

Structural and Functional Characterization of *Streptomyces plicatus* β -N-
acetylhexosaminidase by Comparative Molecular Modeling and Site Directed
Mutagenesis

By

Brian L. Mark

A Thesis Submitted to the Faculty of
Graduate Studies in Partial Fulfilment of the
Requirements for the Degree of

MASTER OF SCIENCE

Department of Biochemistry & Molecular
Biology, University of Manitoba,
Winnipeg, MB, R3E 0W3, Canada

© August, 1998



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-32181-9

Canada

THE UNIVERSITY OF MANITOBA
FACULTY OF GRADUATE STUDIES

COPYRIGHT PERMISSION PAGE

**STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF
STREPTOMYCES PLICATUS B-N-ACETYLHEXOSAMIDASE
BY COMPARATIVE MOLECULAR MODELING AND SITE DIRECTED MUTAGENESIS**

BY BRIAN L. MARK

**A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University
of Manitoba in partial fulfillment of the requirements of the degree**

of

MASTER OF SCIENCE

Brian L. Mark ©1998

**Permission has been granted to the Library of The University of Manitoba to lend or sell
copies of this thesis/practicum, to the National Library of Canada to microfilm this thesis
and to lend or sell copies of the film, and to Dissertations Abstracts International to publish
an abstract of this thesis/practicum.**

**The author reserves other publication rights, and neither this thesis/practicum nor
extensive extracts from it may be printed or otherwise reproduced without the author's
written permission.**

Dedication

To Adrienne and Misty

Acknowledgments

I would especially like to thank Dr. Barbara Triggs-Raine for her outstanding supervision and guidance throughout the duration of this work.

To Adrienne, my wife and friend, like always, your love, support, encouragement mean everything to me.

To my parents, without your love and support I would not have been able to obtain the education I have.

Finally, I would like to acknowledge the following people and organizations for their contributions to the work presented herein:

Dr. Phillips Robbins for providing the *Streptomyces plicatus* β -N-acetylhexosaminidase DNA clone (Clone 36).

Dr. Tim Cao, for completing the initial restriction enzyme map of the DNA clone containing the *Streptomyces plicatus* β -N-acetylhexosaminidase (*SpHex*).

Tim Salo and Greg Wasney, for their technical assistance

Canadian Genetic Diseases Network sequencing core facility for generating the initial *SpHex* nucleotide sequence

Sharon Wong-Madden of New England Biolabs for providing the initial pMAL-c2-*SpHex* fusion construct as well as guidance for optimal protein production.

Dr. Michael James and Dr. Peter Loewen, for providing access and guidance to their computer facilities

Amir Khan, for helping with the comparative molecular modeling, energy minimization and PROCHECK analysis of *SpHex* model.

TABLE OF CONTENTS

Dedication	i
Acknowledgements	ii
TABLE OF CONTENTS	iii
LIST OF FIGURES	viii
LIST OF TABLES	x
LIST OF ABBREVIATIONS	xi
ABSTRACT	xiii
<u>1</u> <u>INTRODUCTION</u>	1
1.0 β-N-acetylhexosaminidase, an O-glycosyl hydrolase	1
1.1 Human β-N-acetylhexosaminidases and the G_{M2} gangliosidoses	6
1.1.0 Biochemistry	7
1.1.0.0 <i>HEXA</i> and <i>HEXB</i> gene structures and isoenzyme products	7
1.1.0.1 Biosynthesis/processing	8
1.1.0.2 Substrate specificity and the G _{M2} activator protein	11
1.1.1 Classification of G _{M2} gangliosidoses	12
1.1.2 Identification of catalytic residues in human β -N-	
acetylhexosaminidase	13
1.2 The <i>Serratia marcescens</i> chitobiase x-ray crystallographic model	16
1.2.0 The <i>SmChb</i> structure and human β -N-acetylhexosaminidase	17
1.3 Structure/function studies of family 20 enzymes	18
1.4 Current study	20

2	<u>MATERIALS AND METHODS</u>	22
2.0	Plasmids	22
2.1	Bacterial culture	22
2.1.0	Bacterial strains	22
2.1.1	Culture media	23
2.1.2	Preparation of electrocompetent <i>E. coli</i>	24
2.1.3	Electroporation of electrocompetent <i>E. coli</i>	24
2.1.4	Growth and storage of transformed <i>E. coli</i>	25
2.1.5	Growth and storage of <i>S. plicatus</i>	25
2.2	Plasmid and genomic DNA isolation	25
2.2.0	Alkaline lysis miniprep of plasmid DNA	25
2.2.1	Large scale Nucleobond preparation of plasmid DNA	26
2.2.2	DNA purification by phenol/chloroform extraction	26
2.2.3	Isolation of <i>S. plicatus</i> genomic DNA	27
2.3	Determination of DNA and protein concentration	27
2.3.0	Spectrophotometric determination of DNA concentration	27
2.3.1	Visual estimation of DNA concentration	27
2.3.2	Colourometric determination of protein concentration	28
2.4	DNA cloning techniques	28
2.4.0	Restriction enzyme digests	28
2.4.1	Agarose gel electrophoresis of DNA	29
2.4.2	Polyacrylamide gel electrophoresis of DNA	29

2.4.3	DNA purification using the GeneClean II kit	30
2.4.4	Alkaline phosphatase treatment of DNA	30
2.4.5	Filling in 3' recessed ends of DNA	31
2.4.6	Ligations	31
2.5	Analysis of <i>S. plicatus</i> genomic DNA by Southern hybridization	31
2.5.0	Preparation of Southern blots	31
2.5.1	Random primer labelling of DNA probes	32
2.5.2	DNA/DNA hybridization	32
2.6	Preparation of single stranded DNA	33
2.7	DNA sequencing	34
2.7.0	Alkali denaturation of double-stranded plasmid DNA	34
2.7.1	Annealing and sequencing reactions	34
2.7.2	Sequencing gel electrophoresis	35
2.8	Comparative molecular modeling	36
2.8.0	Multiple amino acid sequence alignment	36
2.8.1	Model building	36
2.8.2	Energy minimization	37
2.8.3	Quality assessment	37
2.9	Site-directed mutagenesis	37
2.9.0	Annealing of the oligonucleotide to ssDNA	37
2.9.1	Synthesis of complementary DNA strand	37
2.9.2	Polymerase chain reaction (PCR)	38

2.9.3	Restriction enzyme analysis of PCR products	38
2.10	Expression and purification of <i>SpHex</i> -Maltose Binding Protein fusion	39
2.11	Factor Xa digestion of purified <i>SpHex</i> -MBP fusion	39
2.12	Preparation of <i>S. plicatus</i> cytosolic protein fraction	40
2.13	SDS-PAGE analysis	40
2.14	Western blot analysis	40
2.14.0	Generation of Antibodies Against <i>SpHex</i> -MBP Fusion Protein	40
2.14.1	Transfer of proteins resolved by SDS-PAGE onto nitrocellulose membrane	41
2.14.2	Immunodetection of proteins bound to nitrocellulose	41
2.15	Analysis of β - <i>N</i> -acetylhexosaminidase activity	41
2.15.0	In-plate assay of β - <i>N</i> -acetylhexosaminidase activity using 4-methylumbelliferyl- β - <i>N</i> -acetylglucosaminide (4-MUG)	41
2.15.1	β - <i>N</i> -acetylhexosaminidase activity assay and pH profile analysis	42
3	<u>RESULTS</u>	44
3.0	Restriction enzyme mapping of Clone 36	44
3.1	Sequencing the <i>SpHex</i> coding region from Clone 36	44
3.2	Cloning and sequencing the 5'-end of the <i>SpHex</i> gene	44
3.3	Comparison of <i>SpHex</i> to EMBL/SWISS-PROT database	50
3.4	Comparative molecular modeling of <i>SpHex</i>	50
3.5	Construction of the pMAL-c2- <i>SpHex</i> fusion protein vector	79

3.6	Site-directed mutagenesis of the <i>SpHex</i> gene	80
3.7	Purification of <i>SpHex</i> -MBP fusion and cleavage with factor Xa protease	80
3.8	Western blot analysis of recombinant and native <i>SpHex</i>	81
3.9	Kinetic analysis of wild-type and mutant <i>SpHex</i>	82
3.10	<i>SpHex</i> pH profile analysis	90
4	<u>DISCUSSION</u>	95
4.0	Clone 36 and the <i>SpHex</i> -MBP fusion protein expression/purification system ..	95
4.1	Comparative molecular modeling of <i>SpHex</i>	96
4.2	Functional analysis of <i>SpHex</i> by site-directed mutagenesis	101
5	<u>REFERENCES</u>	105

LIST OF FIGURES

Figure 1	Mechanisms of glycosidic bond hydrolysis as proposed by Koshland	2
Figure 2	Structures of G _{M2} ganglioside, chitobiose and the substrate analogs 4-methylumbelliferyl- β - <i>N</i> -acetylglucosaminide (4-MUG) and 4- methylumbelliferyl-6-sulfo- β - <i>N</i> -acetylglucosaminide (4-MUGS)	4
Figure 3	Proteolytic processing schematic for human Hex A	9
Figure 4	Restriction enzyme map and subclones of Clone 36	45
Figure 5	Southern blot autoradiograph of <i>S. plicatus</i> genomic DNA probed with a ³² P-labeled probe complimentary to the 5'-end of clone psHEX-1.8	48
Figure 6	Nucleotide and predicted amino acid sequence of the gene encoding <i>SpHex</i>	51
Figure 7	ClustalW1.7 multiple sequence alignment of family 20 glycosyl hydrolases	57
Figure 8	Amino acid sequence alignment between <i>SpHex</i> and <i>SmChb</i>	65
Figure 9	Ribbon diagram and active site architecture of the <i>SpHex</i> model	68
Figure 10	Ramachandran plot of the <i>SpHex</i> model	71
Figure 11	Comparison of the <i>SpHex</i> model and <i>SmChb</i> X-ray structure	74
Figure 12	Comparison of human Hex A α -chain and the <i>SpHex</i> model	76
Figure 13	SDS-PAGE of purified wild type and mutant <i>SpHex</i> -MBP fusion protein	82

