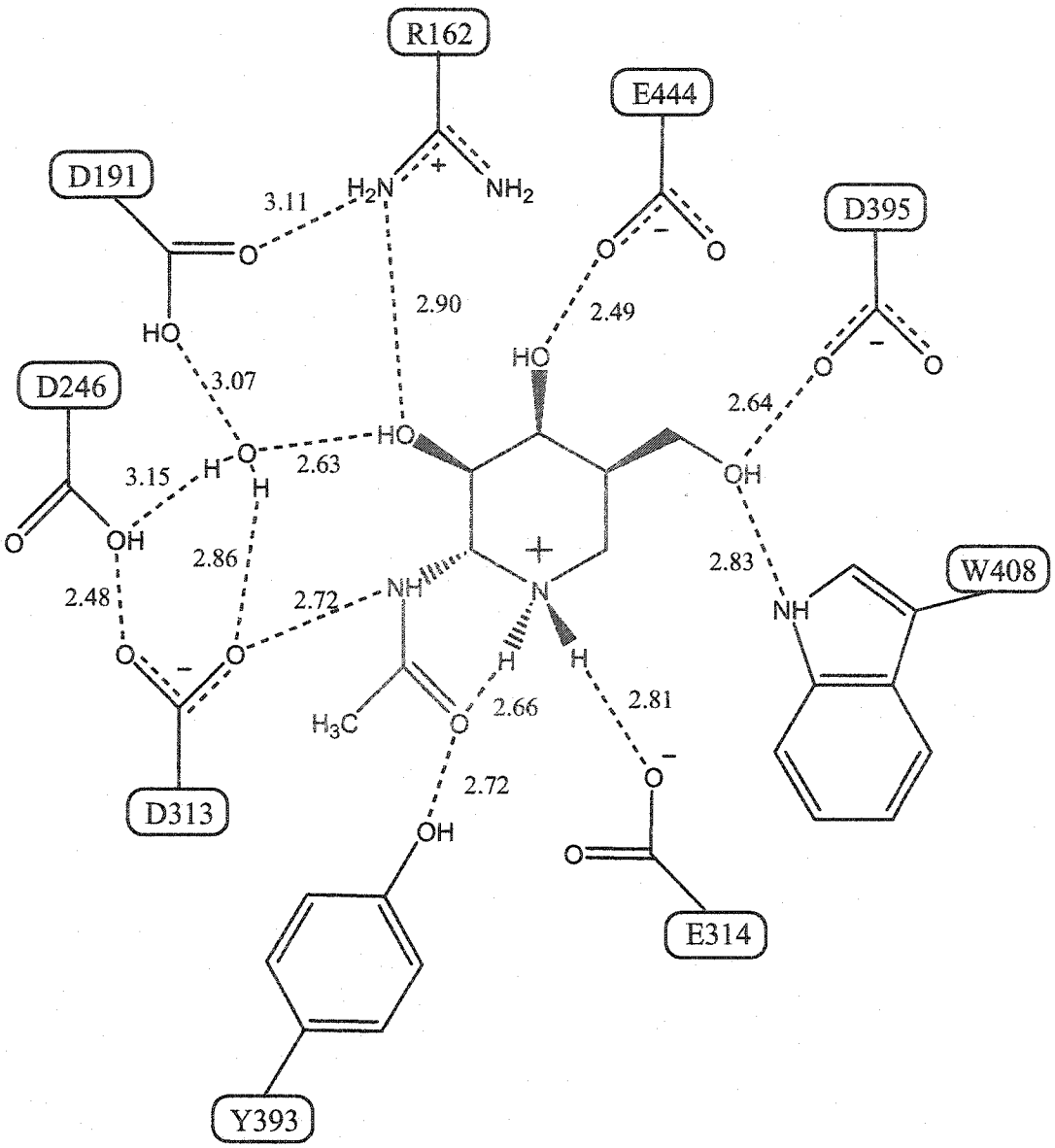


Figure 3.4. Schematic of the hydrogen-bonding interactions occurring between *Sp*HEX and GalNAc-isofagomine (Figure 3.1). Hydrogen bonds are shown as dashed lines and their respective distances indicated in Ångstroms. GalNAc-isofagomine is shown in blue and the *Sp*HEX residue side chains are shown in black and numbered. No attempt has been made to illustrate the true positions of the amino acid side chains.

interactions with Tyr393 and Asp313 occur between *Sp*HEX and the 2-acetamido group of GalNAc-isofagomine and they essentially lock the 2-acetamido group into position within the hydrophobic pocket (Figure 3.4).

The 2-acetamido group of GalNAc-isofagomine is not normally expected to bear a negative charge; however, considering that *Sp*HEX has evolved to stabilize a protonated and positively charged 2-acetamido group nitrogen (N2), it is very possible that the enzyme may polarize the amide of GalNAc-isofagomine beyond its naturally occurring dipole moment, resulting in a further increase of charge density on the carbonyl oxygen. The polarized amide could favorably interact with the formal positive charge on the protonated nitrogen at the 'anomeric' center (as indicated by the 2.7 Å hydrogen bond in Figure 3.4); whereas, the electron deficient C2-acetamido nitrogen (N2) would be stabilized through donation of a hydrogen bond to the caboxylate of Asp313 in a manner similar to the stabilization of the oxazolinium ion intermediate [15]. The hydrogen bond donated from the side chain hydroxyl group of Tyr393 could also help to stabilize the increased electron charge density of the carbonyl oxygen of GalNAc-isofagomine. Such polarization of the amide of the natural substrate would greatly enhance the nucleophilicity of the 2-acetamido carbonyl oxygen atom and increase the efficiency of catalysis by family 20 enzymes.

This structure of GalNAc-isofagomine bound to *Sp*Hex is the first heterocycle with a *galacto* configuration to be solved in complex with a family 20 hexosaminidase. For substrates of *gluco* configuration, it has been observed in both the Michaelis complex [25] and cyclic enzyme intermediate [15], that family 20 glycosidases form a bidentate hydrogen-bonding interaction between O3 and O4 of the substrate and the two nitrogens of the guanidinium group of a conserved Arg residue (Arg162 *Sp*HEX, Arg349 *Sm*CHB) (See Chapter 1, section **D3.3** and Chapter 2, section **B.2**). However, the crystallographic complex between *Sp*HEX and GalNAc-isofagomine demonstrates that the enzyme cannot accommodate a bidentate hydrogen-bonding interaction between Arg162 and the *galacto* configuration of GalNAc-isofagomine (Figure 3.5a). Nevertheless, *Sp*HEX and the related family 20 glycosidases catalyze the

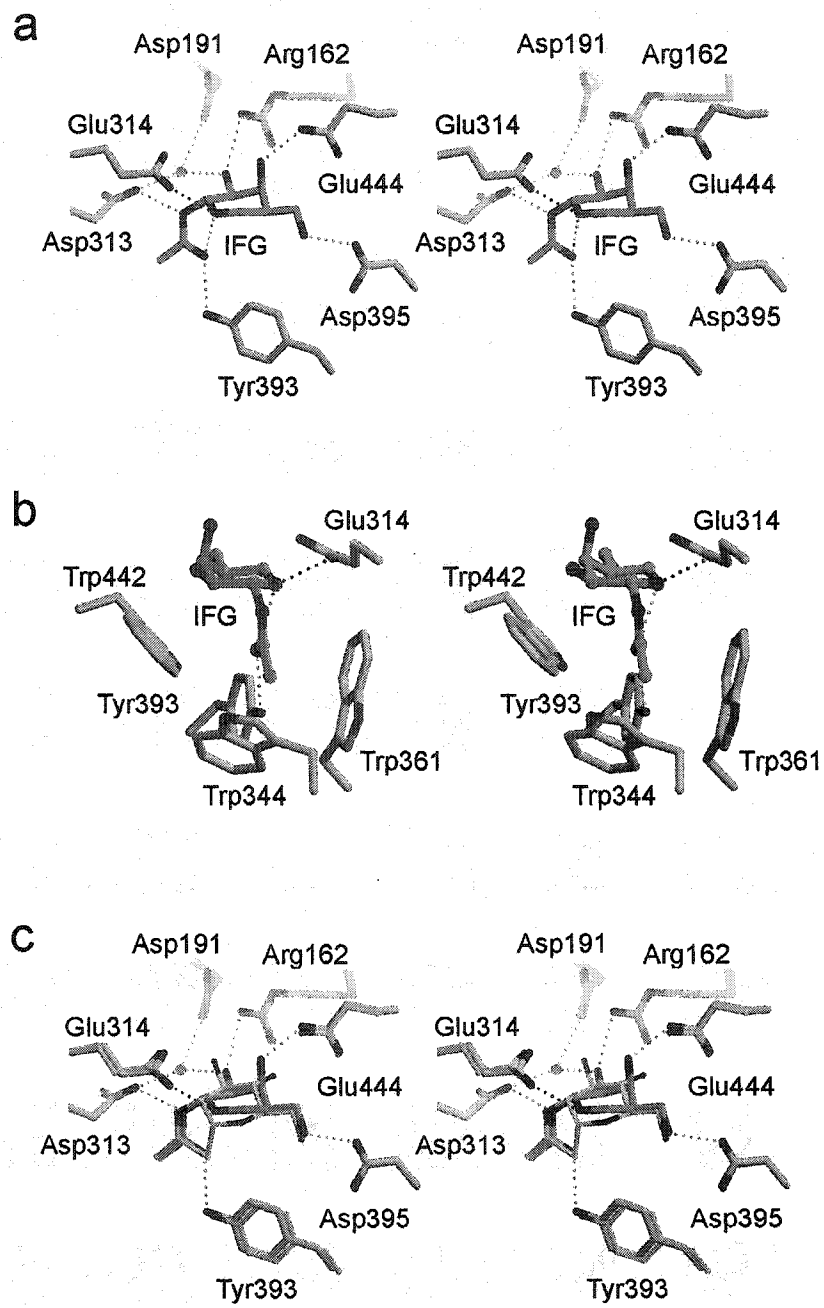


Figure 3.5. SpHEX in complex with GalNAc-isofagomine (IFG). a, Polar SpHEX active site residues that interact with IFG (green carbon atoms). b, Hydrophobic SpHEX active site residues that distort the C2-acetamido group into position for catalysis. c, Superposition of the crystallographic complexes between SpHEX and IFG (green carbon atoms) or NAG-thiazoline [15] (blue carbon atoms). The model of the SpHEX-NAG-thiazoline complex is drawn as ball-and-sticks, whereas the SpHEX-IFG complex is drawn as sticks only. The superposition was carried out in O [19] using residues Trp 344, Trp 361 and Trp 442 (not shown) as described in [15].

removal of terminal N-acetylhexosamine residues of both *gluco* and *galacto* configuration. Glu444 of SpHEX was found to form a hydrogen bond with O4 of GalNAc-isofagomine, and this interaction may compensate for the absence of the bidentate hydrogen bond that occurs between Arg162 and substrates of *gluco* configuration.

A comparison of the structures of *Sp*HEX in complex with either the intermediate analogue NAG-thiazoline or GalNAc-isofagomine demonstrate that O3 forms a hydrogen bond with the guanidino group of Arg162 regardless of whether the sugar ring is of a *gluco* or *galacto* configuration; however, when O4 is in an axial position as observed for the *galacto* azasugar ring GalNAc-isofagomine, it is too far away (3.7 Å) and too poorly positioned to form a hydrogen bond with the guanidinium group of Arg162 (Figure 3.5). Instead, the *Sp*HEX-GalNAc-isofagomine complex shows the axial O4 atom of GalNAc-isofagomine forming a short 2.5 Å hydrogen-bonding interaction with the carboxyl group of Glu444 and not with Arg162 (Figure 3.5). *Sp*HEX Glu444 is conserved in family 20 glycosidases, suggesting that this residue is critical for binding substrates of *galacto* configuration. The carboxylate of this conserved Glu residue has also been shown to form a 2.6 Å hydrogen bond with O4 of substrates and intermediates of *gluco* configuration as observed in the crystal structure of *Sp*HEX in complex with the intermediate analogue NAG-thiazoline [15] and *Sm*CHB in complex with chitobiose [25].

Perhaps the most surprising result from the crystallographic analysis of the complex between *Sp*HEX and GalNAc-isofagomine is the 2.8 Å hydrogen-bonding interaction between the general acid-base residue Glu314 and the equatorial proton of the 'anomeric' nitrogen of the isofagomine inhibitor (Figure 3.4 & 3.5). Such a hydrogen-bonding interaction between the protonated nitrogen at the 'anomeric' center of an isofagomine inhibitor and the acid/base catalytic residue has not been observed in the other known complex of an isofagomine and retaining β-glycosidase [11]. The ability of GalNAc-isofagomine to form such a hydrogen bond with the general acid-base residue of a family 20 glycosidase appears to result from the ability of the catalytic nucleophile to move

in concert with the azasugar ring within the active site pocket of *Sp*HEX (Figure 3.5).

Relative to the position of the cyclic intermediate analogue NAG-thiazoline bound within the active site of *Sp*HEX [15], the azasugar ring of GalNAc-isofagomine is rotated about an axis best defined by atoms C2 and C5 such that the nitrogen at the 'anomeric' center is brought closer to Glu314 while maintaining a 2.7 Å hydrogen bond with the carbonyl oxygen atom of the attached 2-acetamido group (Figures 3.4 & 3.5). Furthermore, compared to the wild type *Sp*HEX structure, Glu314 responds to the binding of GalNAc-isofagomine with a 26° rotation about $\chi 3$ such that $O^{\varepsilon 2}$ is brought to within hydrogen-bonding distance of the nitrogen at the 'anomeric' center of GalNAc-isofagomine. The conformational change in the side chain of Glu314, in combination with the binding mode of GalNAc-isofagomine, reduces the distance between the 'anomeric' center of GalNAc-isofagomine and $O^{\varepsilon 2}$ of Glu314 to 2.8 Å compared to the 4.0 Å distance observed between the equivalent atoms in the *Sp*HEX-NAG-thiazoline complex [15]. Considering the potent inhibition of GalNAc-isofagomine toward *Sp*HEX, this hydrogen-bonding interaction appears to compensate for the missing electrostatic interaction that usually occurs between the protonated endocyclic nitrogen of an isofagomine inhibitor and a 'normal' glycosidase containing an enzymic nucleophile.

The unexpected potency of GalNAc-isofagomine toward *Sp*HEX highlights the need for detailed molecular structure analysis of protein-inhibitor interactions in order to understand clearly the mechanism by which a small molecule inhibits the catalytic activity of its target enzyme. Prior biochemical and structural information suggested that the isofagomine class of inhibitors would act as poor competitive inhibitors of family 20 glycosidases. However, kinetic analysis of the inhibitory activity of GalNAc-isofagomine toward *Sp*Hex proved otherwise, and it was through the analysis of the crystallographic complex formed between these two molecules that detailed insight into the mechanism of this unexpected inhibitory activity was obtained.

D. References

1. Heightman, T.D. and A.T. Vasella, *Recent insights into inhibition, structure, and mechanism of configuration-retaining glycosidases.* Angewandte Chemie-International Edition, 1999. **38**(6): p. 750-770.
2. Gubareva, L.V., L. Kaiser, and F.G. Hayden, *Influenza virus neuraminidase inhibitors.* Lancet, 2000. **355**(9206): p. 827-835.
3. Taylor, G., *Sialidases: Structures, biological significance and therapeutic potential.* Current Opinion in Structural Biology, 1996. **6**(6): p. 830-837.
4. Martin, A.E. and P.A. Montgomery, *Acarbose: An alpha-glucosidase inhibitor.* American Journal of Health-System Pharmacy, 1996. **53**(19): p. 2277-2290.
5. Welborn, T.A., *Acarbose, an alpha-glucosidase inhibitor for non-insulin-dependent diabetes.* Medical Journal of Australia, 1998. **168**(2): p. 76-78.
6. McCarter, J.D., M.J. Adam, and S.G. Withers, *Binding-Energy and Catalysis - Fluorinated and Deoxygenated Glycosides as Mechanistic Probes of Escherichia-Coli (Lacz) Beta-Galactosidase.* Biochemical Journal, 1992. **286**: p. 721-727.
7. McCarter, J.D. and S.G. Withers, *Mechanisms of enzymatic glycoside hydrolysis.* Curr Opin Struct Biol, 1994. **4**(6): p. 885-92.
8. Namchuk, M.N. and S.G. Withers, *Mechanism of Agrobacterium beta-glucosidase: Kinetic analysis of the role of noncovalent enzyme/substrate interactions.* Biochemistry, 1995. **34**(49): p. 16194-16202.
9. Bols, M., *1-aza sugars, apparent transition state analogues of equatorial glycoside formation/cleavage.* Accounts of Chemical Research, 1998. **31**(1): p. 1-8.
10. Ichikawa, Y., et al., *1-N-iminosugars: Potent and selective inhibitors of beta- glycosidases (vol 120, pg 3007, 1998).* Journal of the American Chemical Society, 1998. **120**(23): p. 5854-5854.
11. Notenboom, V., et al., *Detailed structural analysis of glycosidase/inhibitor interactions: Complexes of Cex from Cellulomonas fimi with xylobiose-derived aza-sugars.* Biochemistry, 2000. **39**(38): p. 11553-11563.

12. Uitdehaag, J.C.M., et al., *X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the alpha-amylase family.* Nature Structural Biology, 1999. **6**(5): p. 432-436.
13. Liu, H., Liang X., Shoel, H., Blow, A., Bols, M., *Noeuromycin,1 A Glycosyl Cation Mimic that Strongly Inhibits Glycosidases.* Journal of the American Chemical Society, 2001. **123**(21): p. 5116-5117.
14. Mark, B.L., et al., *Biochemical and structural assessment of the 1-N-azasugar GalNAc-isofagomine as a potent family 20 β-N-acetylhexosaminidase inhibitor.* J. Biol. Chem., 2001. **276**(45): p. 42131-42137.
15. Mark, B.L., et al., *Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase.* Journal of Biological Chemistry, 2001. **276**(13): p. 10330-10337.
16. Leatherbarrow, R.J., *GraFit version 4.0. Data Analysis and Graphics program for the IBM PC.* 1998, Erithacus Software Ltd.
17. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Methods Enzymol, 1997. **276**: p. 307-326.
18. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallogr D, 1998. **54**(Pt 5): p. 905-21.
19. Jones, T.A., et al., *Improved methods for the building of protein models in electron density maps and the location of errors in these models.* Acta Crystallogr A, 1991. **A47**(Pt 2): p. 110-119.
20. Kleywegt, G.J., *Dictionaries for Heteros.* CCP4/ESF-EACBM Newsletter on Protein Crystallography, 1995. **31**: p. 45-50.
21. Gilardi, R.D., Flippen, J. L., Acta Crystallogr B, 1974. **30**: p. 2931.
22. Sinnott, M.L., *Catalytic Mechanisms of Enzymic Glycosyl Transfer.* Chem Rev, 1990. **90**: p. 1171-1202.
23. White, A., et al., *Crystallographic observation of a covalent catalytic intermediate in a beta-glycosidase [see comments].* Nat Struct Biol, 1996. **3**(2): p. 149-54.