Figure 14	<i>SpHex</i> hydrolysis of 4-MUG and 4-MUGS	84
Figure 15	Comparison of factor Xa digested and undigested <i>SpHex</i> -MBP fusion protein	86
Figure 16	Comparison of recombinant and native <i>SpHex</i>	88
Figure 17	Analysis of pH profiles for wild type and mutant <i>SpHex</i>	92

LIST OF TABLES

Table I	ClustalW1.7 multiple sequence alignment of SpHex with 15 family 20 enzymes	54
Table II	PROCHECK summary before and after energy minimization using X-PLOR	70
Table III	Kinetic analysis of wild type and mutant <i>SpHex</i> -MBP fusion protein	91

Abbreviations

α	-	Alpha
Å	-	Angstrom
A ₂₆₀ or A ₅₉₀	-	Absorbance at 260 nm or 590 nm respectively
ATP	-	Adenosine-5'-triphosphate
β	-	Beta
BSA	-	Bovine serum albumin
°C	-	Degrees Celsius
CaCl ₂	-	Calcium chloride
cDNA	-	Complimentary DNA
Ci	-	Curie
cm	-	Centimeter
cm ²	-	Square centimeter
cpm	-	Counts per minute
dATP	-	Deoxyadenosine 5'-triphosphate
dCTP	-	Deoxycytosine 5'-triphosphate
dGTP	-	Deoxyguanosine 5'-triphosphate
dNTP	-	Deoxynucleotide 5'-triphosphate
dTTP	-	Deoxythymidine 5'-triphosphate
ddH ₂ O	-	Double deionized water
Da	-	Dalton
DMSO	-	Dimethylsulfoxide
DNA	-	Deoxyribonucleic acid
DTT	-	Dithiothreitol
EDTA	-	Ethylenediamine tetraacetic acid
ER	-	Endoplasmic reticulum
g	-	Gram
<i>GM2A</i>	-	G _{M2} activator gene
H ₂ O	-	Water
H ₃ PO ₄	-	Phosphoric Acid
HCl	-	Hydrochloric acid
<i>HEXA</i>	-	Human <i>HEXA</i> gene
<i>HEXB</i>	-	Human <i>HEXB</i> gene
Hex A	-	Human β -N-acetylhexosaminidase isoform A
Hex B	-	Human β -N-acetylhexosaminidase isoform B
Hex S	-	Human β -N-acetylhexosaminidase isoform S
k	-	Kilo
KCl	-	Potassium chloride
L	-	litre
LB	-	Luria-Bertani
μ	-	Micro
M	-	Molar
m	-	Milli

MBP	-	Maltose binding protein
MgCl ₂	-	Magnesium chloride
MgSO ₄	-	Magnesium sulfate
min	-	Minute
4-MU	-	4-methylumbelliferone
4-MUG	-	4-methylumbelliferyl-2-acetamido-2-deoxy-β-D-glucopyranoside
4-MUGS	-	4-methylumbelliferyl-β-D-N-acetylglucosamine-6-sulfate
NEB	-	New England Biolabs
N-terminus	-	Amino-terminus
Na ₂ CO ₃	-	Sodium carbonate
Na ₂ HPO ₄	-	Sodium phosphate, dibasic
NaCl	-	Sodium chloride
NaOH	-	Sodium hydroxide
n	-	Nano
OD	-	Optical density
Ω	-	Ohm
PAGE	-	Polyacrylamide gel electrophoresis
PBS	-	Phosphate buffered saline
PCR	-	Polymerase chain reaction
PEG	-	Polyethylene glycol
rms	-	root mean square
RNA	-	Ribonucleic acid
RNase	-	Ribonuclease
S.A.	-	Specific activity
sec	-	Second
SDS	-	Sodium dodecyl sulphate
<i>SmChb</i>	-	<i>Serratia marcescens</i> chitobiase
<i>Sp</i> HEX	-	<i>Streptomyces plicatus</i> β-N-acetylhexosaminidase
SSC	-	Sodium citrate buffer
ssDNA	-	Single stranded DNA
Taq	-	Thermostable DNA polymerase from <i>Thermus aquaticus</i>
TBE	-	Tris borate EDTA buffer
TBS	-	Tris buffered saline
TBST	-	Tris buffered saline with Tween-20
Tris	-	Tris (hydroxymethyl) aminomethane
U	-	Units
V	-	Volts
v	-	Volume
W	-	Watts
w	-	Weight
x	-	Times
g	-	Gravity

ABSTRACT

A gene from *Streptomyces plicatus* encoding a β -N-acetylhexosaminidase (*SpHex*) was sequenced and the encoded protein identified as a member of family 20 glycosyl hydrolases. This family includes human β -N-acetylhexosaminidases whose deficiency results in various forms of G_{M2} gangliosidosis. Based upon the X-ray structure of *Serratia marcescens* chitobiase (*SmChb*), a three-dimensional model of *SpHex* was generated by comparative molecular modeling. The overall structure of the enzyme is very similar to homology modeling derived structures of human β -N-acetylhexosaminidases, with differences being confined mainly to loop regions. From previous studies of the human enzymes, sequence alignments of family 20 enzymes, and analysis of the *SmChb* X-ray structure, putative *SpHex* active site residues were selected and mutated. An Arg162His mutation increased K_m 40 fold and reduced V_{max} 5 fold, providing the first biochemical evidence for this conserved Arg residue (Arg178 in human β -N-acetylhexosaminidase A (HexA) and Arg349 in *SmChb*) as a substrate binding residue in a family 20 enzyme, a finding consistent with our 3-dimensional model of *SpHex*. Glu314Gln reduced V_{max} 296 fold, K_m 7 fold and altered the pH profile, consistent with it being the catalytic acid residue as suggested by our model and other studies. Asp246Asn reduced V_{max} 2 fold and increased K_m only 1.2 fold, suggesting Asp246 may play a lesser role in the catalytic mechanism of this enzyme. Taken together with the X-ray structure of *SmChb*, these studies suggest a common catalytic mechanism for family 20 glycosyl hydrolases.

1 INTRODUCTION

1.0 β -N-acetylhexosaminidase, an O-glycosyl hydrolase

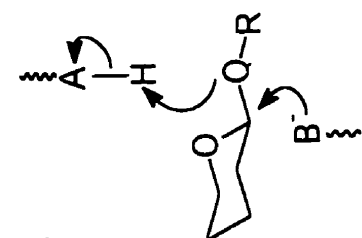
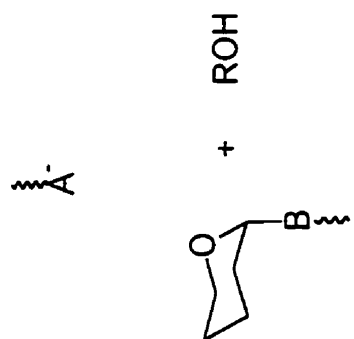
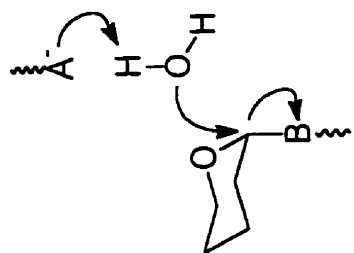
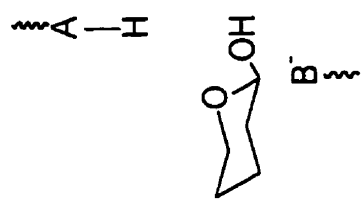
β -N-acetylhexosaminidase is responsible for the hydrolysis of terminal *N*-acetylglucosamine or *N*-acetylgalactosamine residues from the non-reducing end of oligosaccharides and their conjugates in organisms ranging from bacteria to humans (1). The terminal sugar residues are attached to an oxygen of the adjacent sugar by a β -1,4 linkage. Because it is the hydrolysis of the oxygen, or the glycosidic linkage, that results in the removal of these terminal sugar residues from the oligosaccharides and their conjugates, β -N-acetylhexosaminidase is referred to as an *O*-glycosyl hydrolase. β -N-acetylhexosaminidase is only one of hundreds of *O*-glycosyl hydrolases that are required to degrade the tremendous variety of carbohydrate chains and possible linkage configurations that exist. To better understand the evolutionary and mechanistic properties of *O*-glycosyl hydrolases, Henrissat (2; 3) classified these enzymes into families based on amino acid sequence similarity. Consequently, enzymes within a family tend to have similar structures. Although some of these *O*-glycosyl hydrolase families contain enzymes with dissimilar substrate specificities, it has been found that all enzymes belonging to a particular family have the same anomeric specificity and all members of a family either retain or invert the anomeric configuration of the substrate upon hydrolysis (Fig 1)(2). To date, 66 *O*-glycosyl hydrolase families have been identified.

β -N-acetylhexosaminidases and chitobiases (EC 3.2.1.52) belong to family 20 of the *O*-glycosyl hydrolases. In mammals, β -N-acetylhexosaminidase is a lysosomal hydrolase that participates primarily in the degradation of the glycosphingolipid G_{M2} ganglioside (Fig 2)

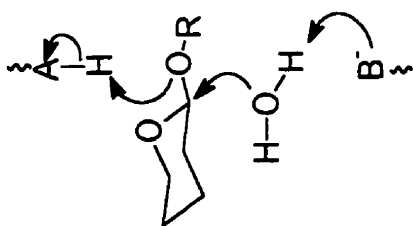
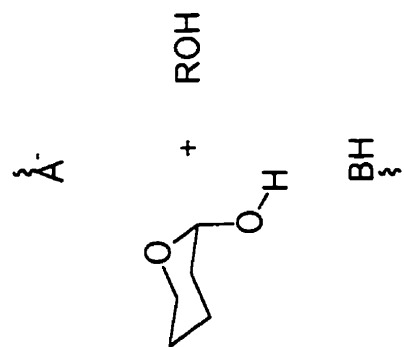
Figure 1 Mechanisms of glycosidic bond hydrolysis as proposed by Koshland (4).

Both mechanisms require a proton donor (acid) and a nucleophile (base) (reviewed in (5)).

Panel A, is a schematic diagram of the **retaining mechanism** whereby the glycosidic oxygen is protonated by the proton donor (A-H) and aglycon departure is assisted by the nucleophile (B⁻). The glycosyl-enzyme intermediate is hydrolyzed by a water molecule at the anomeric carbon of the substrate intermediate, producing a product with the same anomeric configuration as the original substrate. *Panel B*, is a schematic diagram of the **inverting mechanism** whereby no glycosyl-enzyme intermediate is formed. The protonation of the glycosidic bond and attack by a water molecule (activated by the nucleophilic base) occur simultaneously to produce a product with a stereochemistry opposite to that of the original substrate. Figure adapted from (6).

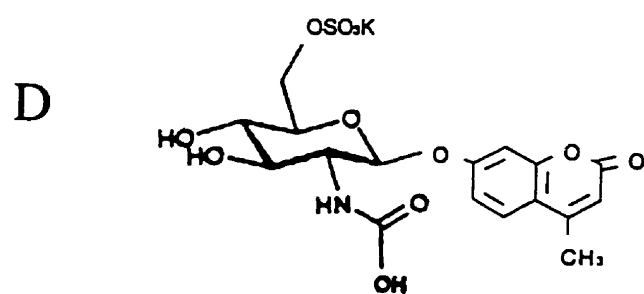
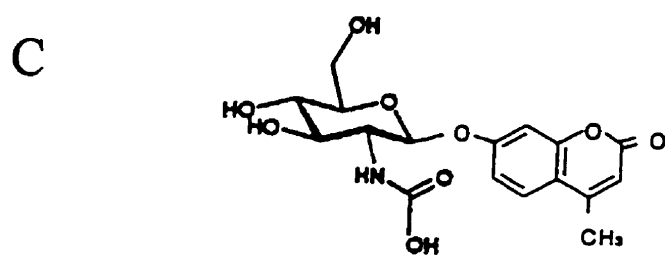
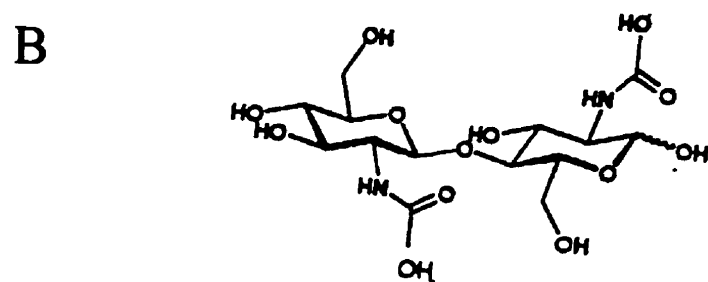
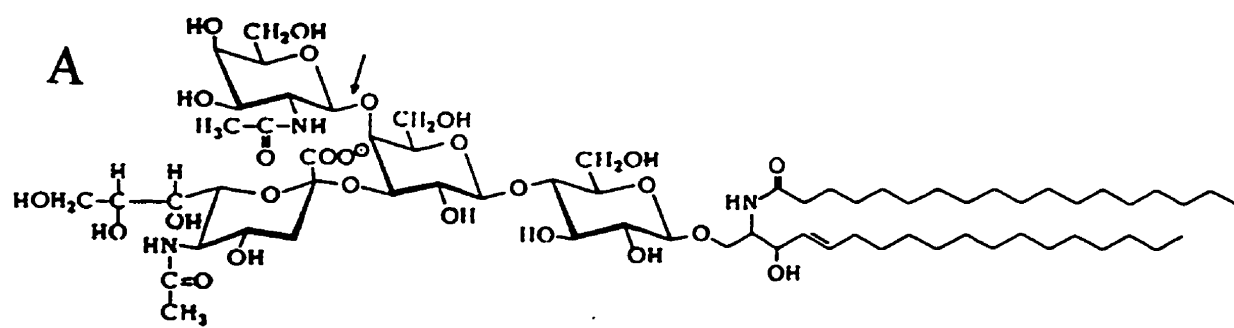


A



B

Figure 2 Structures of G_{M2} ganglioside, chitobiose and the substrate analogs 4-methylumbelliferyl- β -N-acetylglucosaminide (4-MUG) and 4-methylumbelliferyl-6-sulfo- β -N-acetylglucosaminide (4-MUGS). *A*, Structure of G_{M2} ganglioside. The arrow indicates the glycosidic bond hydrolyzed by human β -N-acetylhexosaminidase A. *B*, Structure of chitobiose (diN-acetylglucosamine). *C*, Structure of 4-MUG. *D*, Structure of 4-MUGS. Hydrolysis of 4-MUG and 4-MUGS results in the liberation of 4-methylumbelliferyl (4-MU) and can be measured fluorimetrically.



(7). In bacteria, β -*N*-acetylhexosaminidases are thought to be involved in the degradation of chitin and other carbohydrate isoforms that are used as sources of carbon and nitrogen (1). Unlike β -*N*-acetylhexosaminidases which can catalyze the removal of *N*-acetylglucosamine or *N*-acetylgalactosamine from oligosaccharides and their conjugates, chitobioses hydrolyze the β -1,4 glycosidic linkage between di*N*-acetylglucosamine (chitobiose) residues only (Fig 2). Chitobiase activity is also found in organisms ranging from humans to bacteria. In mammals however, only humans and rats have been reported to have a chitobiase, where it is believed to perform the last catalytic step in the degradation of oligosaccharide side chains from glycoproteins by hydrolyzing chitobiose after its removal from Asn by glycosylasparaginase (8).

1.1 Human β -*N*-acetylhexosaminidases and the G_{M2} gangliosidoses

The primary substrate for human β -*N*-acetylhexosaminidase, G_{M2} ganglioside, is abundant in the plasma membrane of many neuronal cells (reviewed in (9)). The majority of this ganglioside is found in the plasma membrane of neurons within the gray matter of the brain (10); however, the physiological function of this ganglioside in the neuronal plasma membrane is unknown. Nonetheless, human β -*N*-acetylhexosaminidase deficiency leads to a pathological accumulation of G_{M2} ganglioside in lysosomes within the nerve endings and dendrites of neurons throughout the nervous system (9). This accumulation of G_{M2} ganglioside results in a group of heritable neurological disorders collectively known as the G_{M2} gangliosidoses; the most severe and common form is referred to as Tay-Sachs disease. Initial clinical characterization and morphological features of G_{M2} gangliosidosis were first described by Warren Tay and Bernhard Sachs over a century ago; thus, the disorder is

referred to as Tay-Sachs disease (TSD).

TSD is a fatal autosomal recessive disease that has been found to occur in ethnic groups throughout the world; however, the highest incidence of the disease (1/3600) occurs in the Ashkenazi Jewish population where there is a carrier rate of approximately 1 in 35. This contrasts greatly with the non-Jewish carrier rate of 1 in 300 (reviewed in (9)). Fortunately, the frequency of TSD has been reduced by more than 90% since the introduction of enzyme based screening programs in the 1970's and many of the genetic mutations responsible for various disease forms have been documented and used to enhance the screening programs (11; 12). For the reasons stated above, human β -N-acetylhexosaminidase has been the most extensively studied family 20 enzyme since its discovery and association with Tay-Sachs disease in 1969 (7); therefore, it has become a model family 20 enzyme

1.1.0 Biochemistry

1.1.0.0 *HEXA* and *HEXB* gene structures and isoenzyme products

There are 3 isoforms of human lysosomal β -N-acetylhexosaminidase, all derived from combinations of two protein subunits α and β : Hex A is a heterodimer of α and β subunits, Hex B a homodimer of β subunits and the less common and unstable isoform Hex S, is a homodimer of α subunits (13). The α -subunit is encoded by *HEXA*, a 35 kb gene on chromosome 15 and the β subunit is encoded by *HEXB*, a 45 kb gene on chromosome 5 (14-17). Both genes contain coding regions of approximately 1600 bp divided into 14 exons with exons 2-14 having splice junctions at similar positions within the coding sequence (16). There exists a ~60% sequence identity between the coding regions of *HEXA* and *B* at both the nucleotide and amino acid levels (16). Because of this structural similarity between the genes,