24. Mo, F., Jensen, L.H., Acta Crystallogr B, 1978. **34**: p. 1562.
25. Tews, I., et al., *Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease.* Nat Struct Biol, 1996. **3**(7): p. 638-48.
26. Laskowski, R.A., et al., *PROCHECK: a program to check the stereochemical quality of protein structure coordinates.* Journal of Applied Crystallography, 1993. **26**: p. 283-291.

Chapter 4

The role of Asp313 in the catalytic mechanism of *Streptomyces plicatus* β-hexosaminidase

A. Introduction

The salient position of Asp313 in the active site of *Sp*HEX, and its equivalent residue in the active site of *Sm*CHB (Asp539) has prompted studies into the role of this conserved residue in the catalytic mechanism of family 20 glycosidases. The crystallographic Michaelis complex between *Sm*CHB and chitobiose revealed that the carboxylate oxygen atom $O^{\delta 2}$ of Asp539 accepts a 2.9 Å hydrogen bond from the amide nitrogen of the C2-acetamido group of the non-reducing sugar of chitobiose [1] (Figure 4.1). Moreover, the crystallographic complex between *Sp*HEX and the cyclic intermediate complex NAG-thiazoline showed that the equivalent oxygen atom of Asp313 forms a short (2.5 Å) hydrogen-bonding interaction with the thiazoline ring nitrogen of the intermediate analogue [2] (Figure 4.1).

To ensure that Asp313 of *Sp*HEX and Asp539 of *Sm*CHB remain properly oriented and negatively charged throughout the catalytic cycle, the carboxylate oxygen atoms $O^{\delta 1}$ of these residues share a proton via a short hydrogen bond (~2.5 Å) with the carboxyl group of an additional conserved Asp residue within each enzyme active site: Asp246 of *Sp*HEX and Asp448 of *Sm*CHB (Figure 4.1). These additional Asp residues are conserved in all known family 20 glycosidases, and, in addition to typical side chain packing, the orientation of the carboxyl group of this residue is fixed into position in both *Sp*HEX and *Sm*CHB through the acceptance of a hydrogen bond from a spatially conserved main chain nitrogen atom.

Asp313 of *Sp*HEX and Asp539 of *Sm*CHB appear to carry out two important catalytic functions within each enzyme: 1) they orient and hold the C2-acetamido group of the non-reducing sugar into position for catalysis, and 2) they stabilize the positive charge that develops on the C2-acetamido group nitrogen atom upon cyclization and formation of the oxazolinium ion intermediate. To

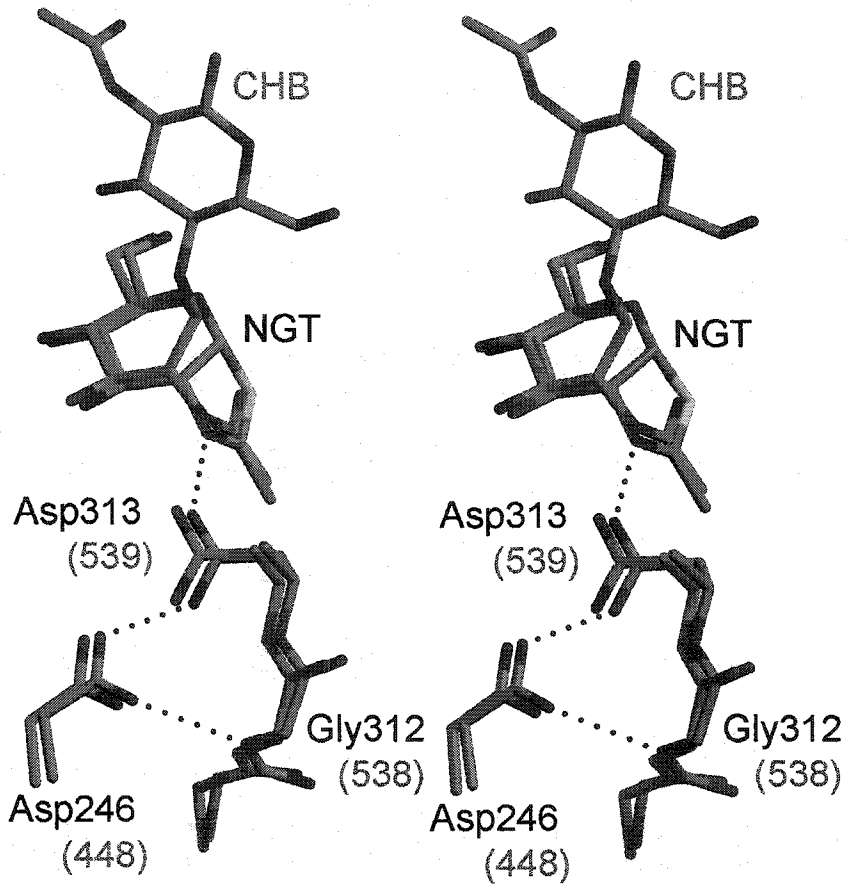


Figure 4.1. Stereographic superposition of the crystallographic structures of *Sm*CHB bound to chitobiose (CHB) as a Michaelis complex [1], and *Sp*HEX bound to the intermediate analogue NAG-thiazoline (NGT). *Sp*HEX carbon atoms are coloured *grey*, while *Sm*CHB carbon atoms are in *orange*. Asp539 of *Sm*CHB and Asp313 of *Sp*HEX accept a hydrogen bond from N2 of the substrate and the reaction intermediate, respectively. A conserved hydrogen bonding network between *Sp*HEX Asp313 (*Sm*CHB Asp539), Asp246 (*Sm*CHB Asp448) and the backbone nitrogen of Gly312 (*Sm*CHB Gly538) helps to maintain the positions of these residues in the active sites of these enzymes.

verify these hypotheses, wild type *Sp*HEX and two variant forms of the enzyme, Asp313Ala (*Sp*HEX-D313A) and Asp313Asn (*Sp*HEX-D313N), were expressed, purified and analyzed kinetically and structurally. The two *Sp*HEX variants were found to have dramatically different catalytic activities as compared to each other and to the wild type enzyme. By determining the crystal structure of wild type *Sp*HEX and each variant in complex with GlcNAc bound as product, the molecular basis for the altered kinetic profiles of each variant was established. Furthermore, these complexes are the first crystal structure determinations of a family 20 glycosidase bound to product, and together with the other studies, they complete the structural analysis of every stable species that occurs along the reaction coordinate of family 20 glycosidases.

All crystallographic complexes described in this chapter were determined by the author of this thesis; however, Spencer Williams and David Vocadlo in the laboratory of Stephen Withers, University of British Columbia, constructed the *Sp*HEX expression vector variants D313A and D313N, and measured all of the enzyme kinetics. Thus, this chapter focuses on the X-ray crystallographic component of this research collaboration, and briefly compares the results of the protein structure determinations with the enzyme kinetic data collected by Drs. Williams and Vocadlo. The research project has been published in its entirety as a paper in the *Journal of Biological Chemistry* [3].

B. Experimental Procedures

B.1. Mutagenesis

Site-directed mutagenesis was carried out by Spencer Williams and David Vocadlo using the *Sp*HEX expression vector p3AHEX-1.8 as template [2]. This expression vector was provided to them by the author of this thesis, and was originally constructed to produce large quantities of *N*-terminal 7xHis tagged *Sp*HEX for X-ray diffraction studies (see Chapter 2, section **A.1**) [2]. As described by Williams et al. [3], the p3AHEX-1.8 vector was mutated using the gene Splicing by Overlap Extension (gene SOEing) method [4], and the resulting

mutant PCR amplicons were subsequently used to construct two new *Sp*HEX expression vectors: p3AHEX-1.8D313A and p3AHEX-1.8D313N [3].

B.2. Protein expression and purification

Wild type *Sp*HEX and the two site-directed variant forms of the enzyme SpHEX-D313A and SpHEX-D313N were overexpressed as N-terminal His_7-tagged fusion proteins in *Escherichia coli* strain BL21 (DE3). The recombinant proteins were affinity purified from the soluble fraction of the bacterial cell lysates using nickel-nitrilotriacetic acid Superflow resin (Qiagen) as described previously [2] (see Chapter 2, section **A.1**). The purified wild type and variant forms of *Sp*HEX were stored at 4°C in elution buffer (20 mM Tris-Cl, pH 8.0, 300 mM NaCl, 10 mM β-mercaptoethanol, 250 mM imidazole pH 8.0) and were stable for several weeks.

B.3. Kinetic Analysis

The *Sp*HEX-catalyzed rates of hydrolysis of 3,5-DNPGlcNAc were measued by Spencer Williams and David Vocadlo as described in [3]. Briefly, all measurements were conducted at 25°C using 25 mM NaH_2PO_4/25 mM sodium citrate buffer (pH 6.0) containing 100 mM NaCl. The linear increase in absoption at 400 nm was monitored continuously using a Pye Unicam spectrophotometer, and the Michaelis-Menton parameters (V_{max} and K_m) were extracted from these data by non-linear least squares fitting to the Michaelis-Menton equation using the program Grafit 4.0 [5].

B.4. Crystallization and data collection

Purified enzyme was dialyzed against 50 mM trisodium citrate, pH 5.0 and 300 mM NaCl and concentrated to approximately 10 mg/ml with a Millipore concentrator. Wild type *Sp*HEX and the two variants were co-crystallized with GlcNAc using the vapor diffusion method at room temperature. Concentrated wild type *Sp*HEX was incubated with 30 mM GlcNAc for 30 min. and then mixed in a 1:1 ratio with mother liquor (2.1 M ammonium sulphate, 100 mM trisodium

citrate, pH 6.0 and 20% glycerol) to create several 8 μl hanging drops from which diffraction quality hexagonal bipyramidal crystals ($P6_122$) were obtained within 2 weeks. The *Sp*HEX-D313A and *Sp*HEX-D313N variants were incubated with 250 mM and 500 mM GlcNAc, respectively, and then individually mixed with same as mother liquor described above but lacking glycerol. Glycerol, which was used as a cryoprotectant for low-temperature data collection, was found to compete with GlcNAc for the active sites of both variant forms of *Sp*HEX. Thus, crystals of the two *Sp*HEX variants in complex with GlcNAc were grown in the absence of glycerol and swept through a cryoprotectant (2.1 M ammonium sulphate, 100 mM trisodium citrate, pH 6.0 and 20% glycerol) for 10 sec just prior to data collection.

Diffraction data from a single crystal of the wild type *Sp*HEX:GlcNAc complex were collected at the Stanford Synchrotron Radiation Laboratory (SSRL) beamline 9-1, whereas diffraction data from single crystals of the *Sp*HEX-D313A:GlcNAc and *Sp*HEX-D313N:GlcNAc complexes were collected with an R-axis IV^{++} image plate detector mounted on a Rigaku rotating anode X-ray generator. All crystals were flash-cooled to 100 K within a N_2(g) stream during data collection, and all diffraction data were processed using DENZO and SCALEPACK [6].

B.5. Structure determination and refinement

Prior to structure determination and refinement, 10% of the diffraction data from the wild type *Sp*HEX:GlcNAc complex were randomly flagged for cross-validation using the free *R* factor [7]. To remain consistent, the same *hkl* indices were also included in cross-validation calculations for the *Sp*HEX-D313A:GlcNAc and *Sp*HEX-D313N:GlcNAc complexes. All crystallographic refinements were carried out using the CNS program [8]. The three complexes were found to crystallize isomorphously with wild-type *Sp*HEX. Thus, using a maximum likelihood target function, rigid body refinement and several rounds of conjugate gradient minimization, were sufficient to position molecular models of wild-type *Sp*HEX (PDB code: 1HP4) and the variants (D313A and D313N mutations

created using the program O [9]) into the asymmetric unit of the appropriate crystallographic complex.

Initial sigma-A weighted $|F_o|-|F_c|$, α_c and $2|F_o|-|F_c|$, α_c electron density maps (where α_c are the calculated phases and $|F_o|$ and $|F_c|$ are the measured and calculated structure factor, respectively), computed prior to solvent modeling, unambiguously defined a GlcNAc molecule bound in the –1 subsite of each of the three complexes. These maps also showed clearly two alternate conformations for the GlcNAc molecule bound to the *Sp*HEX-D313A mutant. GlcNAc models were fit into the appropriate electron density of each complex using the program O, and topology and parameter files of the sugar were generated using Xplo2D [10]. Solvent molecules, which had been removed from all three models prior to positioning into the new asymmetric units, were relocated using the CNS program and manually inspected using the program O. *R*-factor convergence was reached following several subsequent rounds of solvent modeling and maximum likelihood conjugate gradient minimization. Once the atomic B-factors for the *Sp*HEX-D313A:GlcNAc complex had been refined, their values were fixed and the occupancies for the alternate GlcNAc conformations were then refined. The final refinement statistics are presented in Table 4.1. The stereochemical quality of each model was verified using the programs PROCHECK [11] and WHAT_CHECK [12].

B.6. Coordinates

Coordinates and structure factors have been deposited into the Protein Data Bank (accession codes: wild type *Sp*HEX:GlcNAc complex, 1M01; *Sp*HEX-D313A:GlcNAc complex, 1M03; *Sp*HEX-D313N:GlcNAc complex, 1M04).

C. Results and Discussion

C.1. Structural analysis of the SpHEX product complexs

The electron density for the GlcNAc product bound in the –1 subsite of wild-type *Sp*HEX was very well defined (Figure 4.2a). As expected, the C2-acetamido group is positioned beneath the α-face of the pyranose ring and

Table 4.1 Crystallographic statistics

	WT_NAG	D313A_NAG	D313N_NAG
Crystal Information			
Space group	$P6_122$	$P6_122$	$P6_122$
Unit cell dimensions (Å)	a = b = 133.3 c = 176.2	a = b = 132.9 c = 176.9	a = b = 133.7 c = 176.1
Data Collection (values in parentheses refer to the high-resolution shell)			
Detector	MAR345 image plate	RaxisIV^{++} image plate	RaxisIV^{++} image plate
Wavelength (Å)	0.979	1.54	1.54
Resolution (Å)	70.0 - 2.10	40.0 - 1.90	40.0 - 1.95
High-resolution (Å)	2.14 - 2.10	1.96 - 1.90	2.02 - 1.95
Total observations	747643	1251130	1300022
Unique reflections	54264 (2604)	70431 (5572)	65848 (6556)
$\langle I / \sigma \rangle$	39.4 (15.1)	30.7 (5.8)	30.5 (5.5)
Completeness[a] (%)	99.8 (97.4)	96.5 (93.3)	96.9 (98.5)
R_{sym}[b]	0.038 (0.100)	0.060 (0.189)	0.056 (0.256)
Refinement			
Resolution (Å)	70.0 - 2.1	38.4 - 1.90	39.2 - 1.95
R_{work}[c]	0.195	0.194	0.198
R_{free}[c]	0.216	0.214	0.218
Number of atoms			
Protein	3864	3861	3864
heterogen	41	50	29
Water	280	295	287
Average B (Å^2)	22.0	17.9	24.6
RMSD ideal geometery			
bond lengths (Å)	0.005	0.005	0.005
bond angles (°)	1.3	1.3	1.3
Ramachandran plot[d]			
%-most favoured	89.9	90.9	90.1
%-additionally allowed	10.1	9.1	9.9

[a]Calculated by treating Bijvoet pairs as equivalent.
[b]$R_{sym} = \Sigma_{\mathbf{h}} \Sigma_i (|I_i(\mathbf{h}) - \langle I(\mathbf{h})\rangle|) / \Sigma_{\mathbf{h}} \Sigma_i I_i(\mathbf{h})$, where $I_i(\mathbf{h})$ is the ith intensity measurement and $\langle I(\mathbf{h})\rangle$ is the weighted mean of all measurements of $I(\mathbf{h})$.
[c]R_{work} and $R_{free} = \Sigma_{\mathbf{h}} ||F(\mathbf{h})_{obs}| - |F(\mathbf{h})_{calc}|| / |F(\mathbf{h})_{obs}|$ for reflections in the working and test sets (10% of all data) respectively.
[d]Regions defined by PROCHECK [11]

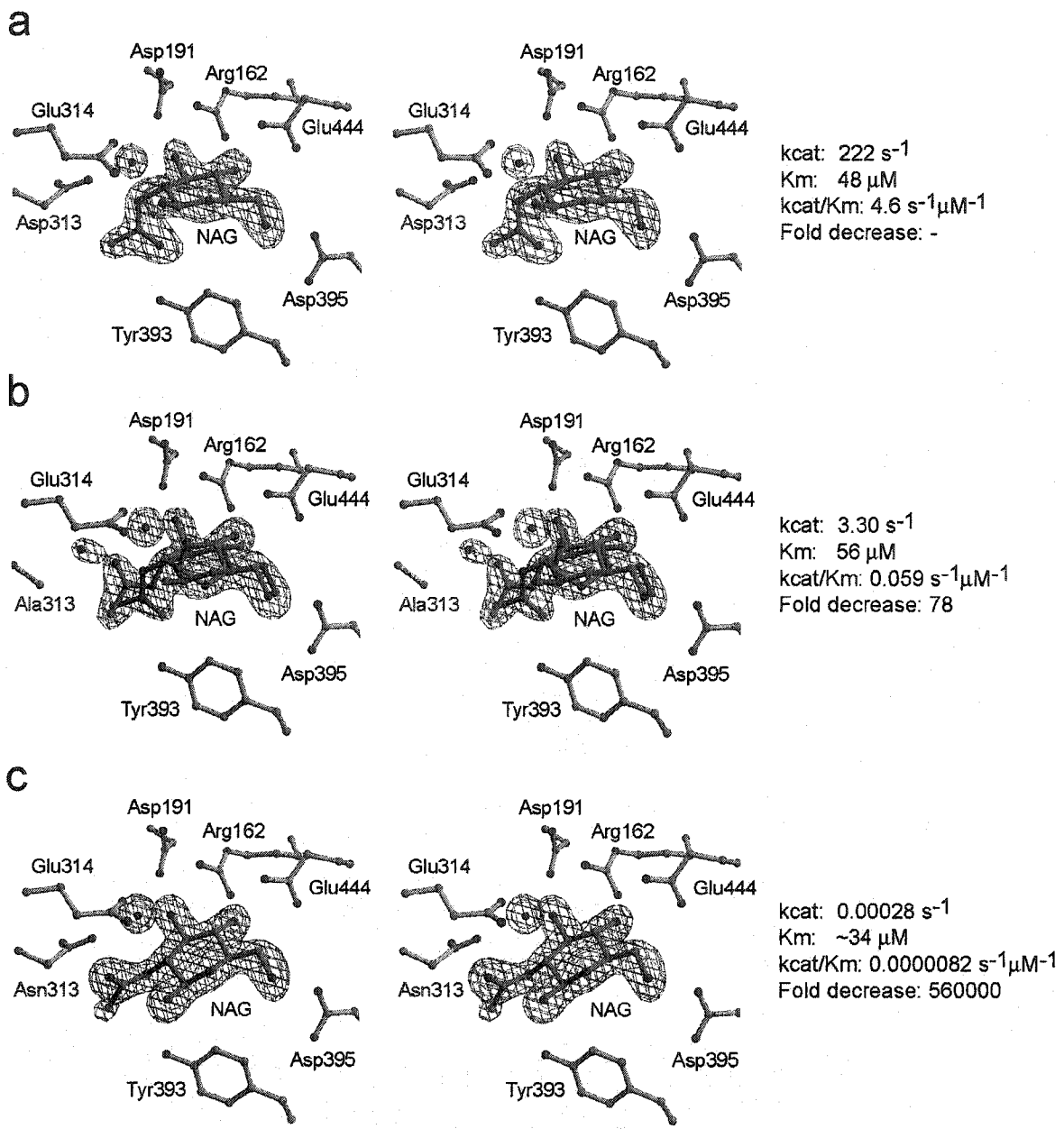


Figure 4.2 Stereographic representations of GlcNAc bound in the active site of wild-type *Sp*HEX (a), *Sp*HEX D313A (b) and *Sp*HEX D313N (c) and their corresponding kinetic parameters*. The ball-and-stick models shown in each panel represent the final refined coordinates of each complex. The initial sigma-A weighted |Fo|-|Fc|, α_C difference electron electron density, calculated prior to modeling the NAG residue and solvent of each complex, is shown in *blue*. *a*, *Sp*HEX carbon atoms are shown in *grey*, whereas the carbon atoms of NAG are shown in *green* (density contoured at 2.4 σ). The Asp 313, is labeled in *red*. *b*, *Sp*HEX carbon atoms shown in *grey*. The mutated residue Ala313 is labeled in *red*. The dominant GlcNAc conformation in the complex (overall occupancy of 0.6) is drawn with *green* carbon atoms. The alternate conformation, which is more catalytically favorable, has an overall occupancy of 0.4 and is drawn with *brown* carbon atoms (density contoured at 2.2 σ). *c*, *Sp*HEX carbon atoms shown in *grey*. Carbon atoms of NAG are shown in *green* (density contoured at 2.4 σ). The mutated residue, Asn 313, is labeled in *red*.* *Kinetics measured by Spencer Williams and David Vocadlo (see section A.3)*

partially held into position by hydrogen-bonding interactions with Asp313 and Tyr393. The position of the acetamido group brings the carbonyl oxygen of this function group to within 2.5 Å of the anomeric carbon, C1 (Figure 2.4a). Overall, hydrogen-bonding contacts exist between all non-carbon atoms except the ring oxygen O5, including a 2.5 Å hydrogen bonding contact between the hydroxyl group on C1 and the carboxyl group of the general acid/base residue Glu314.