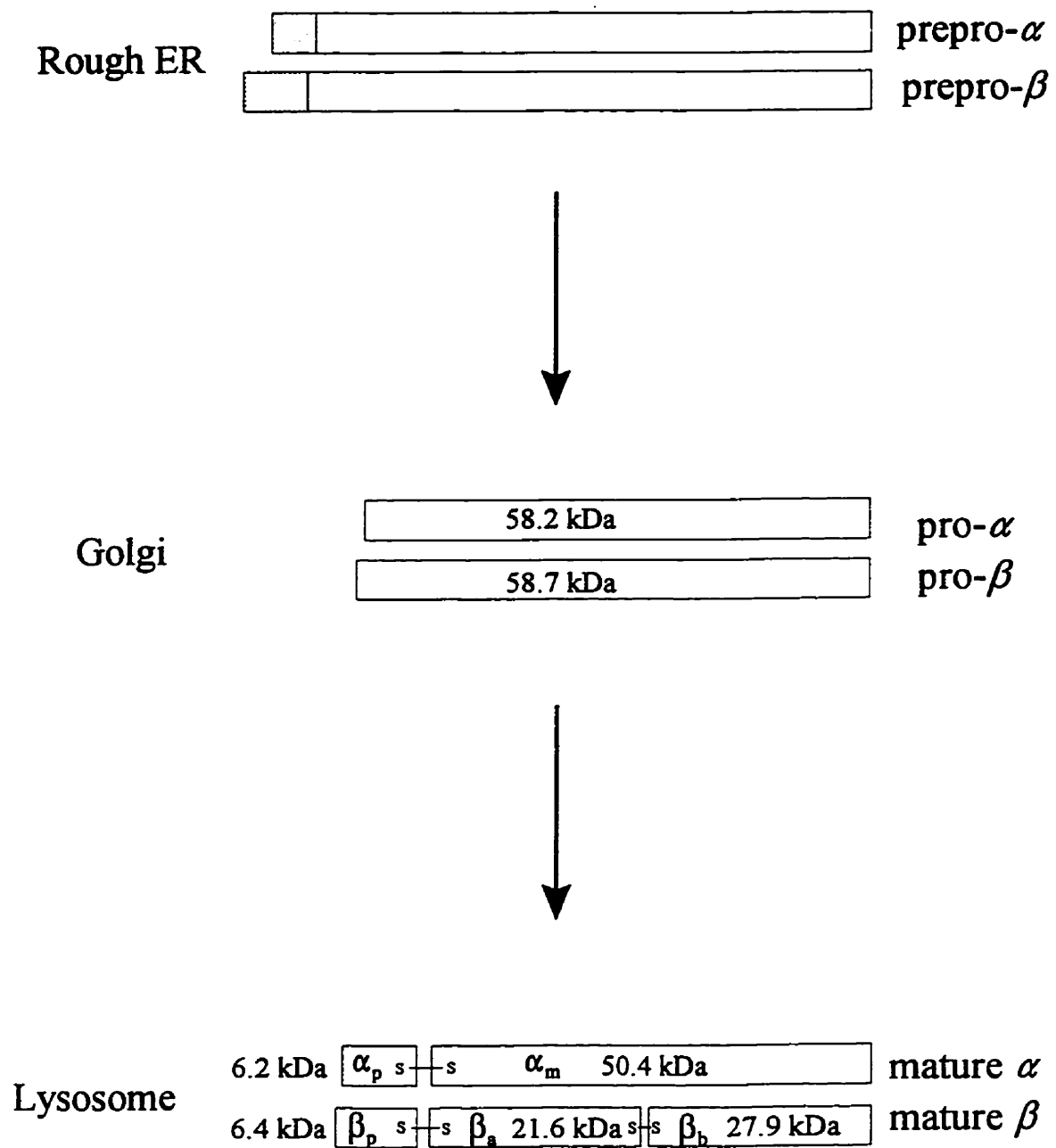
it is generally believed that *HEXA* and *HEXB* arose from a common ancestral gene.

1.1.0.1 Biosynthesis/processing

Human β -*N*-acetylhexosaminidase has a complex biosynthetic pathway (13). The mRNA transcripts from *HEXA* and *HEXB* code for an endoplasmic reticulum (ER) signal peptide, which, upon being translated, assigns the ribosome/mRNA complex to the rough ER where it can complete translation by feeding the entire polypeptide product into the rough ER lumen. During translation, the 23 and 43 amino acid ER signal peptides are proteolytically removed from the newly synthesized prepro- α and prepro- β polypeptides respectively, creating what are known as the pro- α and - β subunits (18; 19). Once in the ER, the pro- α and pro- β polypeptides experience extensive post-translational modifications as they are transported to the lysosome. Both pro-subunits become glycosylated at selected Asn residues (20; 21) and dimerize to form $\beta\beta$ (Hex B) and $\alpha\beta$ (Hex A) isoenzymes before being transported to the Golgi apparatus (22). At least one high mannose oligosaccharide side chain on each pro-subunit will acquire mannose 6-phosphate markers for lysosomal targeting during transit through the Golgi cisternae (23). Finally, the isoenzymes are recognized and bound by the mannose 6-phosphate receptor in the *trans*-Golgi and transported to the lysosome; while en route, the isoenzymes dissociate from the mannose 6-phosphate receptor as the environment within the transport vesicle acidifies (reviewed in (24)).

Once in the lysosome, the human β -*N*-acetylhexosaminidase isozymes undergo limited proteolysis, dephosphorylation and deglycosylation (25). It is the limited proteolysis of the pro- α and - β subunits in the lysosome that gives rise to mature enzyme (25). A schematic of the proteolytic processing of Hex A is given in figure 3.

Figure 3 **Proteolytic processing schematic for human Hex A.** The intracellular compartments through which the α and β -subunits are transported are indicated on the left. The ER signal peptides are shaded. The molecular weights of each subunit fragment are indicated and disulfide bonds are illustrated as “S-S”. To avoid masses contributed by oligosaccharide side chains, molecular weights were calculated from predicted amino acid sequences as given by Mahuran (25).



1.1.0.2 Substrate specificity and the G_{M2} activator protein

Both the α and β subunits are catalytically active once dimerization has occurred, even in the pro-form (26). Only the α -subunit of Hex A is capable of hydrolyzing the terminal *N*-acetylgalactosamine residue from the non-reducing end of the oligosaccharide complex of the G_{M2} ganglioside. Hence the Hex A isoform is crucial to the lysosomal degradation of G_{M2} ganglioside. The reason for this selectivity stems from the presence of a negatively charged sialic acid residue next to the terminal *N*-acetylgalactosamine in the G_{M2} ganglioside structure (Fig 2). The β -subunit can only hydrolyze neutral substrates (27).

The neutral or uncharged synthetic substrate 4-methylumbelliferyl- β -*N*-acetylglucosaminide (4-MUG) is commonly used to measure the catalytic activity of β -*N*-acetylhexosaminidases (Fig 2)(28). Both Hex A and B hydrolyze 4-MUG at a pH optimum of 4.4 and with similar K_m values of ~0.9 mM and V_{max} of 1.8×10^{-4} mol/min/mg and 4.4×10^{-4} mol/min/mg receptively (29-32). Because Hex B is heat stable whereas Hex A is not, the activities of the two enzymes can be differentiated by the introduction of a heat denaturation step during analysis of a sample for β -*N*-acetylhexosaminidase activity (33; 34). Furthermore, the negatively charged synthetic substrate 4-methylumbelliferyl-6-sulfo- β -*N*-acetylglucosaminide (4-MUGS) mimics G_{M2} ganglioside (Fig 2) and is preferentially hydrolyzed by Hex A with a K_m of 0.31 mM and V_{max} 25×10^{-4} mol/min/mg as opposed to Hex B which hydrolyzes this substrate with a K_m of 3.4 mM and a V_{max} of 1.2×10^{-4} mol/min/mg (27); thus, 4-MUGS is used in an assay devised to specifically measure Hex A activity (28).

In 1973, Li *et al.* (35) discovered that a water-soluble lipid-binding cofactor was

required in order for Hex A to catalyze the removal of the terminal *N*-acetylgalactosamine residue from G_{M2} ganglioside. This cofactor, referred to as the G_{M2} activator protein, is a lysosomal protein of molecular weight 22 kDa (36) and is encoded by the *GM2A* gene on chromosome 5 (37). Although the mechanism is poorly understood, G_{M2} activator protein specifically binds negatively charged glycosphingolipids, forming a water-soluble complex (38). Biochemical studies indicate that the activator/-lipid complex in turn, interacts with both subunits of Hex A in such a way as to present the appropriate glycosidic bond to the active site of the α -subunit (27).

1.1.1 Classification of G_{M2} gangliosidoses

Mutations in either *HEXA*, *HEXB* or *GM2A* can give rise to neuronal accumulation of G_{M2} ganglioside in one of three variants of G_{M2} gangliosidosis (reviewed in (9)): Variant B (Tay-Sachs disease), or type I G_{M2} gangliosidosis, refers to *HEXA* mutations that create an α -subunit defect or deficiency resulting in a lack of Hex A activity whereas Hex B activity appears normal. The catalytic activity of Hex A appears to be inversely proportional to the clinical severity of the disease (39). If there is very little or no Hex A activity, G_{M2} accumulation occurs quickly and the onset of neurodegeneration begins during infancy, resulting in what is referred to as infantile onset G_{M2} gangliosidosis or classic Tay-Sachs disease. *HEXA* mutations that have less of an effect on Hex A activity result in less severe forms of variant B G_{M2} gangliosidosis, and cause juvenile onset G_{M2} gangliosidosis, adult-onset or chronic G_{M2} gangliosidosis and Hex A pseudodeficiency. Variant O (Sandhoff disease) or G_{M2} gangliosidosis type II, refers to *HEXB* mutations that create a β -subunit defect or deficiency resulting in a lack of both Hex A and Hex B activity. Variant AB is a less

common form of G_{M2} gangliosidosis and occurs from mutations in the *GM2A* gene resulting in G_{M2} activator deficiency where Hex A and Hex B activities appear normal (32).

1.1.2 Identification of catalytic residues in human β -*N*-acetylhexosaminidase

Even though there is a great deal known about the gene structure, amino acid sequence, biosynthesis and post-translational modification, the active site of human β -*N*-acetylhexosaminidase remains to be completely characterized. Although human Hex B has been crystallized (40), no crystallographic information about the three-dimensional structure has been published because only low resolution data (3.4 Å) could be collected from the Hex B crystals. Until recently, the majority of the information regarding the active site of human β -*N*-acetylhexosaminidase was obtained from studying a particular subset of variant B mutations known as variant B1 mutations. In the majority of Tay-Sachs disease cases, the mutations in *HEXA* are so detrimental to the correct synthesis of the α -subunit that proper dimerization with the β -subunit rarely occurs. However, variant B1 mutations allow for the Hex A isozyme to reach maturity even though it is catalytically inactive toward 4-MUGS or G_{M2} ganglioside (41; 42). These Hex A isoenzymes containing B1 mutations do hydrolyze 4-MUG, demonstrating the presence of a functional β -subunit. Thus, it is thought that B1 mutations inactivate catalytically important residues in the α -subunit without affecting its ability to dimerize and be processed to the lysosome (42).

The most common and well characterized B1 mutation is Arg178His which causes late onset infantile Tay-Sachs disease mainly in people of Portuguese descent (43-46). When the Arg178His allele is found in combination with a totally non-functional allele, the onset of TSD is slightly earlier, indicating that Arg178His allows for enough residual activity to reduce

the severity of the clinical phenotype. Two other mutations in the same codon have been documented. The first is Arg178Cys and is biochemically very similar to Arg178His (46). The second is Arg178Leu and results in acute infantile Tay-Sachs disease (47).

An experiment which helped to substantiate the hypothesis that B1 mutations may be specific for catalytic residues was carried out by Brown *et al.* (48). They demonstrated that when the Arg178His mutation was introduced into the equivalent position in the β -subunit (Arg211His), mature Hex B was produced that was catalytically inactive. Furthermore, the same group demonstrated that the effect of a mutation at the Arg 211 codon in Hex B could be attenuated by substituting the Arg with a more similar residue such as Lys (49). Interestingly, the Arg211Lys produced mature Hex B with a dramatically reduced V_{max} ; however, it was shown that the K_m was unaffected, leading the investigators to conclude that Arg211 was somehow involved in the catalytic process but did not bind substrate.

Another well characterized B1 variant mutation, Asp258His, appeared to affect only the catalytic rate and shift the pH optimum of the α -subunit while not affecting its ability to bind substrate (50). Thus, based on the kinetic data of the Hex A containing the Asp258His mutation, it was postulated that Asp 258 may be the catalytic acid residue needed to hydrolyze the glycosidic linkage.

Two more B1 variant mutations have been documented Ser210Phe (51) and Trp420Cys (52). The role of these residues in the catalytic mechanism of Hex A is unknown except that it was proposed that Trp420Cys introduced a new cystine bridge, thereby affecting the catalytic properties of the α -subunit without affecting its ability to dimerize with the β -subunit (52).

As an alternative to studying naturally occurring mutations in order to identify catalytic residues, Liessem *et al.* identified Glu 355 as being an active site residue of Hex B based on photoaffinity labeling experiments (53). This experiment provided the first direct physical evidence of a residue in the active site of a β -*N*-acetylhexosaminidase. Alignment of amino acid sequences indicated that Glu 355 was highly conserved between the α -subunit of human β -*N*-acetylhexosaminidase, α and β -subunits of the mouse β -*N*-acetylhexosaminidase and the α -subunit of *Dictyostelium discoideum* β -*N*-acetylhexosaminidase. Interestingly, codon 355 in *HEXB* had never been found to be mutated naturally, indicating that studying variant B1 mutations alone would probably not reveal all the key residues involved in catalysis.

With the discovery of more β -*N*-acetylhexosaminidase and chitobiasases having considerable sequence similarity to the human α and β -subunit sequences, beginning with the *Vibrio harveyi* *N,N'*-diacetylchitobiasase sequence in 1989, investigators began to align these amino acid sequences and look for potential catalytic residues by virtue of their conservation. The highest conservation among family 20 enzymes occurs within a region corresponding to exons 5-6 of human *HEXA* and *HEXB* and for this reason was thought to encode the majority of the active site residues (54). In fact, this region includes α -Arg 178 (β -Arg211) and α -Asp258 is located adjacent to this region in exon 7. Tse et al. investigated the role of another invariant acid residue in this region of the β -subunit, β -Asp196 and found that upon substitution to Asn, the catalytic activity of Hex B significantly decreased whereas the K_m appeared unaffected (54). Hence three putative catalytic acid residues, α -Asp258, β -Asp196 and β -Glu 355 had been identified for the α and β -subunits of human β -*N*-

acetylhexosaminidase, two of which were within a highly conserved area of family 20 enzymes and one (Glu 355) encoded in exon 8, substantially downstream of the highly conserved area corresponding to exon 5-6 of human *HEXA* and *HEXB*.

While not definitive, the recent crystallization and determination of the three-dimensional structure of the family 20 enzyme *Serratia marcescens* chitobiase appears to have, by virtue of the structural conservation within family 20, resolved many elusive questions regarding the active site of human β -*N*-acetylhexosaminidase.

1.2 The *Serratia marcescens* chitobiase x-ray crystallographic model

The *Serratia marcescens* chitobiase (*SmChb*) is the only family 20 member for which a three-dimensional structure has been determined (55). *SmChb* is an 818 amino acid protein with a predicted molecular mass of 98.5 kDa (56). The 1.9 Å resolution crystallographic model of *SmChb* depicts a monomeric protein consisting of four domains. Domain III is the largest and forms an 8 stranded β/α -barrel containing the active site of the enzyme at the C-terminal end of the barrel. When the investigators solved the *SmChb* structure complexed with chitobiose, the proton of Glu540 was found to be within 2.5 Å of the oxygen in the glycosidic linkage of the chitobiose substrate, making it the prime candidate for the proton donor, or catalytic acid residue. Furthermore, the non-reducing *N*-acetylglucosamine residue of the chitobiose substrate was found to dock directly to Arg349 via hydrogen bonding, predicting Arg 349 as a key substrate binding residue within the *SmChb* active site. Furthermore, both Glu540 and Arg349 are conserved among all known family 20 enzymes.

Interestingly, the *SmChb* x-ray structure appeared to lack the necessary nucleophilic residue required for either the retaining or inverting catalytic mechanisms depicted in figure

2. Based on the structural information of *SmChb*, Tews *et al.* suggested that *SmChb* uses the retaining mechanism as described in figure 2 and that the acetamido group on the non-reducing *N*-acetylglucosamine residue of the chitobiose substrate (Fig 3) acts as the nucleophile (substrate assisted catalysis)(55). This theory has recently been substantiated by mechanistic studies on *SmChb* (57). Firstly, when using *p*-nitrophenol- β -*N*-acetylglucosaminide as substrate, it was found that the β -anomer of the *N*-acetylglucosaminide product predominated over the α -anomer, indicating that the anomeric configuration was conserved after catalysis and that *SmChb* was using the retaining mechanism described in figure 2. Secondly, it was demonstrated that *SmChb* could be selectively inhibited by oligosaccharide substrates that had been de-*N*-acetylated at their non-reducing end, supporting the theory that the *N*-acetamido group acts as a nucleophile and allows for substrate assisted catalysis to take place.

1.2.0 The *SmChb* structure and human β -*N*-acetylhexosaminidase

In the α subunit of human Hex A, α Glu323 is the homologue to *SmChb* Glu540; the homologous residue in the β subunit is Glu355. Mutagenesis studies based on these data identified Glu323 and Glu355 as likely proton donors in both the α and β subunits of human Hex A and B respectively. Mutation of α -Glu323 to Gln resulted in mature Hex A with normal substrate binding capacity but dramatically reduced catalytic activity towards 4-MUGS, indicating that Glu323 may act as a catalytic acid residue in the α -subunit (58).

Substitution of β -Glu355 to either Ala, Gln, Asp or Trp resulted in mature Hex A with normal substrate binding characteristics but no activity towards 4-MUG; however, all the mutants had normal activity toward G_{M2} ganglioside in the presence of G_{M2} activator protein

suggesting that Glu355 may serve as a catalytic acid residue in the β -subunit (59). In light of this data, it appears that the photoaffinity labeling experiment by Liessem et al. (53) had correctly identified the catalytic acid residue of the β -subunit active site and that the complementary residue in the α -subunit, Glu 323 also serves as a catalytic acid residue. Hence, these studies on Hex A and B provide substantial biochemical support for Glu540 of *SmChb* as the proton donor in the family 20 enzymes.

Using the *SmChb* x-ray structure as a template, Tews *et al.* were able to build three-dimensional homology models of the α and β -subunits of human Hex A (55). Based on the molecular model of the human α -subunit, α -Asp 258, previously believed to be the catalytic acid residue of the α -subunit, was found too far removed from the active site to be considered a catalytic acid residue and more likely to be indirectly involved in binding substrate through hydrogen bonding. Finally, the homology modeling demonstrated that α -Arg178 and β -Arg 211 in both the α and β -subunits respectively, were positioned directly within the active site and were potentially involved in substrate binding in a fashion similar to Arg 349 in *SmChb* as predicted from the x-ray structure. However, there is no biochemical evidence in the literature that supports the suggestion that Arg349 of *SmChb*, or its homologue in other family 20 enzymes, is a substrate binding residue.