Although the bound sugar ring is not in a totally relaxed $^4C_1$ chair conformation, it is also not distorted to the same extent as seen for the skew-boat conformation of the analogous sugar found in the Michaelis complex between chitobiose and *Sm*CHB. Indeed, the pyranose ring of the bound product is very similar in conformation to the piperidine ring of GalNAc-isofagomine bound to *Sp*HEX (Chapter 3). The difference in conformation of the sugar ring bound in the –1 subsite of the Michaelis verses the product complex supports the theory that enzyme-substrate interactions at the +1 subsite contribute significantly to the distortion of the pyranoside ring in the –1 subsite, positioning the scissile bond and leaving-group in a pseudo-axial position prior to bond cleavage. In the absence of an intact scissile bond to a sugar bound in the +1 subsite, the conformation of pyranose ring of the GlcNAc product is not directly influenced by interactions occuring at the +1 sugar binding subsite.

When Asp313 is mutated to an Ala in *Sp*HEX, a pocket is created in the active site that allows the C2-acetamido group of a GlcNAc residue in the –1 subsite to bind in two alternate conformations (Figure 4.2b). The difference electron density for the bound GlcNAc at first glance appears quite unusual; however, upon closer inspection, the density can be understood to arise from the C2-acetamido group existing in two alternate conformations in the crystal structure. One conformation, with a refined occupancy of 0.4, is similar to that seen for GlcNAc bound as product to wild-type *Sp*HEX, and appears to be compatible with catalysis, as evidenced by the substantial level of residual catalytic activity displayed by the *Sp*HEX-D313A variant. The alternate confomer however, with a refined occupancy of 0.6, dominates the crystal structure and

appears not to be compatible with catalysis; the C2-acetamido group is swung out from underneath the α-face of the sugar ring and into a position that would otherwise be occupied by the caboxylate of Asp313. Electron density cannot be seen for the anomeric hydroxyl group of either GlcNAc conformation in this complex, and may reflect large motions of this functional group. The refined positions of the anomeric hydroxyl groups in the models of the alternate GlcNAc conformations resulted from an optimization of the geometry of all the remaining atoms for which there was clear electron density. For the dominant conformation, the pyranose ring refined into a $^4B^1$ boat, whereas for the alternate conformation, the pyranose ring is close to a $^4C_1$ as observed for GlcNAc bound to wild-type *Sp*HEX.

The crystal structure of *Sp*HEX-D313A in complex with GlcNAc supports earlier findings in which the analogous mutation (D539A) was created in *Sm*CHB. When this *Sm*CHB variant was determined as a Michaelis complex with chitobiose, it was observed that the C2-acetamido group of the non-reducing sugar of chitobiose also occupied the position of the missing Asp caboxylate; however, this was the only substrate conformation observed in the complex. Together, these studies indicate that in addition to helping polarize the C2-acetamido group prior to nucleophilic attack, and stabilizing the positive charge that develops on the oxazolinium ring upon cyclization, this conserved Asp residue is important for steering the C2-acetamido group into a position that allows it to provide anchimeric assistance to the catalytic reaction.

As for GlcNAc bound to *Sp*HEX-D313N, the pyranose ring of the sugar is in a $^4C_1$ conformation; however, it is tilted about an axis best defined by atoms C3 and C5 such that the anomeric carbon atom C1 drops 1.1 Å deeper into the active site pocket as compared to the wild type product complex (Figure 4.2c). Although the side chain of Asn313 is in a nearly identical position to that of Asp313 in the wild type complex, the additional bulkiness of the uncharged $CONH_2$ group appears to interfere with the proper positioning of the C2-acetamido group within the active site: the functional group is rotated 117° about the C2-N bond relative to the wild type complex so that the carbonyl oxygen atom

becomes positioned slightly above the plane of the pyranose ring (Figure 4.2 & 4.3). Interestingly, the conformation of GlcNAc bound to this *Sp*HEX variant is very similar to that of the non-reducing sugar of the unbound, small molecule crystal structure of chitobiose [13]. Unfortunately for the *Sp*HEX variant however, this conformation does not allow the C2-acetamido group to participate in catalysis, and because the electron density does not indicate any alternate conformations that would be compatible with the substrate-assisted catalytic mechanism proposed for family 20 enzymes, the crystallographic complex suggests that *Sp*HEX-D313N would be devoid of catalytic activity. Indeed, this conjecture is consistent with kinetic measurements described in section **B.2** below.

An electron density omit map, calculated after refinement of a D313A variant model of *Sp*HEX using structure factor amplitudes from the SpHEX-D313N product complex, indicates that the $CONH_2$ group is oriented with $N^{\delta 2}$ pointing toward the C2-acetamido group, while the carbonyl oxygen atom of the amide side chain shares a proton through a short 2.5 Å hydrogen bond with the carboxyl group of Asp246 (Figure 4.4). Furthermore, difference density and B-factors calculated using a model of the D313N variant refined with the amide of Asn313 flipped 180°, revealed positive $|F_o|-|F_c|$ electron density about the Asn side chain atom interacting with the carboxyl group of Asp246, and negative $|F_o|-|F_c|$ density about the Asn313 atom pointing towards the bound product (Figure 4.4). Thus, the diffraction data were satisfactory enough to discern an electron distribution difference within the Asn313 side chain, and this difference provides evidence to conclude that the $N^{\delta 2}$ atom of Asn313 points into the active site pocket towards the C2-acetamido group of the bound sugar. Finally, the orientation of the Asn313 side chain allows for a hydrogen bonding interaction to form between $N^{\delta 2}$ and the carbonyl oxygen atom of the C2-acetamido group of the bound sugar. Indeed, the attractive strength of this interaction is evidenced by a 24° rotation of the carbonyl oxygen atom out of the plane of the C2-acetamido group and toward atom $N^{\delta 2}$ of Asn313.

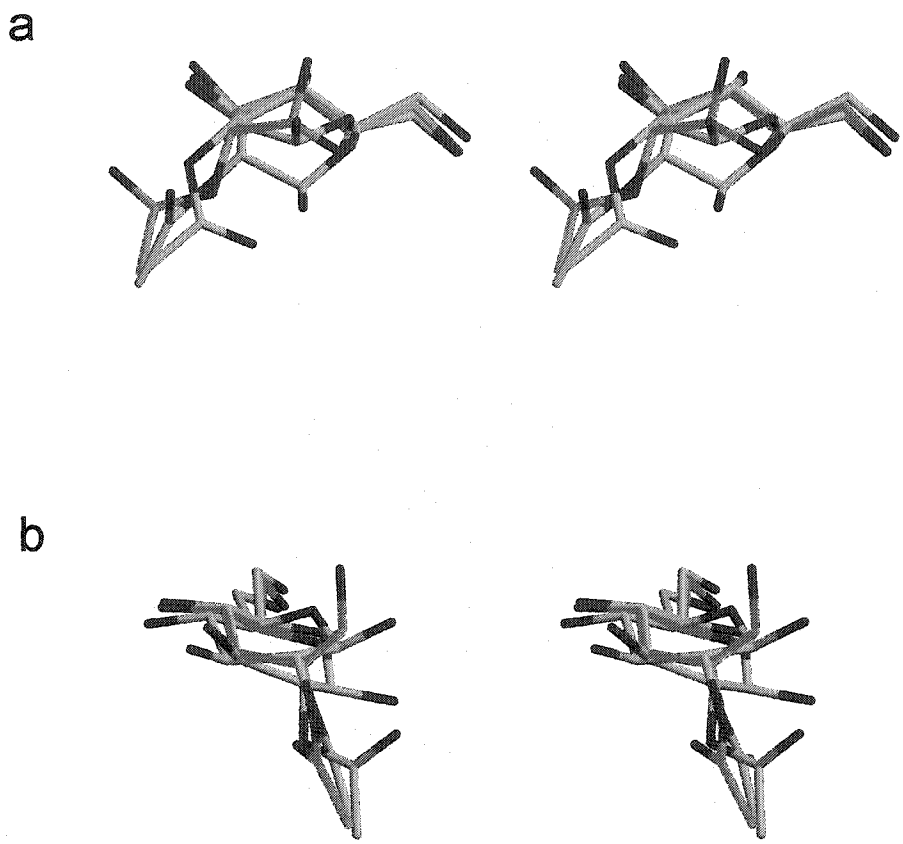


Figure 4.3. Stereographic superposition of GlcNAc as bound in the active site of wild-type *Sp*HEX (*grey*), *Sp*HEX-D313A (*green*), and *Sp*HEX-D313N (*yellow*). Of the two GlcNAc conformations refined in the *Sp*HEX-D313A-GlcNAc complex, only the catalytically incompetent conformer is shown. *Panels A* and *B* are oriented ~90° about the y-axis with respect to each other.

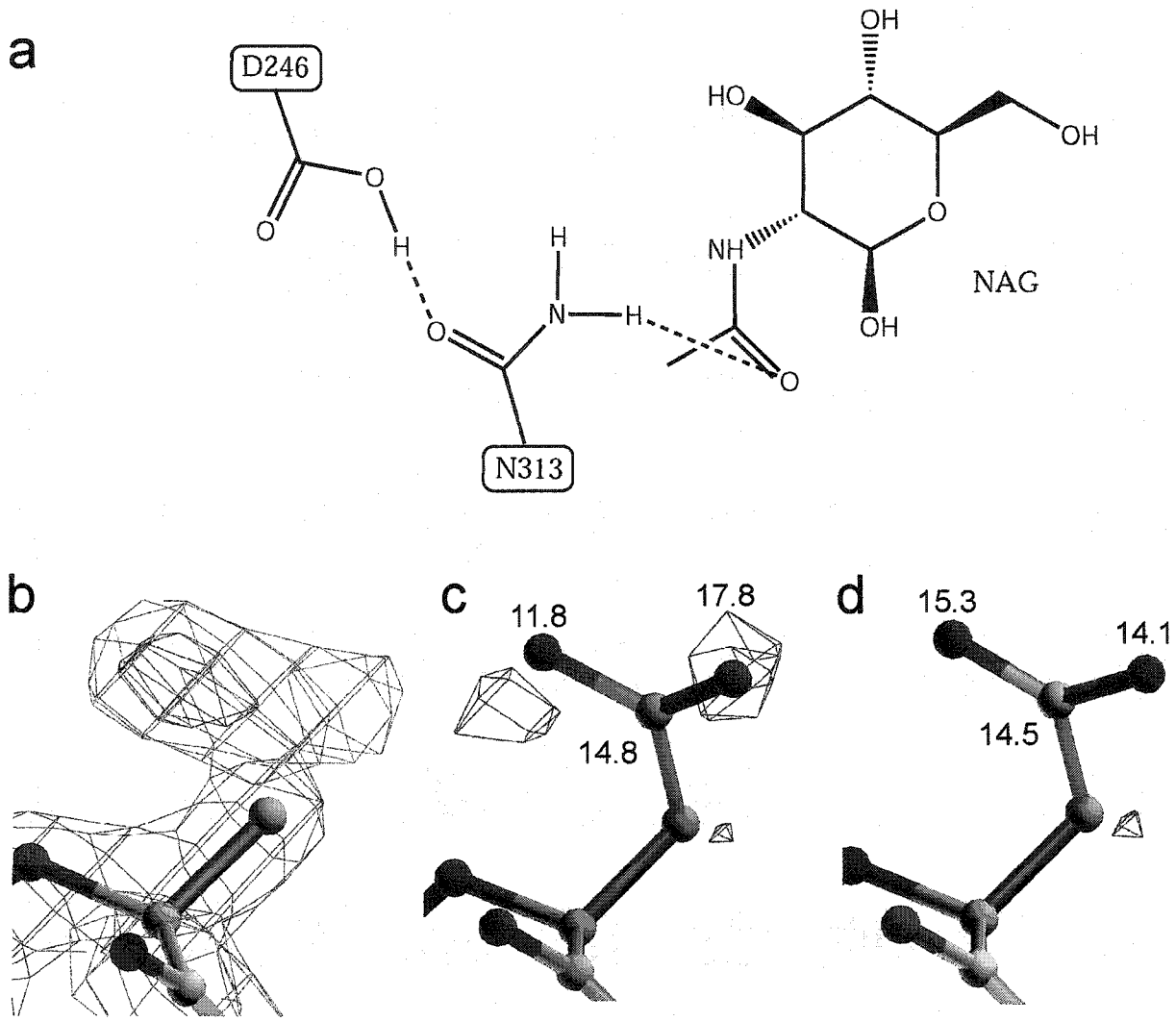


Figure 4.4. Difference electron density and B-factor analysis for the D313N *Sp*HEX variant. *Panel a*, a schematic of the proposed orientation of the amide of N313 and its interactions with Asp246 and the C2-acetamido group of the bound product (GlcNAc). *Panel b*, 2|Fo|-|Fc| density (*red*) and positive |Fo|-|Fc| density (*blue*) of an omit map calculated after refinement of a D313A variant model of *Sp*HEX using structure factor amplitudes from the D313N variant complex. The positive |Fo|-|Fc| density indicates that the oxygen atom of the $CONH_2$ group lies to the *left* and is involved in a short 2.5 Angstrom hydrogen bond with Asp246 as shown in *panel a*. *Panel c*, |Fo|-|Fc| density map and side-chain B-factors calculated after refinement of a model of the D313N variant where the $CONH_2$ group was rotated 180°. The difference density peaks and B-factors suggest that the original orientation of the amide is correct when compared to *panel d*. *Panel d* |Fo|-|Fc| density map and B-factors for the original and final D313N variant model of *Sp*HEX. The lack of difference peaks when contoured to the same σ level as in *panel c* confirms that this orientation is correct. The electron density in panels *c* and *d* are contured as follows: blue density, 2.7 σ; red density, -2.0 σ.

C.2. Kinetic analysis of wild-type and the SpHEX variants D313A and D313N

Kinetic parameters for wild-type *Sp*HEX and the two variants were determined by Drs. Williams and Vocadlo using 3,5-DNPGlcNAc as substrate (section **A.3**) (Figure 4.2). Although the K_m values are quite similar for the wild-type and mutated enzymes, their k_{cat} values are markedly different: 222 s^{-1} for wild type *Sp*HEX, 3.30 s^{-1} for *Sp*HEX-D313A and 0.00028 s^{-1} for *Sp*HEX-D313N. These k_{cat} values correlate very well with the crystallographic observations obtained from the product complexes described above. D313N was at first considered to be a conservative mutation; however, as described above, the slight increase in steric bulk and loss of negative charge resulting from this mutation, was found to reduce substrate turnover far more dramatically when compared to the obviously less conservative mutation D313A. During catalysis, the negatively charged carboxylate of Asp313 helps to stabilize the cyclic intermediate and possibly the transition states that flank it. Both D313A and D313N remove this carboxylate functionality, and thus an important negative charge is also removed from the *Sp*HEX active site. Why then is *Sp*HEX-D313A only 78-fold less active than the wild-type enzyme, while the D313N variant has a catastrophic effect on enzymatic activity, reducing the catalytic rate ~560000 fold? The reason for the huge difference appears to be provided by the crystal structures of their product complexes; although a significant proportion of the GlcNAc bound to *Sp*HEX-D313A can adopt a catalytically active conformation, the increased steric bulk introduced by the Asn313 side chain in *Sp*HEX-D313N does not allow the product (or substrate) to access such a conformation, and the C2-acetamido is essentially always misdirected into a catalytically unfavorable position (Figure 4.2c).

Interestingly, the D539A variant of *Sm*CHB, which is analogous to *Sp*HEX-D313A, has a 1625-fold reduction in catalytic activity compared to wild-type *Sm*CHB when using pNPGlcNAc as substrate [14]. This is a substantially greater rate reduction than what was measured for *Sp*HEX-D313A; however, this difference also appears to correlate well with structural observations. As

opposed to the *Sp*HEX-D313A product complex, where the C2-acetamido group of bound GlcNAc is observed to adopt two conformations, one of which is compatible with catalysis, the crystal structure of *Sm*CHB-D539A in complex with chitobiose demonstrates that the C2-acetamido group of the non-reducing sugar of chitobiose does not adopt the catalytically active confomer, it appears only to occupy the position of the missing Asp carboxyl group.

If, according to crystallographic observations, the C2-acetamido group does not participate in *Sp*HEX-D313N-catalyzed reactions, this variant may use an alternative catalytic mechanism. Indeed, the catalytic activity of this variant could not be not rescued by an exogenous azide nucleophile as could *Sp*HEX-D313A. However, Williams and Vocadlo found that the addition of azide to the reaction mixture enhanced the rate of *Sp*HEX-D313A-catalyzed hydrolysis of 3,5-DNPGlcNAc, and they concluded that the observed chemical rescue was resulting from the direct nucleophilic attack of azide at the anomeric center of the oxazolinium ion intermedate, and thus forming an azide substituted product. Indeed, they were able to isolate such a product and verified its structure by TLC and ^{1}H NMR [3]. Because the catalytic rate of the D313N could not be accelerated by the addition of azide, this variant may not form an oxazolinium ion intermediate; however, the K_m of the variant is very similar to wild-type *Sp*HEX, and this prompted Williams and Vocadlo to conclude that the residual activity may be the result of wild-type *Sp*HEX contamination via spontaneous deamination or translational misincorporation of Asp. The analogous mutation in human Hex B (D354N) was found to reduce the catalytic rate of this enzyme by only 2500-fold [15], not 560000-fold as observed for *Sp*HEX; however, because the enzyme preparation was known to contain at least 0.01% wild-type Hex B [15], Williams and Vocadlo suggest that the D354N mutation is just as deleterious to human Hex B as is the analogous mutation to *Sp*HEX, and that this higher residual activity is due to a high concentration of wild-type enzyme in the Hex B preparation as compared to the *Sp*HEX-D313N preparation [3].