1.3 Structure/function studies of family 20 enzymes

With the advent of high resolution structural data for a family 20 *O*-glycosyl hydrolase, homology modeling of any family 20 member is now possible. Detailed structure/function studies can be performed to verify the catalytic mechanism proposed by Tews *et al.*,(55), and to better understand the structural aspects of substrate specificity.

Modeling of the α and β -subunits of human β -N-acetylhexosaminidase has not only provided insight into the catalytic mechanism of human β -N-acetylhexosaminidase, but has also provided insight into the molecular basis of the different forms of G_{M2} gangliosidosis. Tews *et al.* mapped documented disease causing point mutations onto the homology model of the α -subunit and found that, in general, mutations resulting in Tay-Sachs disease, including the B1 variant mutations, occurred close to or within the active site, and mutations causing juvenile onset and less severe forms of G_{M2} gangliosidosis mapped to locations further away from the active site (55). As Tews *et al.* point out, the benign mutations Arg247Trp and Arg249Trp, responsible for causing Hex A pseudodeficiency, are found on the protein surface and have been shown to have a minor effect on the catalytic activity of Hex A, causing protein instability instead(60).

Unfortunately, structure/function studies on the human β -N-acetylhexosaminidase isoenzymes have been limited by two factors: 1) very low residual enzyme activity of mutant Hex A isoenzymes and 2) difficulty in producing sufficient levels of recombinant Hex A in mammalian expression systems required for detailed kinetic analysis (60; 61). Transient expression of mutant α -subunit cDNAs in monkey COS-7 cells is the most common mammalian expression system used. However, COS-7 cells exhibit a high level of endogenous Hex A and Hex B activity which tends to mask the low levels of β -N-acetylhexosaminidase activity expressed by mutant α -subunit cDNAs (reviewed in (61)). Furthermore, the predominant form of β -N-acetylhexosaminidase produced from the α -subunit cDNAs when expressed in COS-7 cells is a dimer of $\alpha\alpha$ subunits referred to as Hex S, an isoform that has physiological and biochemical properties different from those of Hex

A (61; 62).

In an attempt to overcome the problems associated with using COS-7 cells, mutant α -subunit cDNAs have been expressed in human fetal TSD neuroglial cells (TSD-NG). TSD-NG cells produce no endogenous α -subunits, allowing endogenous β -subunits to dimerize with mutant α -subunits expressed from transfected α -subunit cDNAs (reviewed in (58)). Although the problem of endogenous Hex A activity is overcome in this system, mutant Hex A levels are restricted by the level of endogenous β -subunit, resulting in low levels of Hex A for kinetic analysis. Hex S formation remains high and must be distinguished from Hex A by lengthy chromatofocusing experiments (58).

In order to overcome the low levels of expression inherent in mammalian expression systems, human β -subunit cDNAs have been expressed in the yeast *Saccharomyces cerevisiae* and in the *Spodoptera frugiperda* insect cell line Sf9 (63; 64). Interestingly, active Hex B was found localized to the yeast vacuole after transformation with β -subunit cDNA; however, the level of expression was far too low for any kinetic studies to be carried out (63). Transfection of the insect cell line Sf9 with recombinant baculovirus containing β -subunit cDNA, resulted in high level expression of active Hex B isozyme; however, correct glycosylation and proteolytic processing of Hex B did not occur (64). Nonetheless, the investigators conclude that this expression system appeared to produce properly folded Hex B in sufficiently large quantities to allow for future physical structure analysis of the enzyme by x-ray crystallography.

1.4 Current study

The complexity of structure/function studies of human β -N-acetylhexosaminidase

impelled us to search for bacterial enzymes that might be more amenable to such analyses. *Streptomyces plicatus* β -*N*-acetylhexosaminidase (*SpHex*) was identified as a candidate for study on the basis of substrate specificity.

The DNA clone encoding *SpHex* (Clone 36) was identified during a screen for chitinases using a *Streptomyces plicatus* expression library (65). Using 4-methylumbelliferyl (4-MU) glycosides of *N*-acetylglucosamine oligosaccharides as a substrate for an in-plate assay (66), several clones encoding enzymes that hydrolyzed the substrate were identified (65). Of these, only the *SpHex* enzyme, encoded by Clone 36, required a lag period for the production of 4-MU from these oligosaccharide derivatives of 4-MU. However, *SpHex* gave linear release of 4-MU from the monosaccharide derivative, 4-MUG (65). Since 4-MU was attached to the reducing end of the oligosaccharide substrates and β -*N*-acetylhexosaminidases can only degrade oligosaccharides by removing individual *N*-acetylglucosamine residues starting at their non-reducing ends, the lag period confirmed *SpHex* as a β -*N*-acetylhexosaminidase because it reflected the time required for the enzyme to completely degrade the oligosaccharide substrate before being able to liberate 4-MU.

The *SpHex* was characterized further in an attempt to develop it as a model β -*N*-acetylhexosaminidase to be used for detailed structure/function analyses.

2 Materials and Methods

2.0 Plasmids

pBluescript (pBS⁺) was purchased from Stratagene and was used for subcloning, ssDNA production and sequencing using the T7 and T3 promoters as priming sites.

psHEX-5.0 was a gift from Dr. P. Robbins and consisted of a 5.0 kb fragment (*EcoR* I/*EcoR* I) of *Streptomyces plicatus* DNA encoding a β -N-acetylhexosaminidase activity cloned into pBS⁺.

pmHEX-5.0 was a gift from New England Biolabs (NEB) and consisted of the 5.0 kb fragment (*EcoR* I/*EcoR* I) of *Streptomyces plicatus* DNA containing the coding region of a β -N-acetylhexosaminidase cloned in-frame with the *malE* coding region (maltose binding protein) of pMAL-c2. The pMAL-c2 plasmid is used to produce recombinant proteins fused to the maltose binding protein.

2.1 Bacterial culture

2.1.0 Bacterial strains

Escherichia coli JM109 (Promega) was used for all plasmid DNA manipulations and fusion protein production. Genotype: *recA1*, *supE44*, *endA1*, *hsdR17*, *gyrA96*, *relA1*, *thi*, $\Delta(lac-proAB)$, F'[*traD36*, *proAB*⁺, *laqI*⁺Z Δ M15].

Streptomyces plicatus (American Type Culture Collection # 27800) was used to obtain genomic DNA from which the 5'-end of the gene encoding *SpHEX* was cloned. Crude cytosolic fractions containing native *SpHex* were also obtained from *S. plicatus* and used for immunoblotting experiments.

2.1.1 Culture media

LB Media (2.5% [w/v] Miller's LB Broth Base [Gibco/BRL], autoclaved at 121 °C for 15 min). LBA broth was prepared by adding ampicillin (100 µg/ml) after the liquid had cooled to below 50 °C.

LB agar plates (2.5% [w/v] Miller's LB Broth Base, 0.75% [w/v] Bacto-agar, [Difco] autoclaved at 121 °C for 15 min). Plates were poured once the LB agar solution had cooled to 50 °C. LBA plates were prepared by adding ampicillin (100 µg/ml) as described above.

Low Salt LB (2.0 % [w/v] Lennox L Broth Base [Gibco/BRL], autoclaved at 121 °C for 15 min).

M9 Minimal Media was prepared by autoclaving a solution of M9 minimal salts (Gibco/BRL) (42 mM Na₂HPO₄, 22 mM KH₂PO₄, 8.5 mM NaCl and 19 mM NH₄Cl, pH 7.4). Once cooled, this solution was supplemented with sterile 1 mM MgSO₄, 0.1 mM CaCl₂, 0.05 % NaCl, 0.2 % glucose and 1 mM thiamine (final concentrations). If M9 minimal media plates were to be prepared, 0.75% (w/v) Bacto-agar (Difco) was added prior to autoclaving.

SOC media (1% [w/v] Bacto tryptone, 0.5 % [w/v] Bacto yeast extract, 8.6 mM NaCl, 2.5 mM KCl, pH 7.0, autoclaved at 121 °C for 15 min). Once cooled, the solution was supplemented with 20 mM glucose (final concentration).

2 x TY media (1.6% [w/v] Bacto tryptone, 1.0 % [w/v] yeast extract, 8.5 mM NaCl, pH 7.4, autoclaved at 121 °C for 15 min).

International Streptomyces Project (ISP) media (Difco) (0.4% [w/v] Bacto yeast

extract, 1% [w/v] Bacto malt extract, 0.4% [w/v] Bacto dextrose, autoclaved at 121°C for 15 min). ISP agar plates were made by the addition of 0.75% (w/v) Bacto-agar to the liquid ISP media prior to autoclaving and plates were poured once the solution cooled to 50°C.

2.1.2 Preparation of electrocompetent *E. coli* JM109

E. coli JM109 was streaked from stock onto LB agar and grown overnight at 37°C. A single colony was used to inoculate 5 ml of LB and the culture was grown overnight at 37°C, with shaking (300 rpm), then subcultured into 500 ml of LB and grown to a density of approximately 0.5 - 1.0 OD₆₀₀. The cells were pelleted by centrifugation at 4000 x g for 10 min at 4°C and washed once with 1 volume of sterile cold ddH₂O and pelleted as above. The cells were then washed twice with 0.5 volume and once with 0.02 volume of cold ddH₂O. Finally, the cells were resuspended in 0.002-0.003 volume of sterile-filtered cold 10% glycerol. Aliquots of 40 µl were snap frozen on dry ice and stored at -80°C. Electrocompetent cells were thawed 10 min on ice prior to electroporation.