Family 18, 20 and 56 glycosidases all use the substrate-assisted catalytic mechanism illustrated in the lower pathway of Figure 1.2b, and each contain a

conserved Asp residue that is analogous to Asp313 of *Sp*HEX. The evolutionary conservation of this residue suggests that it plays a crucial role in the catalytic mechanism of these enzymes. Indeed, research presented here, and by others, demonstrates that this Asp residue has a least two important catalytic functions: 1) it helps orient the neighboring C2-acetamido of the substrate so that it can provide anchimeric assistance to the catalytic reaction, and 2) the environment surrounding the conserved Asp ensures that its carboxyl group maintains a negative charge, which in turn provides a favorable electrostatic environment that stabilizes the oxazolinium ion intermediate and quite possibly the transition states that lead to its formation and breakdown [3].

D. References

1. Tews, I., et al., *Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease.* Nat Struct Biol, 1996. **3**(7): p. 638-48.
2. Mark, B.L., et al., *Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase.* Journal of Biological Chemistry, 2001. **276**(13): p. 10330-10337.
3. Williams, S.J., et al., *Aspartate 313 in the Sterptomyces plicatus hexosaminidase plays a critical role in substrate assisted catalysis by orienting the 2-acetamido group and stabilizing the transition state.* Journal of Biological Chemistry, 2002. **(in press)**.
4. Horton, R.M., et al., *Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction.* Biotechniques, 1990. **8**(5): p. 528-35.
5. Leatherbarrow, R.J., *GraFit version 4.0. Data Analysis and Graphics program for the IBM PC.* 1998, Erithacus Software Ltd.
6. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Methods Enzymol, 1997. **276**: p. 307-326.
7. Brunger, A.T., *Free R-Value - a Novel Statistical Quantity for Assessing the Accuracy of Crystal-Structures.* Nature, 1992. **355**(6359): p. 472-475.

8. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallogr D, 1998. **54**(Pt 5): p. 905-21.
9. Jones, T.A., et al., *Improved methods for the building of protein models in electron density maps and the location of errors in these models.* Acta Crystallogr A, 1991. **A47**(Pt 2): p. 110-119.
10. Kleywegt, G.J., *Dictionaries for Heteros.* CCP4/ESF-EACBM Newsletter on Protein Crystallography, 1995. **31**: p. 45-50.
11. Laskowski, R.A., et al., *PROCHECK: a program to check the stereochemical quality of protein structure coordinates.* Journal of Applied Crystallography, 1993. **26**: p. 283-291.
12. Vriend, G., *WHAT IF: a molecular modeling abd drug design program.* J. Mol. Graph., 1990. **8**: p. 52-56.
13. Mo, F., Jensen, L.H., Acta Crystallogr B, 1978. **34**: p. 1562.
14. Prag, G., et al., *Structures of chitobiase mutants complexed with the substrate Di-N-acetyl-d-glucosamine: the catalytic role of the conserved acidic pair, aspartate 539 and glutamate 540.* J Mol Biol, 2000. **300**(3): p. 611-7.
15. Hou, Y., et al., *Characterization of Glu and Asp residues in the active site of human b-hexosaminidase B.* Biochemistry, 2001. **40**: p. 2201-2209.

Chapter 5

Crystal structure of human β-hexosaminidase B: understanding the molecular basis of Tay-Sachs and Sandhoff disease

A. Introduction

Most of our understanding about the structural and functional relationships inherent in family 20 glycosidases has come from the study of prokaryotic enzymes. This is especially true in regards to our understanding of the molecular basis of how these enzymes catalyze glycosidic bond hydrolysis. Studying prokaryotic family 20 glycosidases alone however, places a limit on what there is to know about the whole enzyme family. The biochemistry of eukaryotic family 20 enzymes demonstrates that they contain unique molecular features not found in prokaryotic family 20 glycosidases.

Perhaps the foremost difference between prokaryotic and eukaryotic family 20 glycosidases is that unlike prokaryotic family 20 enzymes, which exist as active monomers, eukaryotic family 20 enzymes dimerize to form a variety of functionally distinct isoenzymes. Furthermore, eukaryotic family 20 glycosidases hydrolyze glycolipid substrates that need to be solublized and presented to the isoenzymes by lipid transport proteins, a level of complexity not found in prokaryotic enzymes. Unfortunately, significant sequence divergence between prokaryotic and eukaryotic family 20 glycosidases limits the amount of insight that can be gained from comparative molecular modeling studies of the eukaryotic enzymes. For example, the aa sequence identity between active site regions of the prokaryotic enzymes *Sp*HEX or *Sm*CHB and the active site region of human Hex is less than 30%, and quickly drops to the level of noise elsewhere. Comparative molecular modeling studies that use these enzymes as structural templates have not been able to answer questions concerning the mechanisms of G_{M2}-ganglioside processing or the protein:protein interactions involved in isoezyme dimerization. Understanding the structural basis of subunit

dimerization is of particular interest because it is the mechanism that initiates the catalytic activity of the human Hex isoenzymes.

Presented here is the X-ray crystal structure of human Hex B, the first crystal structure to be determined for a eukaryotic family 20 glycosidase. The project originally began in the late 1980's, using native protein isolated directly from human placenta in the laboratory for Dr. Don Mahuran, University of Toronto. By 1992, crystals of the isoenzyme had been grown and were found to diffract to moderate resolution (3.2 Å) at room temperature using synchrotron radiation [1]. The large unit cell dimensions of the Hex B crystals ($a = b = 114.2$ Å, c=402.2 Å) necessitated the use of synchrotron radiation during data acquisition, and screening for heavy atom derivatives proved difficult, especially at a time when access to synchrotron radiation for protein crystallography experiments was limited. These crystallographic challenges were compounded by serious technical problems associated with the purification of Hex B, and interest in the project began to wane until is was essentially dropped in the early 1990's. In 2000, the project was revitalized with a renewed supply of placental Hex B from Dr. Mahuran's laboratory, and the belief that recent technological advances such as cryo-crystallography and access to third generation synchrotron sources could overcome many of the problems that had faced the project in its early years.

Indeed, these modern technologies were instrumental to the successful determination of the Hex B crystal structure, which now provides exciting new insight into the molecular structure of eukaryotic family 20 glycosidases and human disease. The Hex B structure, currently refined to 2.2 Å resolution, has been determined alone and in complex the mechanistic inhibitors GalNAc-isofagomine and NAG-thiazoline. From these, and the known X-ray structure of the G_{M2}-activator, a molecular model of Hex A in complex with the activator and G_{M2}-ganglioside has been built. Together, the crystallographic and modeling data demonstrate how the α and β subunits dimerize to form Hex A or Hex B, how these isoenzymes hydrolyze diverse substrates, and how many documented

point mutations cause Sandhoff disease (β-subunit mutations) and Tay-Sachs disease (α-subunit mutations).

B. Experimental Procedures

B.1. Purification and crystallization

Hex B was purified from human placenta by Amy Leung in the laboratory of Don Mahuran using the methods described in [2], and crystallized from ammonium sulfate solutions using the vapor diffusion method as described previously [1]. The protein was used in its native glycosylation state for all crystallization experiments. Ellipsoidal crystals having a hexagonal cross-section perpendicular to the longest crystal axis (space group $P6_122$) appeared after about 3 weeks from 6 μl hanging drops and were used as seeds in fresh crystallizations systems that produced large, diffraction quality crystals within 6 months.

B.2. Structure determination

All diffraction data were collected at the Advanced Photon Source beamline 14-BM-C (BioCars) from single crystals that were briefly soaked (10 sec.) in cryosolvent (30% glycerol, 60% Ammonium sulphate, 50 mM potassium phosphate, pH 8.0) then flash-cooled to 100 K within a N_2 (g) stream. Intensity data were processed using DENZO and SCALEPACK [3]. Structure factor phases were determined experimentally using the multiple isomorphous replacement (MIR) method. Crystals were derivatized by soaking them overnight in solutions of 3.6 mM methyl mercury acetate ($CH_3HgCOCH_3$) or 2.6 mM potassium platinum tetra-chloride (K_2PtCl_4). Difference Patterson and Fourier searches of diffraction data collected from native and heavy atom derivatized Hex B crystals were carried out using the program SOLVE [4]. Two sites were located within the asymmetric unit for each heavy atom type. Each derivative was of sufficient quality to calculate an MIR map that clearly defined the overall shape of the Hex B subunits to a resolution of 2.8 Å. Statistical density modification using the program RESOLVE [5] greatly improved the quality of the

Table 5.1 Crystallographic statistics

Crystal information					
Space group	$P6_122$				
Matthews Coefficient	3.24 (2 molecules / Asymmetric unit)				
Solvent content (%)	~ 62				
Data Collection (values in parentheses refer to the high-resolution shell)					
Data Set	Native	Pt	Hg	NGT complex	IFG complex
Unit cell dimensions (Å)	a=b=112.47 c = 397.87	a =b=111.52 c = 398.23	a=b=111.70 c = 396.80	a=b=112.40 c = 397.23	a=b=112.39 c = 397.30
Wavelength (Å)	1.00	1.00	1.00	0.900	0.900
Resolution range (Å)	35.0 – 2.40	35.0 – 3.00	35.0 – 2.90	37.0 – 2.50	35.0 – 2.20
High-resolution (Å)	2.44 – 2.40	3.05 – 3.00	2.95 – 2.90	2.57 – 2.50	2.26 – 2.20
Total observations	2772717	800088	769328	784049	1153759
Unique reflections	58844 (2894)	33068 (1460)	30367 (1644)	52315 (4235)	74245 (6161)
$\langle I / \sigma \rangle$	21.9 (4.89)	22.9 (7.73)	20.8 (6.31)	23.2 (5.08)	22.3 (4.93)
Completeness[a] (%)	99.5 (99.7)	99.5 (99.9)	97.0 (98.6)	99.7 (100.0)	99.4 (100.0)
R_{sym}[b]	0.090 (0.314)	0.071 (0.272)	0.081 (0.314)	0.077 (0.367)	0.094 (0.457)
R_{iso}[c]		0.207 (0.209)	0.175 (0.199)		
Heavy atom sites		2	2		
Phasing power					
Centric		0.46	0.48		
Acentric		0.55	0.52		
Overall Figure of Merit					
Solve	0.29 (Overall Z-score = 14.75)				
Resolve	0.57 (with 2-fold non-crystallographic symmetry restraints)				
Refinement					
R_{work}[d]	0.201			0.202	0.193
R_{free}[d]	0.231			0.229	0.218
Number of atoms					
Protein	7762			7762	7762
Heterogen	79			107	107
Water	226			244	321
Average B ($Å^2$)	25.1			24.2	20.9
RMSD from ideal geometry					
bond lengths (Å)	0.011			0.009	0.011
bond angles (°)	1.59			1.56	1.57
Ramachandran plot[e]					
%-most favored	91.7			91.0	91.6
%-additionally allowed	8.3			9.0	8.4

[a]Calculated by treating Bijvoet pairs as equivalent.
[b]$R_{sym} = \Sigma_h \Sigma_i (|I_i(h) - \langle I(h)\rangle|) / \Sigma_h \Sigma_i I_i(h)$, where $I_i(h)$ is the ith intensity measurement and $\langle I(h)\rangle$ is the weighted mean of all measurements of $I(h)$.
[c]$R_{iso} = \Sigma_h ||F(h)_{deriv}| - |F(h)_{native}|| / |F(h)_{native}|$
[d]R_{work} and $R_{free} = \Sigma_h ||F(h)_{obs}| - |F(h)_{calc}|| / |F(h)_{obs}|$ for reflections in the working and test sets (5% of all data) respectively.
[e]Regions defined by PROCHECK [6]

experimental map (see Table 5.1). The density-modified map was used to obtain a crude estimate of the non-crystallographic two-fold symmetry (NCS) operator relating the two β-subunits comprising the asymmetric unit. This NCS operator was refined by RESOLVE and used to produce a high quality, 2-fold averaged map into which the molecular model of Hex B was built using the program O [7].

B.3. Model building and refinement of native Hex B

Prior to the refinement process, the program UNIQUE [8] was used to calculate a complete data set to 2.0 Å resolution using the space group $P6_122$ and unit cell dimensions determined for crystals of native Hex B (Table 5.1). A random subset of *hkl* indices (5%) was selected from this calculated data set using FREERFLAG [8] and set aside for cross-validation using the free *R* factor [9]. A hand-built model of native Hex B was refined using a high-temperature, simulated annealing protocol guided by a maximum-likelihood target function within the CNS program [10]. Due to the high temperature simulation, molecular dynamics were restricted to torsion angles only [11]. This initial round of model refinement included experimental phase information as output from SOLVE and was restrained to the 2-fold NCS operator refined by RESOLVE. Following simulated annealing, several rounds of conjugate-gradient minimization were carried out, resulting in a combined reduction of *R*-work (*R*-free) from 0.41 (0.40) to 0.29 (0.34). Subsequent iterative rounds of model building and TLS refinement [12] were carried out using O and REFMAC5 [8], respectively, until R-factor convergence was reached (Table 5.1). NCS restraints and experimental phases were not included during the later rounds of refinement using REFMAC5. Waters were picked using ARP/wARP [8].

B.4. Model building and refinement of the NAG-thiazoline·Hex B and GalNAc-isofagomine·Hex B complexes.

NAG-thiazoline and GalNAc-isofagomine were separately soaked into Hex B crystals to create complexes isomorphous with native Hex B crystals. Using a maximum likelihood target function within REFMAC5, rigid body refinement and

several rounds of conjugate gradient minimization, were sufficient to position a solvent free molecular model of native Hex B into the asymmetric unit of each crystallographic complex. Initial sigma-A weighted $|F_o|-|F_c|$, α_c and $2|F_o|-|F_c|$, α_c electron density maps (where α_c are the calculated phases and $|F_o|$ and $|F_c|$ are the measured and calculated structure factor amplitudes, respectively), computed prior to solvent modeling, unambiguously defined the conformation of GalNAc-isofagomine or NAG-thiazoline bound in the –1 subsites of each Hex B active site. Models of the inhibitors were built into the appropriate electron density of each complex using the program O and the coordinates were subjected to iterative rounds of TLS refinement and solvent modeling until R-factor convergence was reached (Table 5.1).

B.5. Coordinates

Coordinates and structure factors have been deposited into the Protein Data Bank: Native Hex B, 1NOU; Hex B:GalNAc-isofagomine, 1NOW; Hex B:NAG-thiazoline, 1NPO.

B.6. Comparative molecular modeling of Hex A and docking of the GM2-ganglioside/activator protein

Using the program Swiss-PDBviewer [13], a pairwise sequence alignment was generated between the mature α and β-subunit aa sequences. This alignment was used as a guide for substituting the aa sequence of the α-subunit onto the C^{α} coordinates of a β-subunit from the Hex B structure. Loops 280-283 and 396-398, not present in the β-subunit structure, were modeled using the program MODELLER [14]. Side chain positions in the α-subunit model were optimized manually. Although not energy minimized, the α-subunit model has acceptable stereochemistry as determined by PROCHECK (Ramachandran plot: 90.5% within limits, 8.3% allowed, 1.2% generous, 0.0% disallowed) [6]. Because the α-subunit model replaces one of the β-subunits in the Hex B crystal

structure, the remaining β-subunit of Hex B was used to create the complete Hex A αβ heterodimer with a dimer interface resembling closely that of Hex B.

The activator crystal structure contains three copies of the protein per asymmetric unit [15]. A mobile loop (Val153-Leu163) forms part of the opening to the hydrophobic cavity of the activator. For two of the three copies in the asymmetric unit (monomers A and B), the loop is in a 'open' conformation, whereas in the third copy (monomer C) the loop is in an 'closed' conformation [15]. All three models of the activator (monomers A, B & C) were individually docked onto the Hex A homology model using the program suite 3D-DOCK [16, 17]. Interestingly, when based on optimal electrostatics and surface complementarity, only monomer A docked onto the Hex A model such that the opening to the hydrophobic cavity aligned with the α-subunit active site. This suggests that the open conformation of the activator is the form that binds to Hex A.

Inspection of the complex revealed that the mobile loop of the docked activator occluded part of the active site so that G_{M2}-ganglioside could not be modeled into the complex without numerous steric clashes with the activator. Due to the apparent mobility of this loop, and lack of crystallographic information about how the activator interacts with G_{M2}-ganglioside, the geometry of the loop was not adjusted, and space for G_{M2}-ganglioside in the active site was instead made through small manual rigid body movements of the whole activator structure relative to Hex A. Using the program O, the oligosaccharide portion of a model of G_{M2}-ganglioside was then fit into the α-subunit active site based on the Michaelis complex of chitobiose bound to *Sm*CHB (Figure 1.4) [18]. The modeled conformation of the oligosaccharide and the manually adjusted position of the activator protein allowed for the lipid tail of G_{M2}-ganglioside to enter the hydrophobic cavity of the activator protein.

C. Results and Discussion

C.1. Structure of human Hex B

Mature Hex B (M_r 112,000) was purified from human placenta and crystallized in its native mature (lysosomal) form [1]. Its crystal structure was determined using the multiple isomorphous replacement (MIR) method with data to 2.8 Å resolution. A model of the Hex B homodimer was readily built into the experimental electron density and subsequently was refined with data to 2.4 Å resolution (Table 5.1).

Two β-subunits, related by non-crystallographic two-fold symmetry, comprise the asymmetric unit of the crystals used to determine the Hex B structure. The subunits share a buried surface area of 1612 $Å^2$, and when aligned structurally, their Cα atoms have an r.m.s. difference of 0.3 Å; however, the contacts made between these subunits are the result of crystal packing and do not represent the biological homodimer interface of the Hex B enzyme. Instead, two biologically relevant Hex B homodimers were found to lie side-by-side on a common crystallographic two-fold axis (space group $P6_122$) (Figure 5.1), such that each biological homodimer contributed one β-subunit to the asymmetric unit. The β-subunits of one of the biological homodimers share a buried surface area of 2694 $Å^2$, whereas the subunits of the other share a buried surface area of 2737 $Å^2$. By choosing either of these crystallographically related subunit pairs as the biologically relevant Hex B homodimer, the structural basis for much of the observed biochemistry of this enzyme became apparent.