2.1.3 Electroporation of electrocompetent *E. coli* JM109

Thawed electrocompetent cells were mixed with 1µl (~1-10 ng) of plasmid DNA suspended in sterile ddH₂O. The mixture was transferred to an ice-cold 1 mm gap electroporation cuvette and subjected to an electrical charge gradient using a BTX Inc. Electro cell manipulator 600 (settings: 2.5 kV high voltage, 129 Ω resistance, 1.3-1.5 kV charging voltage). Immediately after electroporation, the cells were resuspended in 960 µl 37°C SOC media and incubated at 37°C for 1 hr. Electroporated cells (10-200 µl) were plated onto LBA agar and grown at 37°C overnight. Single colonies were chosen for further

manipulations.

2.1.4 Growth and storage of transformed *E. coli* JM109

Unless otherwise stated, single colonies of transformed *E. coli* were selected from LBA agar and grown in 5 ml LBA medium overnight at 37°C with shaking (300 rpm). The cultures (~1.0 OD₆₀₀) were mixed with 80% filter sterilized glycerol to give a final concentration of 48 % glycerol and stored at - 80°C as stock.

2.1.5 Growth and storage *S. plicatus*

Lyophilized *S. plicatus*, purchased from ATCC, was resuspended in 3 ml of ISP medium and grown at 25°C for 3 days with shaking (200 rpm). The bacteria were subcultured into 250 ml of ISP medium and grown for a further 5 days. A single spore was retrieved from the culture and spread onto ISP agar. The plated *S. plicatus* was grown for 3 days at room temperature and then stored at 4°C as stock. The liquid culture was pelleted at 5000 x g for 10 min at 4°C and stored at -20°C after aspiration of the supernatant. To ensure viability over time, once every two weeks, 5-10 ml of ISP media was inoculated with a loopful of *S. plicatus* stock and grown for 3 days at room temperature with shaking (200 rpm). A single spore was plated onto fresh ISP agar, allowed to grow for 3 days and then stored at 4°C as fresh stock.

2.2 Plasmid and genomic DNA isolation

2.2.0 Alkaline lysis minipreparation of plasmid DNA

The alkaline lysis minipreparation of plasmid DNA was performed as described by Sambrook *et al.* (67). To destroy contaminating RNA in the plasmid DNA minipreparation, 1 ul of 10 mg/ml RNase A was added to the suspension and allowed to incubate at 37°C for

30 min before the miniprep was frozen at -20°C.

2.2.1 Large scale plasmid DNA preparation

This isolation procedure is according to instruction provided with the Nucleobond-AX column (Nest Group Inc.). A 250 ml *E.coli* culture, grown overnight at 37°C with shaking (300 rpm), was pelleted at 5000 x g for 10 min at 4°C. The supernatant was aspirated and the pellet resuspended in 12 ml of S1 buffer (50 mM Tris-Cl, 10 mM EDTA, 100 µg/ ml RNase A, pH 8.0). To lyse the cells, 12 ml of S2 buffer (200 mM NaOH, 1 % SDS) was added. The suspension was gently mixed and after a 5 min incubation at room temperature, 12 ml of S3 buffer (2.6 M potassium acetate, pH 5.2) was added. The suspension was gently mixed and allowed to incubate on ice for 5 min. The cell debris was pelleted by centrifugation at 12000 x g for 1 hr at 4°C. The supernatant was loaded onto an AX 500 Nucleobond column that had been equilibrated with 5-8 ml of N2 buffer (100 mM Tris, 15% ethanol, 900 mM KCl, pH 6.3). The column was washed with 24 ml of N3 buffer (100 mM Tris, 15 % ethanol, 1150 mM KCl, pH 6.3), and the plasmid DNA was eluted from the column with 6 ml of N5 buffer (100 mM Tris, 15% ethanol, 1000 mM KCl, pH 8.5). The eluate was mixed with 4.2 ml of 2-propanol and centrifuged at 12000 x g for 1 hr. The supernatant was aspirated and the pellet washed with 10 ml of ice cold 70% ethanol and allowed to air dry briefly before being resuspended in 200-400 µl of sterile ddH₂O.

2.2.2 DNA purification by phenol/chloroform extraction

To remove contaminating protein from plasmid DNA preparations, an equal volume of a 1:1 phenol/chloroform solution was added, the mixture was vortexed, and then centrifuged at 14000 x g for 20 sec to separate the organic and aqueous phases. The aqueous

phase was collected and the extraction repeated if precipitate was visualized at the interface of the two phases. This was followed by an extraction with an equal volume of chloroform. The aqueous phase was mixed with 0.1 volume of sodium acetate, pH 4.8, 2.5 volumes of 95% ethanol and placed at -20°C for 1 hr to precipitate the plasmid DNA. The precipitated plasmid DNA was pelleted by centrifugation at 14000 x g for 20 min at 4°C. The pellet was washed with 1-2 ml of ice cold 70% ethanol, briefly allowed to air dry and resuspended in an appropriate volume of sterile ddH₂O.

2.2.3 Isolation of *S. plicatus* genomic DNA

Isolation of *S. plicatus* genomic DNA was based on the method of Nisen *et al.* (68). However, the isolated genomic DNA was not dialyzed after phenol/chloroform extraction. The DNA was isolated from 10 ml cultures of *S. plicatus* that had been grown for 3 days (section 2.1.5). The genomic DNA was resuspended in 1 x TE buffer and stored at 4°C.

2.3 Determination of DNA and protein concentration

2.3.0 Spectrophotometric determination of DNA concentration

Assuming ddH₂O containing 50 µg/ml double stranded DNA has an absorbance of 1.0 at 260 nm and ddH₂O containing 40 µg/ml single stranded DNA has an absorbance of 1.0 at 260 nm (1 cm path length) (67) then the following calculation can be used to quantify unknown concentrations of double or single stranded DNA:

$$[\text{DNA}] (\mu\text{g}/\mu\text{l}) = A_{260\text{ nm}} \times 50(\text{dsDNA}) \text{ or } 40(\text{ssDNA}) \div \text{volume } (\mu\text{l}) \text{ of DNA sample added}$$

2.3.1 Visual estimation of DNA concentration

DNA Mass TM Ladder (Gibco/BRL) consisted of an equimolar mixture of six blunt ended DNA fragments ranging from 0.1 to 2 kb in length. Agarose gel electrophoresis

(Section 2.4.1) of this mixture separated out 6 bands containing 200, 120, 80, 40, 20 and 10 ng of DNA. The concentration of DNA sample could be visually estimated by comparing its fluorescent intensity to bands of the Low DNA Mass Ladder sample after electrophoresis in a 0.9% agarose gel containing ethidium bromide.

2.3.2 Colourometric determination of protein concentration

Protein concentration was determined using a Bio-Rad protein assay kit based on the Bradford method (69). To produce a standard curve, bovine γ -globulin stock (1 mg/ml) was diluted to 5, 10, 15 and 20 μ g/ml. This typically gave an absorbance range between 0.2 to 0.6 at 595 nm. Protein concentrations in test samples were determined only from dilutions falling within this linear range that was used to make the standard curve. Linear regression analysis was performed using the program Origin (Microcal).

2.4 DNA cloning techniques

2.4.0 Restriction enzyme digests

All restriction enzymes were from New England Biolabs (NEB) or Gibco/BRL and digests were performed according to their instructions. Briefly, depending on the NEB restriction enzyme used, the reaction mixture contained 1 x NEB reaction buffer 1,2,3 or 4 and 100 μ g/ml BSA, 0.1-5 μ g of DNA and 1-5 units of restriction enzyme. For Gibco/BRL restriction enzymes, digestion reactions contained the appropriate Gibco/BRL enzyme buffer diluted 10 fold, 0.1-5 μ g of DNA and 1-5 units of restriction enzyme. To ensure that the glycerol concentration did not become exceedingly high, the concentration of restriction enzyme never exceeded 10% of the total reaction volume. Digestion reactions were incubated at temperatures recommended by the manufacturer for 1-3 hrs.

2.4.1 Agarose gel electrophoresis of DNA

Agarose gel electrophoresis was used to separate plasmid DNA and DNA fragments based on size for analysis and further manipulations using a Gibco/BRL Horizon™ 11-14 electrophoresis apparatus. Agarose gels consisted of 0.7-1.0% molecular biology grade agarose (Gibco/BRL), 1 x TBE (0.1 M Tris, 0.09 M boric acid, 0.2 mM EDTA, pH 8.0) and 10 mg/ml ethidium bromide. 1 x TBE was used as running buffer. DNA samples to be separated by electrophoresis were mixed 10:1 with 10 x loading buffer (0.25% (v/v) bromophenol blue, 0.25% (v/v) xylene cyanol, 30% (v/v) glycerol). To determine the size of DNA fragments separated in the gel, Gibco/BRL 1 Kb DNA Ladder (100 ng/μl in 1 x loading buffer) was separated by electrophoresis in parallel with the test samples. DNA Mass Ladder could also be added to one of the wells if the DNA concentration needed to be estimated (Section 2.3.2). DNA was visualized using a Hoefer Mighty Bright™ UV transilluminator and photographed with a Polaroid MP4 Land Camera. The resolution range of agarose gel electrophoresis was found to be between 7-0.1 kb.