C.2. Subunit structure

Each β-subunit of Hex B is a kidney-shaped, two-domain protein (Figure 5.2). The three polypeptides present in the lysosomal form of each mature β-subunit, β_p, β_b, β_a (Figure 5.4), could be traced as independent chains through the electron density. However, residues 50-54 of β_p and 553-556 of β_a, which constitute the extreme N and C termini of the mature amino acid (aa) sequence, respectively, could not be modeled due to insufficient density. The

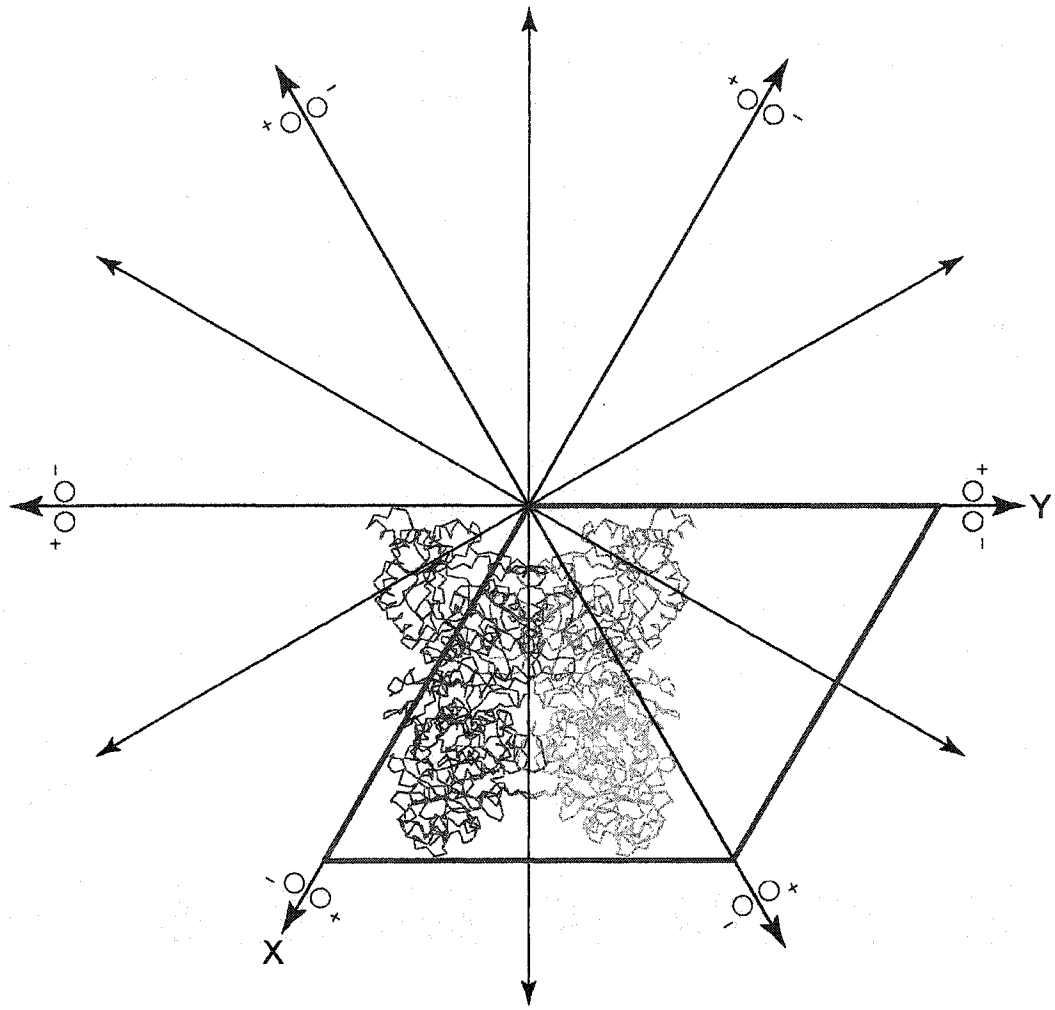


Figure 5.1 The molecular packing of one asymmetric unit (green and yellow β-subunits) with another asymmetric unit (blue and magenta β-subunits) within the human Hex B crystal form used for structure determination. The two asymmetric units are related by an implied crystallographic two-fold axis, and the unit cell axes are shown in red. The green and yellow subunits of the asymmetric unit do not comprise the biological dimer. Instead, the asymmetric unit is comprised of subunits from two independent biological homodimers which lie on a common crystallographic two-fold axis. The blue and green subunits comprise one biological homodimer, while the magenta and yellow subunits comprise the other biological homodimer. The stereographic projection 622 has been superimposed onto the unit cell to illustrate the symmetry elements.

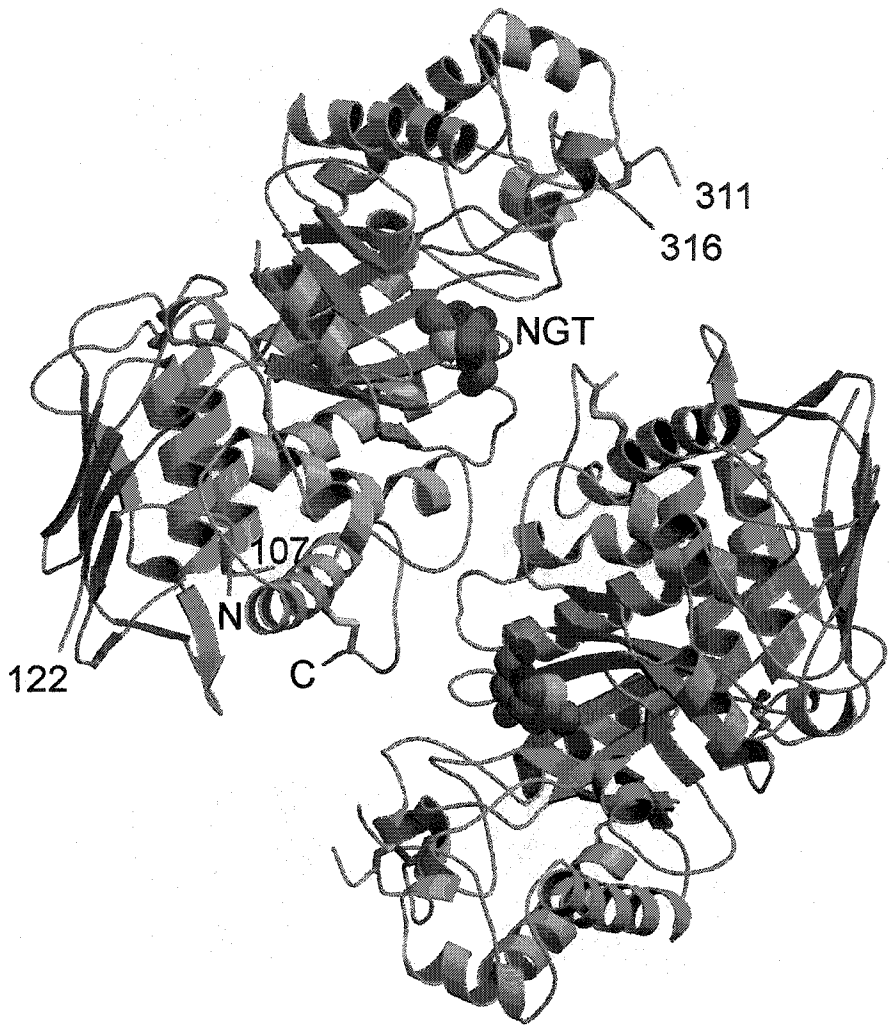


Figure 5.2. Ribbon diagram of human b-hexosaminidase B. The β-subunits of the Hex B homodimer are coloured with domain I in *green* and domain II in *blue* (the 8 parallel strands of the β-barrel of domain II are colored *sky blue*). What appear to be common structural features of family 20 glycosidases is the absence of regular α-helices at positions α5 and α7 of the (β/α)8-barrel structure of domain II and an additional C-terminal helix following helix α8. This additional helix packs between domains I and II, spatially orienting the two domains relative to each other. Helix α7 consists of only 2 turns and is part of an extended loop that forms a major portion of the dimer interface. The subunits are related at the dimer interface by a crystallographic two-fold symmetry axis running perpendicular to the page. The N and C-termini created as a result of posttranslational processing are numbered by residue. The labels N and C denote the extreme N (residue 55) and C (residue 552) termini visible within the electron density. The disulphide bonds Cys91-Cys137, Cys309-Cys360 and Cys534-Cys551 are drawn in *brown*, *magenta* and *yellow*, respectively. The analogue of the reaction intermediate NAG-thiazoline, bound in the active site of each subunit is drawn as a space filling model with carbon atoms in grey, oxygen in *red*, nitrogen in *blue* and sulphur in *yellow*. The active sites of each subunit are located 37 Å apart. Note the remarkable structural similarity of each subunit to *Sp*HEX (Figure 2.4).

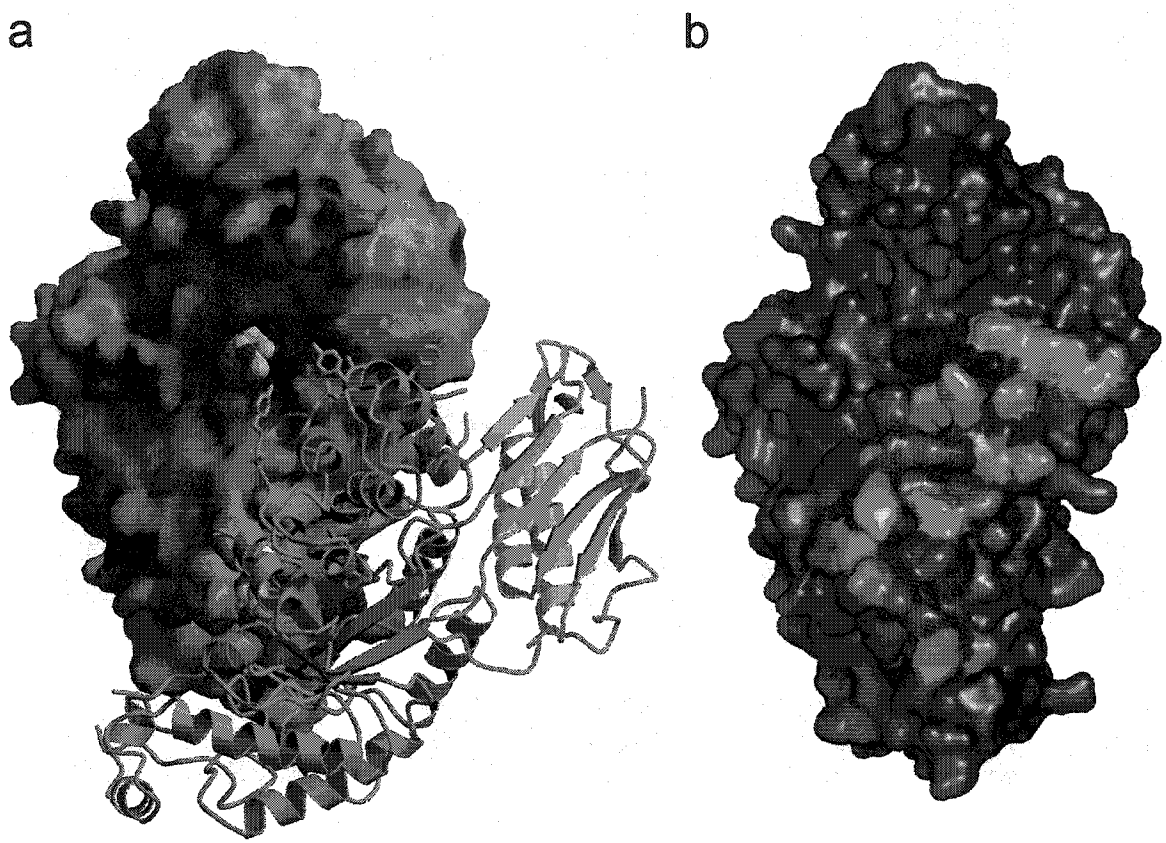


Figure 5.3. Electrostatic potential surface map and dimer interface of human Hex B. *Panel a*, A solvent accessible surface, drawn over one b-subunit and coloured with regions of positive charge in *blue* and negative charge in *red*, reveals an overall negative charge about the active site (figure created using the program GRASP [53]). The other subunit of the homodimer is represented by a ribbon diagram with domain I in *green* and the catalytic (β/α)8 domain II in *yellow.* The intermediate analogue NAG-thiazoline, bound in the active site of each subunit is shown as a space filling model with carbon atoms in *grey*, oxygen in *magenta*, *nitrogen* in *blue* and sulphur in *yellow. Panel b,* Surface rendering of a single β-subunit showing the extensive surface area buried at the dimer interface as determined using the CNS program [10]. Polar side chains are coloured *green*, hydrophobic side chains are *orange*, backbone atoms are *blue,* charged residues in *red* and residues not involved in dimerization are coloured *grey.* The active site pocket is coloured *magenta.* Panel drawn using PYMOL [54] and colour scheme adopted from [55].

posttranslational proteolytic cleavages that produce these three chains from their single precursor remove 2 surface accessible loops (as previously predicted [19, 20]) from each β-subunit, resulting in the three disulfide-linked polypeptides that comprise the mature subunit (Figure 5.4). The locations of the disulfide bonds are consistent with results obtained by MALDI-MS [21]. Although *N*-glycan electron density was observed on Asn residues 84, 142, 190 and 327, only two sugars (di-GlcNAc) were built into density extending from position 190; the remaining sugars extending from this position, and all the other *N*-linked glycans were too unstructured to model.

Domain I (residues 50 to 201) of each subunit consists of a 6 stranded anti-parallel β-sheet that buries two, parallel α-helices against domain II. Domain II (202-556) is a $(\beta/\alpha)_8$-barrel structure that houses the active site within loops extending from the C-termini of the strands that constitute the β-barrel. The β-subunits of Hex B have a fold remarkably similar to homologous domains found in the two other Family 20 glycosidases that have had their structures determined; the 506 aa β-hexosaminidase from *S. plicatus* (*Sp*HEX) [22] and the 818 a.a. chitobiase from *S. marcescens* (*Sm*CHB) [18]. *Sp*HEX is a monomeric, two-domain protein with a fold nearly identical to one mature β-subunit of Hex B; 373 Cα atoms of the two proteins have a r.m.s. difference of only 1.6 Å. *Sm*CHB consists of four domains, where domains II and III have 344 Cα atoms with a r.m.s. difference of only 1.4 Å compared to a mature β-subunit of Hex B. These structural comparisons, along with multiple sequence alignments, provide sufficient evidence to suggest that the two-domain structure observed for the mature β-subunit represents a fundamental fold present in all family 20 glycosidases found in species ranging from prokaryotes to humans.

C.3. Dimer interface

The α- and β-subunits of Hex are believed to be enzymatically inactive as monomers. The crystal structure of Hex B not only reveals why dimerization is crucial for catalytic activity, but also comparative molecular modeling studies of

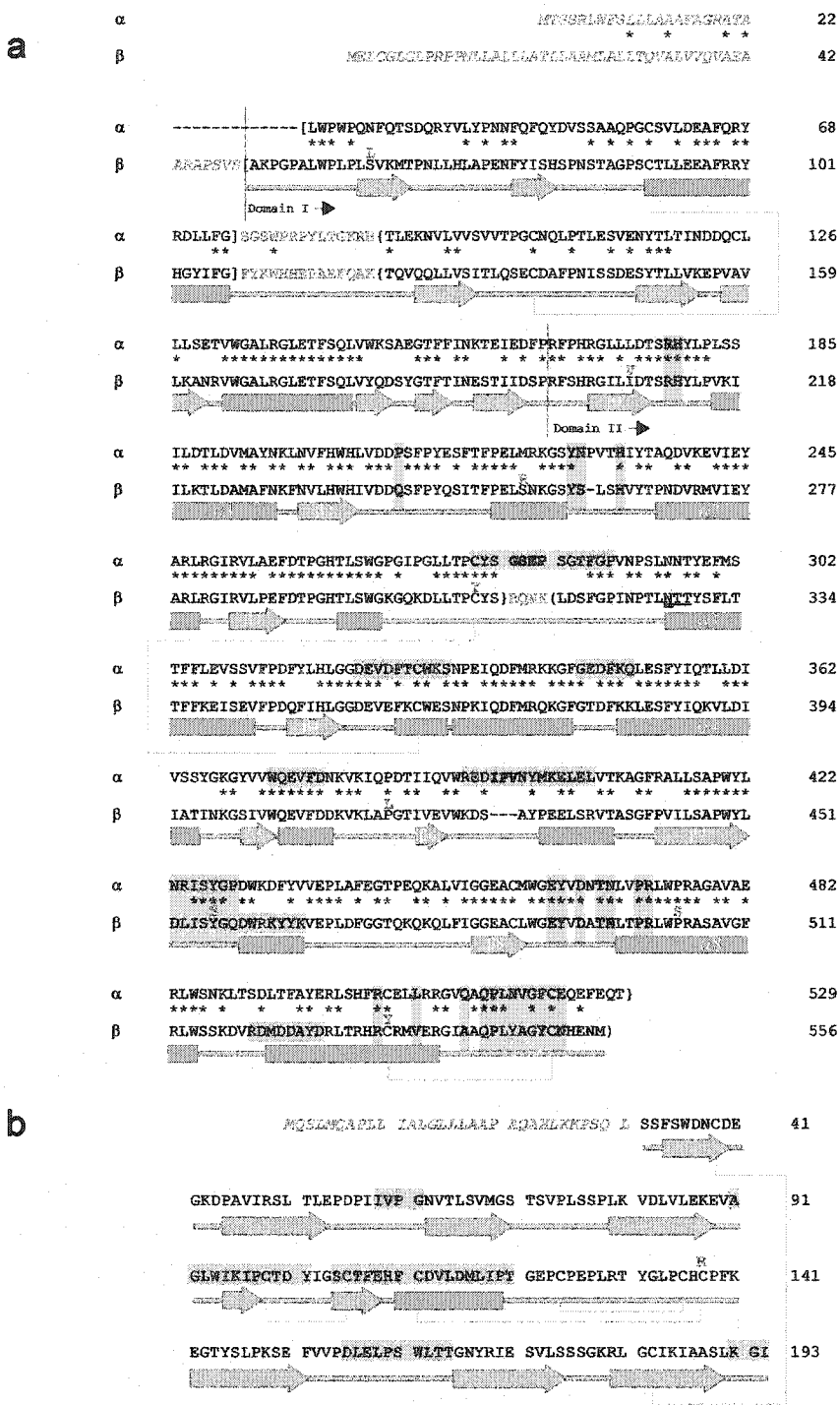


Figure 5.4 Pairwise sequence alignment and secondary structure of subunits α and β (a) and sequence of the Activator with secondary structure [15] (b). Residues coloured in *blue* in panels *a* & *b* are removed during posttranslational processing. Secondary structure is indicated as follows: Helices are *green* boxes, strands are *blue* arrows and disulfide bridges are shown as *grey* lines. Residues boxed in *yellow* are involved in dimerization as determined from the Hex B crystal structure and as predicted for the Hex A isoform. Residues boxed in *blue* are predicted to form the structural binding epitopes on Hex A and the Activator protein (as determined from visual inspection of the Hex A model - figure 5.6 a & b). The unique α-subunit loops 280-283 (GSEP) and 396-398 (IPV) are shown in bold type and are predicted to interact with the Activator. Known β-subunit point mutations are shown directly above the β-subunit sequence in purple.

Hex A suggest that the dimer interface forms the docking site for the G_{M2}·activator complex (see below). The β-subunits of Hex B dimerize with their active sites facing towards one another, but are offset by ~120° about an axis perpendicular to the crystallographic two-fold (Figures 5.2 and 5.3). This creates a large dimer (79 x 90 x 87 Å) with a continuous "U" shaped cleft between the two active sites of the enzyme, providing each active site unobstructed access to the solvent.

Because of the crystallographic two-fold symmetry in the Hex B dimer, each subunit experiences identical protein:protein interactions at the dimer interface. The extensive interface, formed exclusively between the catalytic $(\beta/\alpha)_8$-barrel domains, covers a patch on the monomer surface adjacent to the active site of each subunit, and several residues from one subunit structurally complete and stabilize active site residues of the other subunit (Figure 5.5). The family 20 glycosidase inhibitor GalNAc-isofagomine [23] was soaked into crystals of Hex B to form a complex and its position within the active site was used to distinguish important catalytic residues of Hex B. Figure 5.5 shows precisely how residues from the partnering subunit stabilize and arrange the active site residues that interact with the inhibitor. This arrangement is identical for both subunits and without these cooperative interactions at the dimer interface, numerous van der Waals contacts and hydrogen-bonding interactions required to stabilize more than half of each active site would be absent, likely rendering the lone subunits inactive.

In particular, Tyr456 and Tyr547, along with neighbouring residues, create a pocket complementary to the active site residues Glu491 and Tyr492 of the partnering subunit. Glu491 selectively stabilizes sugars of *galacto*-configuration within the active site pocket of Hex B as evidenced by the hydrogen-bonding interaction formed between the carboxyl group of this residue and O4 of GalNAc-isofagomine. Hence, interactions from the partnering subunit are indirectly involved in stabilizing the *galacto* pyranoside configuration most readily catalyzed by human Hex A and B. The functional importance of Tyr456 is evidenced by the naturally occurring mutation Tyr456Ser, a mutation that affects dimerization,

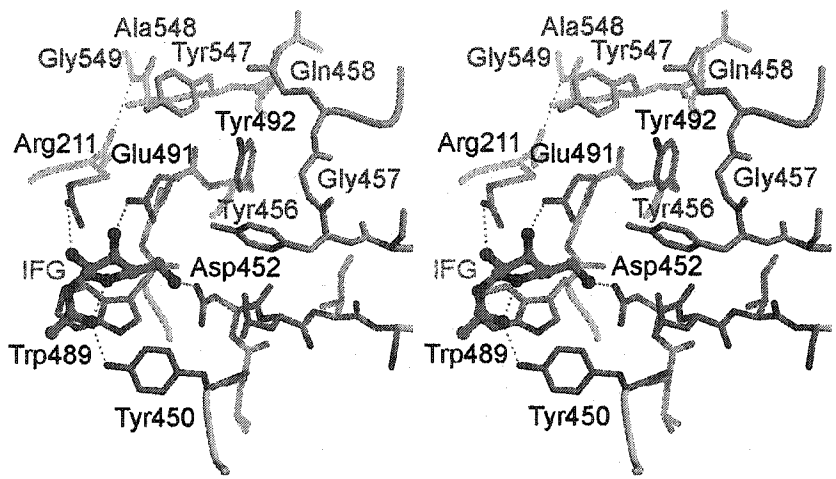


Figure 5.5. Stereographic representation of the dimer interface near the β-subunit active site. Active site residues (*grey*) stabilized by interactions from residues of the partnering subunit (*yellow*). The two-fold symmetry at the dimer interface results in both active sites experiencing the same stabilizing effects from the associated monomer. The crystallographically determined position of GalNAc-isofagomine (IFG) in the active site of each subunit demonstrates that 4 of the 6 hydrogen bonds between the enzyme and inhibitor depend on stabilizing interactions from the partnering subunit. In the absence of the protein:protein interactions that are formed upon dimerization, Arg211, Glu491, Asp452 and Tyr450 are most likely too unstructured to be catalytically active.

creating a transport deficiency from the ER. The patient with this mutation experienced a mild chronic form of Sandhoff disease which was likely made possible through his second unidentified mutant allele [24-26]. Moreover, just two resides from Tyr547 lies Gly549, the backbone nitrogen of which donates a stabilizing hydrogen bond to the carbonyl oxygen atom of Arg211 of the partnering subunit. This Arg residue is conserved in all known family 20 glycosidases and has been shown to be crucial for substrate binding and the stabilization of reaction intermediates (Figures 5.5 & 5.6) [18, 27, 28]. The importance of this Arg residue to the catalytic mechanism of human Hex A is underscored by the naturally occurring variant B1 mutations (Chapter 1, section **D.1**) of this residue that results in a transport-competent, mature Hex A with a catalytically inactive α-subunit, but an active β-subunit. Patients with mutations at αArg178 (aligns with βArg211) have near normal levels of Hex A protein and activity when assayed with artificial β-GlcNAc-containing substrates, but lack the ability to hydrolyze artificial substrates containing a β-GlcNAc-6-SO_4 residue or G_{M2} [29-34].

Both Gly549 and Tyr547 reside on a loop that terminates at the C-terminus of polypeptide β_a. This loop forms numerous interactions at the dimer interface and is structurally stabilized by a previously reported disulphide bond between Cys534-Cys551 [20]. All residues downstream of Cys551 to the C-terminus of the subunit (residue 556) are disordered. The stabilizing effect of this disulfide bond must allow for efficient and stable dimerization to take place. The loss of this disulphide bond by the natural missense mutation βCys534Tyr results in the acute form of Sandhoff disease [35]. Residues buried at the dimer interface of Hex B are highlighted in the pairwise alignment between the α- and β-subunits shown in Figure 5.4. Assuming Hex A and B dimerize in an analogous manner, the alignment demonstrates that not all residues at the dimer interface of Hex A and B are identical which may explain the differences in dimer stability observed between the isozymes (ββ>αβ>αα) (reviewed in [36]).

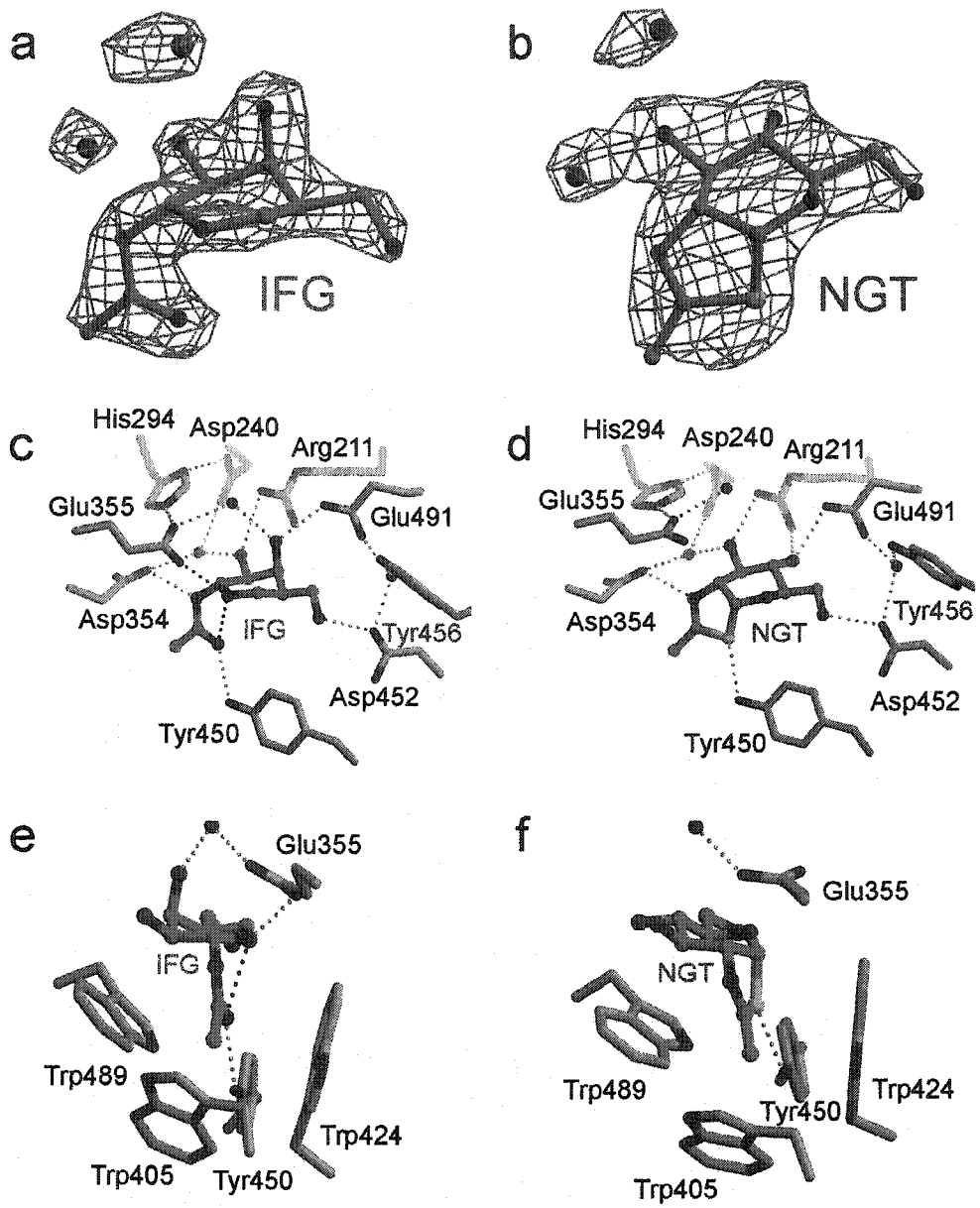


Figure 5.6. Hex B in complex with the transition state mimic GalNAc-isofagomine (IFG) (*a,c,e*) or the intermediate analogue NAG-thiazoline (NGT) (*b,d,f*). *Panels a and b*, Unrefined 2.2 Å and 2.5 Å resolution sigma-A weighted |Fo|-|Fc|, α_C electron density map containing the refined model of IFG and NAG-thiazoline (NGT) bound in the active site of a Hex B β-subunit, respectively (electron density contoured at 2.5 σ). Panels c and d, The b-subunit active site showing the extensive hydrogen-bonding interactions between IFG and NGT and the enzyme, respectively. The *magenta* hydrogen-bond between Glu355 and the ring nitrogen of IFG is believed to compensate for the missing hydrogen bond that occurs between an isofagomine and a 'normal' β-retaining glycosidase [23,38]. Tyr456 (*yellow*) comes into the active site from the partnering subunit. *Panels e and f,* Trp residues create a hydrophobic pocket into which the C2-acetamido group becomes appropriately positioned for intramolecular nucleophilic attack at the anomeric carbon atom of the terminal sugar. The water molecule located above the β-face of the azasugar ring of IFG and NGT may represent the incoming water that undergoes base catalyzed (activated by Glu355) nucleophilic attack at the anomeric center of the cyclized intermediate.

C.4. Active site structure and catalytic mechanism

By soaking Hex B crystals with the transition state mimic GalNAc-isofagomine [23] and with the reaction intermediate analogue NAG-thiazoline [22, 37], enzyme-inhibitor complexes (Figure 5.6) were obtained, the structures of which clearly demonstrate that human Hex B uses a substrate-assisted catalytic mechanism in accordance with results previously described for other Family 20 glycosidases (Figure 1.2b, *lower pathway*) [18, 22].

C.4.1. Hex B· GalNAc-isofagomine complex

Although crystallographic studies confirm that the piperidinium ring of 1-*N*-azasugars of the isofagomine class do not mimic the planar conformation for atoms C1, C2, O5 and C5 of a pyranoside ring expected during the oxocarbenium ion-like transition state, they do mimic the electrostatic nature of the transition state via a protonated and positively charged nitrogen as their 'anomeric' center [23, 38]. The protonated endocyclic nitrogen of isofagomine azasugars typically form a strong electrostatic interaction with the enzyme nucleophile of β-retaining glycosidases, thereby making them potent inhibitors of these enzymes. For Hex however, the interaction is different; in the absence of an enzyme nucleophile, the protonated nitrogen atom of the ring forms instead a strong electrostatic, hydrogen-bonded interaction with the general acid-base residue (Figure 5.6c) [23]. This novel interaction imparts sufficient binding energy to make isofagomines potent competitive inhibitors of family 20 glycosidases, and the interaction can be observed in the human Hex B·GalNAc-isofagomine complex as a 2.8 Å hydrogen bond donated from the ring nitrogen of the inhibitor to the carboxyl group of Glu355, a residue previously suspected of being the general-acid-base residue of Hex B [39, 40]. The mutation βGlu355Gln reduces k_{cat} 5000-fold with only a small effect of K_m, an observation consistent with its role as the acid/base catalyst in the mechanism of Hex B [41].

As expected, the C2-acetamido group of the GalNAc-isofagomine is twisted beneath the α-face of the azasugar ring and locked into position by two hydrogen-bonding interactions from residues Asp354 and Tyr450 that flank either

side of a tryptophan lined pocket in which the C2-acetamido is seated (Figure 5.6e). Finally, there exists a 2.6 Å intramolecular hydrogen bond between the carbonyl oxygen atom of the C2-acetamido group and the protonated ring nitrogen of the inhibitor; this enzyme-induced interaction represents the trajectory of nucleophilic attack that would lead to the cyclized oxazolinium ion intermediate on the hydrolytic pathway of a good substrate (Figure 1.2b, *lower pathway*).

C.4.2. Hex B·NAG-thiazoline complex

NAG-thiazoline is a relatively stable analogue of the hydrolytically unstable oxazoline intermediate that is generated along the reaction coordinate of family 20 glycosidases [37]. Indeed, it has been demonstrated that jack bean Hex will synthesize NAG-thiazoline from a precursor molecule in which the C2-acetamido carbonyl oxygen atom of the terminal, non-reducing sugar was replaced by a sulphur atom [37]. The thiazoline ring of NAG-thiazoline is held within the tryptophan lined pocket of the Hex B active site by hydrogen-bonding interactions to Asp354 and Tyr450 in a manner similar to the C2-acetamido group of GalNAc-isofagomine (Figure 5.6d). Prior to cyclization, Asp354 and Tyr450 are thought to polarize the 2-acetamido amide, thereby increasing the charge density and nucleophilicty of the carbonyl oxygen atom and promoting nucleophilc attack [23]. Indeed, the mutation βAsp354Asn reduces the k_{cat} of Hex B 2500-fold while leaving the K_m essentially unchanged [41]. Upon cyclization, the carboxylate of Asp345 adopts an additional function by stabilizing the positive charge that develops on N2 of the oxazoline ring [22]. Structural and kinetic analysis of the equivalent Asp residue in *Sp*HEX (Asp313) [42, 43] and *Sm*CHB (Asp539) [42, 43] demonstrate that this residue not only appears to stablizie the transition states leading to and from the oxazoline intermediate, it also helps to orient the C2-acetamido during catalysis.

The pyranose ring conformation of NAG-thiazoline is best described as a $^4C_1$ chair; this is in contrast to the high-energy boat conformation seen for the terminal sugar residue in the Michaelis complex of *Sm*CHB bound to chitobiose [18]. This difference in conformation indicates that the anomeric carbon atom C1

undergoes the greatest nuclear motion during the catalytic reaction and is consistent with the electrophilc migration of C1 recently described for HEW-lysozyme [44].

In both inhibitor complexes of Hex B, a water molecule is held into position above the β-face of each inhibitor by the general acid-base residue, βGlu355, and may represent the incoming water molecule that is activated by βGlu355 to attack the anomeric center of the bound sugar, producing a product with β-configuration (Figure 5.6e & f). However, in order for this water to be in an ideal position to attack the anomeric C1 atom, it would need to be associated with, and activated by the other carboxyl oxygen atom of βGlu355. In the presence of a natural oxazolinium ion intermediate, it is believed that this water molecule would move into this ideal position so that βGlu355 could abstract a proton from it and the resulting hydroxide ion could attack the anomeric center and complete the reaction (Figure 1.2b, *lower pathway*). The hydrophobic pocket in which the oxazoline ring sits, appears to protect the intermediate from solvolysis via unwanted pathways; water can only attack from the β-face of the intermediate, effectively reinverting the anomeric configuration of the intermediate to produce a product with retained anomeric stereochemistry [22].

Finally, having inhibitors of both *gluco*- and *galacto*-configuration bound in the Hex B active site demonstrates how, through a joint effort by Arg211 and Glu491, Hex B can accommodate both sugar configurations within its active site. Because Arg211 appears to play the dominant role for binding sugars of *gluco*-configuration (Figure 5.6d), it is not surprising to find that the mutation βGlu491Gln does not significantly change the kinetic profile of Hex B when assayed for activity using substrates of *gluco*-configuration [41]; however, the GalNAc-isofagomine·Hex B complex (Figure 5.6c) indicates that such a mutation would have deleterious kinetic effects when catalyzing the hydrolysis of substrates of *galacto*-configuration.

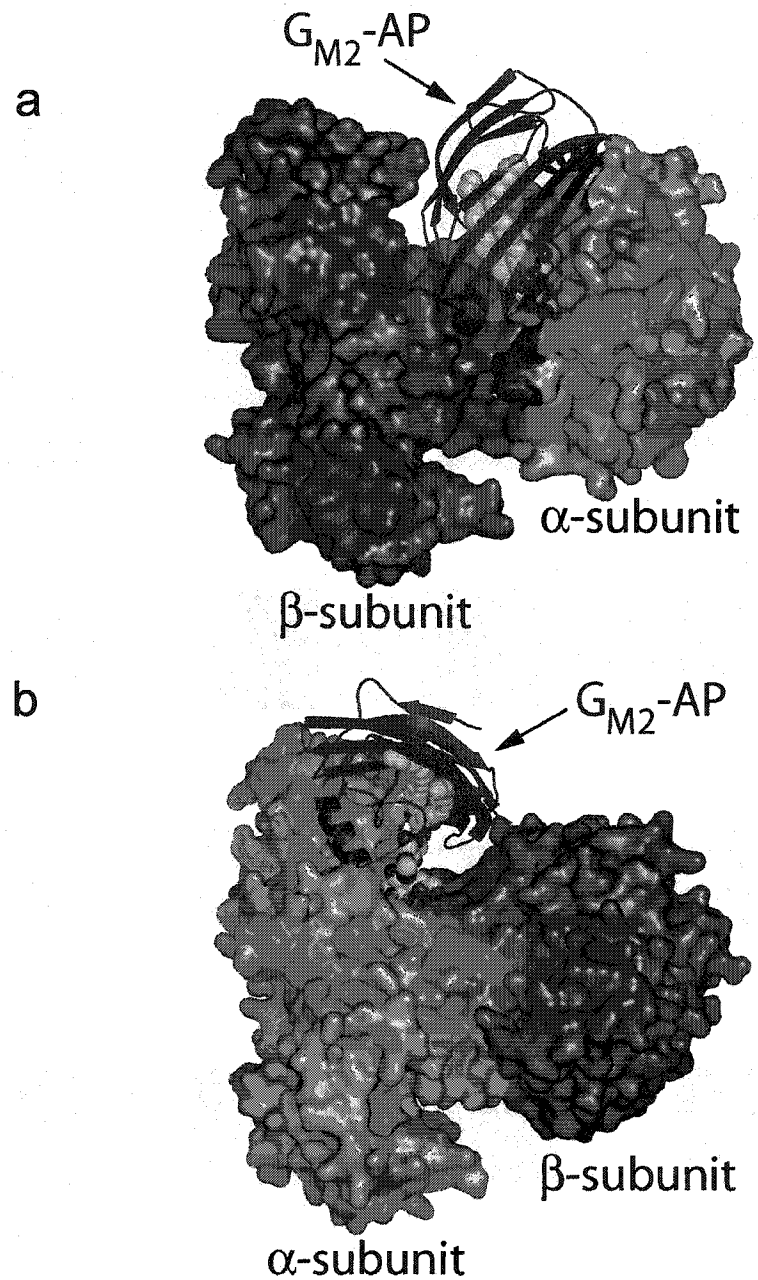


Figure 5.7. Predicted model of human Hex A·GM2·activator quaternary complex. *Panels a and b,* Two views of the predicted quaternary complex. Residues of the α-subunit identical to those of the β-subunit are coloured *blue*, non-identical residues are coloured light brown. Most of the conserved a.a. in the α- and β-subunits are located in (β/α)8-barrel of domain II. The β-subunit is coloured *grey*, with residues of the active site distinguished in *orange*. The GM2· activator protein complex (GM2-AP) docks into a large groove between the two subunits so that the terminal non-reducing GalNAc sugar on GM2 can be presented to the α-subunit active site and removed. Two surface loops (*magenta* and *green*), present only on the α-subunit, interact with the docked activator protein and appear to be involved in creating a docking site unique to the α-subunit. The *magenta* colored loop is proteolytically removed from the β-subunit during posttranslational processing and may represent a modification that regulates the metabolic function of this subunit.

C.5. Predictive modeling of Hex A and the HexA·G_{M2}-ganglioside·activator complex.

The primary structures of the α and β subunits are 60% identical, making the α-subunit an excellent candidate for comparative molecular modeling using the β-subunit structure as a template (Figure 5.7). Most of the conservation between the α and β-subunits is found within their catalytic domains (Figures 5.4 & 5.7). However, residues of the β-subunit involved in dimerization are also highly conserved in the mature α-subunit. With dimerization being a common functional requirement for the human isozymes, the αβ heterodimer almost certainly dimerizes the same way as observed for the Hex B (ββ) homodimer. Given a homologous dimer interface, the Hex A heterodimer model is essentially identical in overall shape to Hex B.

Interestingly, the predicted αβ dimer interface forms a large groove into which the activator structure [15] can dock. The novel β-cup topology of the activator consists primarily of β-sheet structure forming a hollow hydrophobic cavity that is accessible to the solvent through a hole at only one end of the protein [15]. The activator is believed to form a 1:1 complex with G_{M2} [45] and two hypotheses have been made for how the activator presents G_{M2} to Hex A [46]. In the first hypothesis, the activator binds G_{M2} in the intralysosomal membrane or vesicle and lifts it a few Ångstroms out of the membrane to allow Hex A access to its terminal β-GalNAc. In the second hypothesis, the activator fully removes G_{M2} from the membrane forming a soluble complex which then interacts with Hex A. Clearly, the closed β-cup formation of the activator indicates that the ceramide tail must be fully bound within the cup's hydrophobic interior (hypothesis 2), presumably leaving the oligosaccharide portion of the glycosphingolipid exposed through the opening so that the terminal GalNAc sugar can be presented to the α-subunit active site and removed. Indeed, our results from an automated docking study found that the optimal surface and electrostatic complimentarity between the Hex A model and the activator occurs when the entrance to the hydrophobic cavity of the activator is positioned directly

over the α-subunit active site (Figure 5.7). The predicted model of the active complex is consistent only with hypothesis 2 discussed above; thus, it appears that G_{M2} must first be fully removed from its membrane environment by the activator prior to docking onto Hex A.

The predicted G_{M2}:activator:Hex A docking interactions are supported by previous biochemical findings demonstrating that the activator interacts with the middle section of the α-subunit and the carboxyl half of the β-subunit of Hex A [47, 48]. Our predicted model of the quaternary complex indicates that the docked activator interacts with discrete patches on the α-subunit surface comprised of groups of residues found between α280-400 and with patches of residues between β465-545 of the β-subunit sequence. This latter observation is consistent with the finding that elements of the β-subunit are required for effective binding of the activator to Hex A [49].

Two natural mutations, Cys138Arg of the activator, and Pro504Ser of the β-subunit, are believed to specifically disrupt activator binding to Hex A [50, 51]. Cys138 forms a disulphide bond to Cys112 within the activator structure and stabilizes the position of an α-helix spanning residues 111-120 [15]. Our model of the quaternary complex suggests that residues of this α-helix form part of a surface epitope that binds directly to the α-subunit of Hex A. It appears that the activator mutation Cys138Arg could severely compromise this intermolecular interaction without affecting its ability to form a complex with GM2-ganglioside. This was exactly the biochemical phenotype found after analysis of the recombinant activator mutant (Cys138Arg) produced from bacteria [50].

Residue Pro504 of the β-subunit introduces a kink into helix α8 of the $(\beta/\alpha)_8$-barrel. This kink is required for proper packing of helix α8 against two loops that, according to our model of the quaternary complex, interact directly with the docked activator protein. Thus, disruption of the interactions between these two loops and helix α8 via the mutation Pro504Ser appears to adversely affect this portion of the binding epitope between Hex A and the activator. This prediction fits the observed biochemical phenotype for the Hex A mutant

(βPro504Ser), which displays a compromised ability to associate with the G_{M2}·activator complex, but can still hydrolyze simple water-soluble substrates [51].

In addition to subtle differences between the α and β-subunits that give rise to selective binding of the activator to the α-subunit of Hex A, two small loops, 280-283 and 396-398, observed only on the mature α-subunit, were found to interact directly with the docked activator (Figure 5.7). Loop 280-283 (Gly-Ser-Glu-Pro) aligns with the β-subunit residues 312-315 (Arg-Gln-Asn-Lys) that are removed during post-translational processing, suggesting that the selective removal of this loop from β-subunit helps to specify to which active site the activator presents substrate.

The active site structures of the α and β-subunits are essentially identical; however, three aa changes: βAsp453->αAsn423, βLeu454->αArg424 and βAsp427->αGlu394 appear to provide the α-subunit with the distinct characteristic of being able to accommodate a negative charge near the terminal non-reducing β-linked hexosamine of its substrates. A model of the Michaelis complex between the G_{M2} oligosaccharide and the α-subunit active site suggests that the negatively charged carboxylate of the sialic acid on G_{M2} is stabilized by the positively charged guanidino group of Arg424, which is in turn positioned by Glu394 (Figure 5.8). Removal of the positively charged guanidino group (αArg424Gln) results in an α-subunit with a 9-fold increase in K_m, and a slight decease in V_{MAX} relative to wild type, when assayed for its ability to catalyze the removal of GlcNAc-6-SO_4^- from an artificial substrate [52]. These kinetic data indicate that the positively charged guanidino group of Arg 424 can stabilize the negatively charged 6-sulphate group on the terminal GlcNAc-6-SO_4^- that is removed from the substrate by Hex A. Furthermore, the mutation αAsn423Asp results in α-subunit activity having a K_m elevated 6-fold relative to wild type when assayed for activity toward the same 6-sulphated substrate [52]. This mutation introduces a negatively charged carboxylate into the α-subunit active site at position 423 and this negative charge appears to reduce the binding affinity of

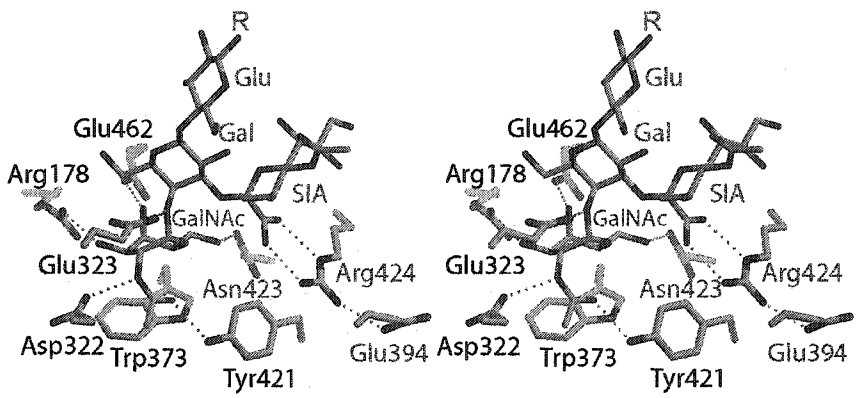


Figure 5.8. Model of the GM2 oligosaccharide (*yellow*) bound to the α-subunit active site (*gray*). The distorted boat conformation of the terminal GalNAc to be removed (Gal, labeled in *blue*) and the pseudo-axial orientation of the scissile bond and leaving group are based on crystallographic observations of the Michaelis complex of chitobiose bound to *Sm*CHB [18]. By incorporating these conformational restraints into the model, only one reasonable position could be found for the sialic acid residue (labeled SIA) within the active site pocket. Once positioned, the negatively charged carboxylate of the sialic acid, which can only be accommodated by the α-subunit, was found to come within hydrogen bonding distance of Arg424, a positively charged residue that is unique to the α-subunit (the β-subunit contains a Leu at this position). Glu394 and Asn423 (which are both Asp residues in the β-subunit) and are believed to help hold Arg424 into position. Arg 424, in turn, stabilizes the negatively charged caboxylate of the sialic acid of the substrate via electrostatic and hydrogen-bonding interactions. The general acid-base residue, Glu323 (Glu355 in the β-subunit), can be seen interacting with the glycosidic oxygen atom of the scissile bond.

the negatively charged terminal 6-sulphated sugar due to electrostatic repulsion of the substrate by the carboxylate of Asp423.

Together, these kinetic results indicate that disruption of the electrostatic environment provided by Arg424 and Asn423, reduces the ability of the α-subunit active site to accommodate negatively charged substrates, and provides biochemical evidence to substantiate the model presented here. The cleft that accommodates the sialic acid residue of G_{M2} is not present in other Family 20 glycosidases [18, 22], but is instead occupied by protein structure comprising sugar binding subsites that for G_{M2} at least, appear to be present on the activator protein. Furthermore, the conformation of the modeled G_{M2} substrate places the negatively charged carboxylate of its sialic acid in close proximity to the hydroxymethyl group of the terminal GalNAc residue to be removed. The close spatial arrangement of these two functional groups suggests that αArg424 and αAsn423 stabilize not only the negatively charged sialic acid moiety of G_{M2}, they also stabilize natural and artificial substrates that have the hydroxymethyl group of their non-reducing sugar replaced by a negatively charged 6-sulphate group. Biochemical confirmation of this hypothesis has been obtained using an αArg424Gln Hex A variant, which exhibited an increase in K_m for the G_{M2}·activator complex of 3-4-fold (Sharma *et al.*, unpublished). Substrates containing a 6-sulphate group are used to selectively measure α-subunit activity; however, it has been unclear how the α-subunit of Hex A could accommodate natural and artificial substrates containing negatively charged functional groups that appear to be in spatially distant positions from each other, *i.e* GlcNac-$\underline{6SO_4}$-X and GalNAc-Gal(SA)-Glc-Ceramide (GM2).

C.6. Conclusions

The structural data provide explanations for the mechanisms for many of the known biochemical properties of human Hex, including the structural basis for how many of the naturally occurring point mutations in *HEXA* and *HEXB* cause Tay-Sachs and Sandhoff disease, respectively. The X-ray crystal structure of Hex B demonstrates clearly why dimerization is necessary for enzymatic

function. Determining the structure of Hex B in complex with the mechanism-based inhibitors GalNAc-isofagomine and NAG-thiazoline not only confirm that human Hex uses a substrate-assisted catalytic mechanism in accordance with other family 20 glycosidases, but also the complexes reveal the structural basis for how the enzyme binds terminal amino sugars of both *gluco* and *galacto* configuration. Furthermore, the comparative molecular modeling of Hex A and subsequent activator docking studies provide insight into why only the α-subunit active site can stabilize negatively charged GlcNAc-6-sulphate containing substrates, and the negatively charged sialic acid moiety on G_{M2}, as well as insight into the structural basis for how Hex A associates with the G_{M2}·activator complex.

D. References

1. Church, W.B., et al., *Crystallization of human β-hexosaminidase B.* J. Mol. Biol., 1992. **227**: p. 577-580.
2. Mahuran, D.J. and J.A. Lowden, *The subunit and polypeptide structure of hexosaminidase from human placenta.* Can. J. Biochem., 1980. **58**: p. 287-294.
3. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Methods Enzymol, 1997. **276**: p. 307-326.
4. Terwilliger, T.C. and J. Berendzen, *Automated MAD and MIR structure solution.* Acta Crystallogr D, 1999. **55**(Pt 4): p. 849-61.
5. Terwilliger, T.C., *Maximum-likelihood density modification.* Acta Crystallogr D Biol Crystallogr, 2000. **56**(Pt 8): p. 965-72.
6. Laskowski, R.A., et al., *PROCHECK: a program to check the stereochemical quality of protein structure coordinates.* Journal of Applied Crystallography, 1993. **26**: p. 283-291.
7. Jones, T.A., et al., *Improved methods for the building of protein models in electron density maps and the location of errors in these models.* Acta Crystallogr A, 1991. **A47**(Pt 2): p. 110-119.

8. Collaborative Computational Project Number 4, *The CCP4 Suite: Programs for Protein Crystallography.* Acta Crystallogr D, 1994. **50**: p. 760-763.

9. Brunger, A.T., *Free R-Value - a Novel Statistical Quantity for Assessing the Accuracy of Crystal-Structures.* Nature, 1992. **355**(6359): p. 472-475.

10. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallogr D, 1998. **54**(Pt 5): p. 905-21.

11. Rice, L.M. and A.T. Brunger, *Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement.* Proteins, 1994. **19**(4): p. 277-90.

12. Winn, M.D., M.N. Isupov, and G.N. Murshudov, *Use of TLS parameters to model anisotropic displacements in macromolecular refinement.* Acta Crystallogr D Biol Crystallogr, 2001. **57**(Pt 1): p. 122-33.

13. Guex, N. and M.C. Peitsch, *SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling.* Electrophoresis, 1997. **18**(15): p. 2714-2723.

14. Fiser, A., R.K. Do, and A. Sali, *Modeling of loops in protein structures.* Protein Science, 2000. **9**: p. 1753-1773.

15. Wright, C.S., S.C. Li, and F. Rastinejad, *Crystal structure of human GM2-activator protein with a novel beta-cup topology.* J Mol Biol, 2000. **304**(3): p. 411-22.

16. Gabb, H.A., R.M. Jackson, and M.J.E. Sternberg, *Modelling Protein Docking using Shape Complimentarity, Electrostatics and Biochemical Information.* Journal of Molecular Biology, 1997. **272**: p. 106-120.

17. Katchalski-Katzir, E., et al., *Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques.* Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(6): p. 2195-2199.

18. Tews, I., et al., *Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease.* Nat Struct Biol, 1996. **3**(7): p. 638-48.

19. Hubbes, M., et al., *The amino-terminal sequences in the pro-α and -β polypeptides of human lysosomal β-hexosaminidase A and B are retained in the mature isozymes.* FEBS. LETT., 1989. **249**: p. 316-320.

20. Sagherian, C., et al., *Proteolytic processing of the proβ chain of β-hexosaminidase occurs at basic residues contained within an exposed disulfide loop structure.* Biochem. Cell Biol., 1993. **71**: p. 340-347.

21. Schuette, C.G., J. Weisgerber, and K. Sandhoff, *Complete analysis of the glycosylation and disulfide bond pattern of human beta-hexosaminidase B by MALDI-MS.* Glycobiology, 2001. **11**(7): p. 549-56.

22. Mark, B.L., et al., *Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase.* Journal of Biological Chemistry, 2001. **276**(13): p. 10330-10337.

23. Mark, B.L., et al., *Biochemical and structural assessment of the 1-N-azasugar GalNAc-isofagomine as a potent family 20 β-N-acetylhexosaminidase inhibitor.* J. Biol. Chem., 2001. **276**(45): p. 42131-42137.

24. Banerjee, P., et al., *Molecular basis of an adult form of beta-hexosaminidase B deficiency with motor neuron disease.* Biochem Biophys Res Commun, 1991. **181**(1): p. 108-15.

25. Banerjee, P., et al., *Preferential beta-hexosaminidase (Hex) A (alpha beta) formation in the absence of beta-Hex B (beta beta) due to heterozygous point mutations present in beta-Hex beta-chain alleles of a motor neuron disease patient.* J Biol Chem, 1994. **269**(7): p. 4819-26.

26. Redonnet-Vernhet, I., et al., *Significance of two point mutations present in each HEXB allele of patients with adult GM2 gangliosidosis (Sandhoff disease): Homozygosity of Ile^{207}->Val substitution is not associated with a clinical or biochemical phenotype.* Biochim. Biophys. Acta., 1996. **1317**: p. 127-133.

27. Mark, B.L., et al., *Structural and functional characterization of Streptomyces plicatus beta-N-acetylhexosaminidase by comparative molecular modeling and site-directed mutagenesis.* J Biol Chem, 1998. **273**(31): p. 19618-24.

28. Hou, Y., et al., *The role of beta-Arg211 in the active site of human beta-hexosaminidase B.* Biochemistry, 2000. **39**: p. 6219-6227.

29. Akerman, B.R., et al., *A mutation common in non-Jewish Tay-Sachs disease: frequency and RNA studies.* Human Mutation, 1992. **1**(4): p. 303-9.

30. Brown, C.A., et al., *Introduction of the α subunit mutation associated with the B1 variant of Tay-Sachs disease into the β subunit produces a β-hexosaminidase B without catalytic activity.* J. Biol. Chem., 1989. **264**: p. 21705-21710.

31. dos Santos, M.R., et al., *GM2-gangliosidosis B1 variant: analysis of beta-hexosaminidase alpha gene mutations in 11 patients from a defined region in Portugal.* Am J Hum Genet, 1991. **49**(4): p. 886-90.

32. Nakano, T., et al., *A point mutation in the coding sequence of β-hexosaminidase α gene results in defective processing of the enzyme in an unusual G_{M2}-gangliosidosis variant.* J. Neurochem., 1988. **51**: p. 984-987.

33. Nakano, T., et al., *A new point mutation within exon 5 of β-hexosaminidase α gene in a Japanese infant with Tay-Sachs disease.* Ann. Neurol., 1990. **27**: p. 465-473.

34. Tanaka, A., K. Ohno, and K. Suzuki, *G_{M2}-Gangliosidosis B1 variant: A wide geographic and ethnic distribution of the specific β-hexosaminidase α chain mutation originally identified in a Puerto Rican patient.* Biochem. Biophys. Res. Comm., 1988. **156**(2): p. 1015-1019.

35. Kuroki, Y., et al., *A novel missense mutation (C522Y) is present in the beta-hexosaminidase beta-subunit gene of a Japanese patient with infantile Sandhoff disease.* Biochem Biophys Res Commun, 1995. **212**(2): p. 564-71.

36. Mahuran, D.J., *Biochemical consequences of mutations causing the GM2 Gangliosidoses.* Biochim. Biophys. Acta., 1999. **1455**: p. 105-138.

37. Knapp, S., et al., *NAG-thiazoline, an N-Acetyl-beta-hexosaminidase inhibitor that implicates acetamido participation.* J Am Chem Soc, 1996. **118**: p. 6804-6805.

38. Notenboom, V., et al., *Detailed structural analysis of glycosidase/inhibitor interactions: Complexes of Cex from Cellulomonas fimi with xylobiose-derived aza-sugars.* Biochemistry, 2000. **39**(38): p. 11553-11563.

39. Fernandes, M.J.G., et al., *Identification of candidate active site residues in lysosomal beta-hexosaminidase A.* J. Biol. Chem., 1997. **272**(2): p. 814-820.

40. Pennybacker, M., et al., *Evidence for the involvement of Glu-355 in the catalytic action of human beta-hexosaminidase B.* J Biol Chem, 1997. **272**(12): p. 8002-6.

41. Hou, Y., et al., *Characterization of Glu and Asp residues in the active site of human β-hexosaminidase B.* Biochemistry, 2001. **40**: p. 2201-2209.

42. Williams, S.J., et al., *Aspartate 313 in the Sterptomyces plicatus hexosaminidase plays a critical role in substrate assisted catalysis by orienting the 2-acetamido group and stabilizing the transition state.* Journal of Biological Chemistry, 2002. **(in press)**.

43. Prag, G., et al., *Structures of chitobiase mutants complexed with the substrate Di-N-acetyl-d-glucosamine: the catalytic role of the conserved acidic pair, aspartate 539 and glutamate 540.* J Mol Biol, 2000. **300**(3): p. 611-7.

44. Vocadlo, D.J., et al., *Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate.* Nature, 2001. **412**(6849): p. 835-8.

45. Conzelmann, E., et al., *Complexing of glycolipids and their transfer between membranes by the activator protein for degradation of lysosomal ganglioside G_{M2}.* Eur. J. Biochem., 1982. **123**: p. 455-464.

46. Sandhoff, K., K. Harzer, and W. Fürst, *Sphingolipid activator proteins*, in *The Metabolic Basis of Inherited Disease*, C.R. Scriver, et al., Editors. 1995, McGraw-Hill: New York. p. 2427-2441.

47. Pennybacker, M., et al., *Identification of domains in human beta-hexosaminidase that determine substrate specificity.* J Biol Chem, 1996. **271**(29): p. 17377-82.

48. Tse, R., et al., *Identification of Functional Domains within the α and β Subunits of β-Hexosaminidase A Through the Expression of α-β Fusion Proteins.* Biochemistry, 1996. **35**: p. 10894-10903.

49. Sandhoff, K., E. Conzelmann, and H. Nehrkorn, *Specificity of human liver hexosaminidase A and B against glycosphingolipids G_{M2} and G_{A2}: purification of the enzymes by affinity chromatography employing specific elution.* Hoppe-Seyler's Z. Physiol. Chem., 1977. **358**: p. 779-787.

50. Xie, B., et al., *Biochemical Characterization of the $Cys^{138}Arg$ Missense Mutation Associated with the AB-Variant Form of G_{M2} Gangliosidosis: Evidence that Cys^{138} is Required for the Recognition of the G_{M2} Activator: G_{M2}-Ganglioside Complex by β-Hexosaminidase A.* Biochemistry, 1998. **37**: p. 814-821.

51. Hou, Y., et al., *A $Pro^{504}Ser$ substitution in the β-subunit of β-hexosaminidase A inhibits α-subunit hydrolysis of G_{M2} ganglioside, resulting in chronic Sandhoff disease.* J. Biol. Chem., 1998. **273**: p. 21386-21392.

52. Sharma, R., et al., *Identification of the 6-sulfate binding site unique to alpha-subunit- containing isozymes of human beta-hexosaminidase.* Biochemistry, 2001. **40**(18): p. 5440-6.

53. Nicholls, A., K.A. Sharp, and B. Honig, *Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons.* Proteins, 1991. **11**(4): p. 281-96.

54. DeLano, W.L., *The PyMOL Molecular Graphics System.* 2002, DeLano Scientific: San Carlos, CA, USA.

55. DeLano, W.L., *Unraveling hot spots in binding interfaces: progress and challenges.* Curr Opin Struct Biol, 2002. **12**(1): p. 14-20.

Chapter 6
Summary

A. The catalytic mechanism of family 20 glycosidases

To date the structures of three family 20 glycosidases, *Sp*HEX [1], *Sm*CHB [2] and human Hex B (manuscript submitted), have been determined. A comparison of these structures demonstrates a remarkable degree of structural conservation within this glycosidase family. All three structures contain a core protein fold comprised of two domains: The primary domain is a catalytic $(\beta/\alpha)_8$-barrel structure that houses an active site within loops that extend from the C-termini of the β-strands. Attached to the N-terminus of the $(\beta/\alpha)_8$-barrel domain, is a domain with α/β topology that includes a solvent exposed, anti-parallel β-sheet that buries two, roughly parallel, α-helices against the $(\beta/\alpha)_8$-barrel domain. Sequence alignments suggest that this latter domain is present in all family 20 glycosides; however, its function remains unknown.

Together, crystallographic studies of *Sp*HEX and *Sm*CHB have provided 'snapshots' of every stable enzymatic complex that occurs along the reaction coordinate of a family 20 glycosidase: First, there is the Michaelis complex of *Sm*CHB bound to its natural substrate chitobiose [2]. This crystallographic complex demonstrates perhaps the two most remarkable features of the catalytic mechanism of family 20 enzymes: 1) Family 20 enzymes distort the pyranoside ring of the terminal sugar (bound in subsite –1) from a relaxed $^4C_1$ chair conformation into a skew-boat conformation so that the scissile bond and leaving group become held in a pseudo-axial orientation, 2) In addition to pyranoside ring distortions, residues within the active sites of family 20 enzymes distort the bound substrate so that the carbonyl oxygen atom of the C2-acetamido group becomes appropriately positioned for nucleophilic attack at the anomeric center. Distortion of the terminal sugar ring from a $^4C_1$ chain into a $^4B^1$ or skew-boat conformation prior to bond hydrolysis is thought to maximize electron donation from the ring oxygen, O5, of the terminal sugar to the antibonding orbital at the

electron-deficient anomeric center. This electron donation stabilizes much of the positive charge developing at the anomeric center during the transition states [3, 4]. In addition to satisfying the requirements of stereoelectronic theory, distorting the terminal sugar into a boat conformation allows the C2-acetamido group to rotate freely underneath the α-face of the terminal sugar, while at the same time moving the hydrogen atom of the anomeric center out of the line of attack of the newly positioned nucleophile.

The arrangement of residues in the active site of family 20 enzymes is such that the scissile bond and the leaving group of the substrate must become held in a pseudo-axial orientation relative to the terminal sugar before the scissile glycosidic oxygen can become correctly aligned with the enzymic general acid/base residue. This suggests that general acid/base catalysis may have been incorporated into the catalytic mechanism of family 20 enzymes only after these enzymes had evolved considerable catalytic proficiency by inducing conformational changes in the substrate (including C2-acetamido group participation). Indeed, removal of the general acid/base residue Glu314 of *Sp*HEX by introducing the mutation Glu314Gln has been shown to reduce the V_{max} of this enzyme only 296-fold [5]. However, it is possible that general acid/base catalysis co-evolved over time with the rate-enhancing substrate distortion mechanisms described above. Interestingly, structural studies demonstrate that the active site residues of family 20 enzymes move very little during catalysis, even though they force the substrate to proceed through a radical conformational itinerary in order to take advantage of stereoelectronic weaknesses in the β-glycosidic linkage upon which they act.

The second 'snapshot' along the reaction coordinate of a family 20 glycosidases is of the crystallographic complex of *Sp*HEX bound to the cyclic intermediate analogue NAG-thiazoline (chapter 2) [1]. The most notable feature of this complex is the conformation of the pyranose ring of the bound intermediate analogue. The pyranose ring of the enzyme bound analogue is in a relaxed $^4C_1$ chair conformation, indicating that during the transition from the Michaelis complex to the cyclized enzyme intermediate, the anomeric center of

the substrate must migrate approximately 0.9 Å to a position where it can then covalently react with the carbonyl oxygen atom of the C2-acetamido group. One usually thinks of a nucleophile as moving in for the 'attack'; however, in this case, it is the electrophile that moves, and the nucleophile remains stationary. Superposing the *Sp*HEX:NAG-thiazoline complex onto the *Sm*CHB:chitobiose complex confirms that the active site residues do not move significantly during the first step of the double displacement reaction. The only nuclear motion that occurs during this first step is the migration of the anomeric carbon atom towards the waiting carbonyl oxygen atom of the C2-acetamido side chain. The second step of the reaction would presumably involve the migration of the anomeric center from its position in the $^4C_1$ chair conformation of the intermediate, to a position above the plane of the sugar ring, where a water molecule (activated by general acid/base) is waiting to react with it to form a hemiacetal product (Figure 1.2b, *lower pathway*). Movement of the anomeric center during glycosidic bond hydrolysis is now known as 'electrophilic migration', and it appears to be common to all configuration retaining β-glycosidases [6].

The only side chain movement that has been observed for a family 20 enzyme in response to a bound ligand is seen in the crystallographic complex of *Sp*HEX bound to GalNAc-isofagomine (Chapter 3) [7]. In this complex, $\chi 3$ of the general acid/ base residue Glu314 is rotated by approximately 20° so that the carboxyl group of this catalytic residue may accept a hydrogen bond from the ring nitrogen of the azasugar inhibitor. Because this inhibitor mimics the electrostatic character of the oxocarbenium ion-like transition state, perhaps slight side chain movements occur during transition state formation and breakdown. These movements are not seen in the crystallographic Michaelis complex, intermediate complex, or product complexes of these enzymes.

The third and final step along the reaction coordinate is the product complex, and insight into the structure of this complex has come from crystal structure determinations of *Sp*HEX bound to GlcNAc (product) [8]. Furthermore, the structures of two *Sp*HEX variants (Asp313Ala and Asp313Asn), determined in complex with GlcNAc (bound as product), have provided detailed insight into

the role of a conserved Asp residue (*Sp*HEX Asp313) in the catalytic mechanism of family 20 glycosidases (Chapter 4) [8]. As seen in the Michaelis complex of *Sm*CHB bound to chitobiose [2], the C2-acetamido group of GlcNAc bound to SpHEX is twisted underneath the α-face of the pyranose ring. The position of the substrate nucleophile appears to remain in a static position even after formation of the hemiacetal; however, the bulky C1 hydroxyl group does not allow the pyranose ring of the product to relax into $^4C_1$ chair conformation when bound in the enzyme active site, and this inability to bind in a low energy conformation may be what allows the incoming substrate to easily out-compete the product for binding to the active site. Although the substrate also binds to family 20 enzymes in a high energy conformation, the substrate makes significant binding interactions at the +1 subsite that are not available to the product, and this difference may dramatically increase substrate binding affinity towards the enzyme relative to product. Site-directed mutagenesis experiments in combination with structural analyses revealed that Asp313 in *Sp*HEX plays a vital role in substrate-assisted catalysis by properly positioning the C2-acetamido with the active site group during catalysis. This residue is conserved in all known family 20 glycosidases and mutagenesis studies in *Sm*CHB human Hex B confirm that this residue plays a similar catalytic role in these enzymes as well.

There is a substantial amount of evidence demonstrating that anchimeric assistance by C2-acetamido group participation is a common mechanistic feature of the glycosidases belonging to families 18, 20 and 56 [2, 9-11], and future structural and functional investigations may demonstrate this type of substrate-assisted catalysis to be an integral mechanistic component of glycosidases from other sequence-related families.

Substrate-assisted catalysis may have indeed occurred during the evolutionary history of many enzymes. It has been suggested for the serine proteinases at least, that the spontaneous formation of the classic triad of catalytic residues (Ser-His-Asp) would have been a highly unlikely evolutionary event. Alternatively, it was proposed that the serine proteinases once contained only a catalytic diad (Ser and Asp) and operated via an anchimeric assistance

mechanism whereby the His residue of the catalytic triad was provided by the substrate [12]. It was only later in evolutionary history that these proteinases incorporated within their active sites a suitably positioned His residue that could take over from the substrate, thereby dramatically broadening substrate specificity [12]. We may be observing a similar evolutionary path for the glycosidases from families 18, 20, and 56; however, it is also possible that existing evolutionary pressures demand that these enzymes maintain an active site architecture and catalytic mechanism that is dependent and specific for the presence of the C2-acetamido group. Regardless of the evolutionary fate of the family 18, 20, and 56 glycosidases, these enzymes presently have an active site structure that carries out a catalytic mechanism displaying all the general characteristics of the double-displacement mechanism first proposed by Koshland [11]. They provide unique support for the general mechanism of retaining β-glycosidases that includes substrate distortion, formation of a covalent intermediate, and electrophilic migration of C1 along the reaction coordinate.

B. Family 20 glycosidases and human disease

More than 120 years have passed since Warren Tay first described the clinical manifestations of G_{M2}-gangliosidosis [13]; however, it was only 33 years ago that the link between hexosaminidase activity and G_{M2}-gangliosidosis was established [14]. Since then, the genes encoding human Hex (*HEXA* and *HEXB)*, and the activator protein (*GM2A)* have been cloned and sequenced, and numerous disease causing mutations in these genes have been documented and their biochemical phenotypes studied [15]. Most importantly, the frequency of Tay-Sachs and Sandhoff disease has been reduced by more than 90% worldwide since the introduction of enzyme based screening programs in the late 1970's [15, 16].

For the past year now, we have been able to understand human Hex structure and function at the molecular level by studying the X-ray crystal structure of human placental Hex B. The crystal structure provides exciting new

insight into the structural basis for how eukaryotic family 20 β-hexosaminidases function, and how many of the documented *HEXB* mutations result in loss of enzyme function. The enzyme for this study was purified directly from human placenta [17] and its crystal structure exhibits all of the complex post-translational modifications of the mature enzyme. Each β-subunit (and presumably each α-subunit) consists of two domains (the core protein fold for family 20 enzymes), and one of the two domains is an $(\beta/\alpha)_8$-barrel structure that houses the enzyme active site. Hence, each of the human hexosminidase isoforms (αβ, ββ, αα) contains two functional active sites.

The Hex B structure was solved in complex with mechanism-based inhibitors [7, 18] and confirms earlier suggestions that the human hexosaminidases use the same substrate-assisted catalytic mechanism employed by bacterial family 20 glycosidases [1, 2]. The structure also provides novel information about how β-subunits dimerize and about why this initializes catalytic activity. The extensive dimer interface, formed exclusively between the catalytic $(\beta/\alpha)_8$-barrel domains, covers a patch on the monomer surface directly adjacent to the active site of each subunit. The residues from one subunit structurally complete and stabilize the active site of the other subunit. Many of the point mutations that cause Sandhoff disease disrupt the intricate interplay that occurs at this dimer interface.

Using the Hex B structure as a template, and the previously published crystal structure of the G_{M2}-activator protein, a model of Hex A in complex with the activator and ganglioside was built. Together, the crystallographic and modeling data demonstrate how α and β subunits dimerize to form Hex A or Hex B, how these isoenzymes hydrolyze diverse substrates, and how many documented point mutations cause Sandhoff disease (β-subunit mutations) and Tay-Sachs disease (α-subunit mutations).

The crystallographic studies on human Hex B lay the groundwork for future studies that investigate the molecular biology of G_{M2}-ganglioside degradation. Our understanding of the biochemistry of G_{M2}-gangliosidosis could be greatly enhanced by determining the X-ray crystal structures of Hex A alone

and in complex with the G_{M2}-activator protein, and perhaps even a complex including G_{M2}-ganglioside substrate. Such experimentally derived information regarding the protein-protein interactions involved in the G_{M2}-ganglioside degradation machinery would provide detailed insight into the molecular basis of how many of the known mutations cause disease. Furthermore, a solid understanding of the protein structure-function relationships of the system would allow the possibility of making phenotype predictions from newly identified genetic mutations. There is still much to be learned.

C. References

1. Mark, B.L., et al., *Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase.* Journal of Biological Chemistry, 2001. **276**(13): p. 10330-10337.
2. Tews, I., et al., *Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease.* Nat Struct Biol, 1996. **3**(7): p. 638-48.
3. Deslongchamps, P., *Stereoelectronic Effects in Organic Chemistry.* 1983, New York: Pergamon Press.
4. Kirby, A.J., *Stereoelectronic Effects on Acetyl Hydrolysis.* Accounts of Chemical Research, 1984. **17**: p. 305-311.
5. Mark, B.L., et al., *Structural and functional characterization of Streptomyces plicatus beta-N-acetylhexosaminidase by comparative molecular modeling and site-directed mutagenesis.* J Biol Chem, 1998. **273**(31): p. 19618-24.
6. Vocadlo, D.J., et al., *Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate.* Nature, 2001. **412**(6849): p. 835-8.
7. Mark, B.L., et al., *Biochemical and structural assessment of the 1-N-azasugar GalNAc-isofagomine as a potent family 20 β-N-acetylhexosaminidase inhibitor.* J. Biol. Chem., 2001. **276**(45): p. 42131-42137.

8. Williams, S.J., et al., *Aspartate 313 in the Sterptomyces plicatus hexosaminidase plays a critical role in substrate assisted catalysis by orienting the 2-acetamido group and stabilizing the transition state.* Journal of Biological Chemistry, 2002. **(in press)**.
9. Tews, I., et al., *Substrate-Assisted Catalysis Unifies Two Families of Chitinolytic Enzymes.* J Am Chem Soc, 1997. **119**: p. 7954-7959.
10. Markovic-Housley, Z., et al., *Crystal structure of hyaluronidase, a major allergen of bee venom.* Structure Fold Des, 2000. **8**(10): p. 1025-35.
11. Koshland, D.E., *Stereochemistry and the mechanism of enzymatic reactions.* Biol. Rev., 1953. **28**: p. 416-436.
12. Carter, P. and J.A. Wells, *Engineering enzyme specificity by "substrate-assisted catalysis".* Science, 1987. **237**(4813): p. 394-9.
13. Tay, W., *Symmetrical changes in the region of the yellow spot in each eye of an infant.* Trans. Ophthalmol. Soc. U.K., 1881. **1**: p. 155-157.
14. Okada, S. and J.S. O'Brien, *Tay-Sachs disease: generalized absence of a beta-D-N- acetylhexosaminidase component.* Science, 1969. **165**(894): p. 698-700.
15. Gravel, R.A., et al., *The G_{M2} gangliosidoses*, in *The Metabolic and Molecular Bases of Inherited Disease*, C.R. Scriver, et al., Editors. 1995, McGraw-Hill: New York. p. 2839-2879.
16. O'Brien, J.S., et al., *Tay-sachs disease. Detection of heterozygotes and homozygotes by serum hexosaminidase assay.* N Engl J Med, 1970. **283**(1): p. 15-20.
17. Mahuran, D.J. and J.A. Lowden, *The subunit and polypeptide structure of hexosaminidase from human placenta.* Can. J. Biochem., 1980. **58**: p. 287-294.
18. Knapp, S., et al., *NAG-thiazoline, an N-Acetyl-beta-hexosaminidase inhibitor that implicates acetamido participation.* J Am Chem Soc, 1996. **118**: p. 6804-6805.