2.4.2 Polyacrylamide gel electrophoresis of DNA

Polyacrylamide gel electrophoresis was used for accurate analysis of multiple DNA fragments smaller than 200 bp using a Bio-Rad Mini-Protean™ II cell apparatus. DNA fragments were prepared as described in section 2.4.3 and separated by electrophoresis through 1 mm 8-20% acrylamide:bisacrylamide (29.2:0.8) made with 1 x TBE. 1 x TBE was also used as running buffer. For determination of molecular size 1 Kb DNA Ladder was separated by electrophoresis in parallel with the test samples. Gels were run in 1 x TBE at 100-150 V until the bromophenol blue dye front migrated to approximately 1 cm from the

bottom of the gel. The gel was then removed from the apparatus, stained in a 0.1% ethidium bromide solution for 15 min and photographed as described in section 2.4.1.

2.4.3 DNA purification using the GeneClean II kit

Purification of DNA fragments from agarose gels was accomplished using a GeneClean II kit (Bio101). Briefly, following electrophoresis (section 2.4.1), separated DNA fragments were visualized using a hand held UV light source and individual bands excised from the gel. The gel pieces were weighed, mixed with 4.5 volumes of 6 M NaI (supplied with kit), 0.5 volumes of TBE modifier (supplied with kit) and incubated at 55°C for approximately 5 min. Depending on the amount of DNA in the sample, 5-10 µl of Glassmilk® (concentrated silica matrix particles supplied with kit) were mixed with the melted agarose gel solution and the mixture was incubated on ice for 5-10 min.

The Glassmilk®, now bound to the DNA, was pelleted by centrifugation at 14000 x g for 5 sec, resuspended in 10 to 50 volumes of New Wash (supplied with kit) and pelleted by centrifugation at 14000 x g for 5 sec. The Glassmilk® was washed twice more with New Wash before being resuspended in 10-20µl of sterile ddH₂O, incubated at 55°C for at least 5 min then pelleted by centrifugation at 14000 x g for 5 sec. The supernatant, now containing the eluted DNA, was removed from the pelleted Glassmilk® and was ready for further manipulation.

2.4.4 Alkaline phosphatase treatment of DNA

Calf intestinal alkaline phosphatase (CIP) (NEB) was used to remove 5' phosphate groups from DNA fragments. Reaction volumes were typically 10-20µl and contained 0.1-0.5 µg of DNA and 1-2 units of CIP in 50 mM NaCl, 10 mM Tris-Cl, 10 mM MgCl₂, 1 mM

dithiothreitol, pH 7.9. Reactions were incubated at 37°C for 1 hr after which the DNA was purified by phenol/chloroform extraction (Section 2.2.2).

2.4.5 Filling in 3' recessed ends of DNA

DNA polymerase I, Large (Klenow) fragment (NEB) lacks 5'→3' exonuclease activity; however, it retains its 3'→5' polymerase activity and was used to fill in 3' recessed ends of dsDNA. Reaction volumes were typically 10-20 µl and contained 0.1-0.5 µg of DNA and 3-8 units of Klenow fragment in 10 mM Tris-Cl, pH 7.5, 5 mM MgCl₂, 7.5 mM dithiothreitol, supplemented with 1 mM dNTPs (NEB). Reactions were incubated at 37°C for 1 hr.

2.4.6 Ligations

T4 DNA ligase (Boehringer Mannheim) was used to catalyze the formation of phosphodiester bonds between juxtaposed 3'-hydroxyl and 5'-phosphate ends of dsDNA fragments. Reaction volumes were typically 10-20 µl and contained 0.1-0.5 µg of DNA and 1-10 units of T4 DNA ligase in 66 mM Tris-Cl, 5 mM MgCl₂, 1 mM dithiothreitol and 1 mM ATP, pH 7.5. Reactions were incubated at 14°C overnight and 1 µl of the reaction was electroporated into *E. coli* (Section 2.1.3).

2.5 Analysis of *S. plicatus* Genomic DNA by Southern hybridization

All procedures used for genomic DNA analysis were based on methods described by Sambrook *et al.* (67).

2.5.0 Preparation of Southern blots

Genomic DNA (4-10 µg) was digested with appropriate restriction enzymes and separated by electrophoresis through a 0.7% agarose gel as described in section 2.4.1. The DNA was transferred to Hybond-H nylon membrane (Amersham) by capillary transfer for 12-

24 hr using 10 x SSC (1.5 M NaCl, 0.15 M trisodium citrate) as buffer. After the transfer was complete, the DNA was cross-linked to the nylon membrane using a Bio-Rad Gene Linker UV chamber.

2.5.1 Random primer labeling of the DNA probe

The DNA fragment to be labeled (30-100ng) was combined with 5 µg of random hexanucleotides in a total volume of 14 µl, boiled for 2-3 min and then placed on ice. Once cooled to 4°C, 2.5 µl of 0.5 mM dATP, 2.5 µl of 10 x Klenow buffer (100 mM Tris-Cl, pH 7.5, 50 mM MgCl₂, 75 mM dithiothreitol), 5.0 µl of [α -³²P] dATP (50 µCi) and 1.0 µl of Klenow fragment (3-8 units) was added to the randomly primed DNA probe and the reaction incubated for 3 hr at room temperature.

The nucleotides were separated from the labeled DNA using a Pharmacia nick column. The column was washed with 5 ml of 1 x TE (10 mM Tris-Cl, pH 7.5, 1 mM EDTA). The labeling reaction was stopped by the addition of 1 µl of 0.5 M EDTA, pH 7.5 and brought to 100 µl total volume with ddH₂O before being loaded onto the column. Once the labeling reaction had entered into the column, 100 µl aliquots of 1 x TE were added to the column and 6 fractions were collected. A 2 µl aliquot from each fraction was counted using a Beckman LS 3801 scintillation counter and the fraction with the highest specific activity of labeled DNA was used for subsequent hybridizations.

2.5.2 DNA/DNA hybridization

The Southern blot was prehybridized at 65°C for 3 hr in prehybridization solution consisting of 6 x SSC, 0.1% (w/v) BSA, 1% (w/v) Ficoll 400, 0.1% (w/v) polyvinylpyrrolidone, 0.5 % SDS, 100 µg/ml denatured, fragmented salmon sperm DNA in

a Robbins Model 2000 Micro Hybridization incubator. The prehybridization solution was replaced with an equal volume of hybridization solution consisting of 6 x SSC, 0.5% SDS, 100 µg/ml denatured, fragmented salmon sperm DNA and at least 10⁶ cpm/ml labeled probe. The hybridization was carried out at 65°C for 12-18 hr. The ³²P-labeled DNA fragments were detected by autoradiography for 12-24 hr at -80°C using Kodak X-Omat AR X-ray film and a Fisher Biotech L-plus intensifying screen.

2.6 Preparation of single stranded DNA

Single stranded DNA (ssDNA) was used as template for site directed mutagenesis and DNA sequencing. Preparation of ssDNA was adapted from the method of Carter (70). *E. coli* containing the F' episome were selected on M9 minimal media agar (section 2.1.1) and used to make electrocompetent cells (Section 2.1.2) for ssDNA production. A single colony of *E. coli* (F') transformed with phagemid was selected and used to inoculate 5 ml 2 x TY/ampicillin (100 µg/ml) culture that was grown overnight at 37°C with shaking (300 rpm). A 2 ml aliquot of the culture was used to inoculate 20 ml of 2 x TY/ampicillin and 100 µl of M13KO7 helper phage (NEB) was added. After incubation for 90 min at 37°C with shaking (300 rpm), kanamycin sulfate (50 -70 µg/ml) was added and the culture incubated for a further 12 hr as described above.

The *E. coli* were pelleted by centrifugation at 6000 x g for 7-10 min at 4°C. The pellet was discarded and the supernatant was mixed with 3.6 ml of 20% (w/v) polyethylene glycol (PEG) 6000, 2.5 M NaCl and incubated at room temperature for 30 min. Precipitated phagemid were pelleted by centrifugation at 12000 x g for 10-20 min at 4°C and the supernatant discarded. The phagemid pellet was resuspended in 1 ml of TE buffer (10 mM

Tris-Cl, pH 8.0, 1 mM EDTA) containing 0.1 mg/ml RNase A and the packaged ssDNA was phenol/chloroform extracted as described in section 2.2.2 except that the ssDNA was precipitated for 20 min at room temperature with 0.5 volumes 7.5 M ammonium acetate and 3 volumes 95% ethanol. The purified ssDNA pellet was resuspended in 50 µl of sterile ddH₂O and stored at -20°C.

2.7 DNA sequencing

DNA sequencing was performed using a T7 Sequencing™ Kit (Pharmacia Biotech). based on the method of Sanger *et al* (71).

2.7.0 Alkali denaturation of double-stranded plasmid DNA

Phenol/chloroform extracted double stranded plasmid DNA (2-3 µg in 32 µl ddH₂O) was combined with 8 µl of 2 M NaOH and incubated for 10 min at room temperature. The denatured DNA was precipitated by adding 7 µl of 3 M sodium acetate, 4 µl of ddH₂O, 120 µl of 95% ethanol and placing at -80°C for at least 15 min. The DNA was pelleted by centrifugation at 14000 x g for 20 min at 4°C, washed once with ice cold 70% ethanol, briefly air dried and resuspended in 10 µl of ddH₂O. If DMSO was to be added to the annealing reaction described below, the DNA was resuspended in 7µl of dd H₂O.

2.7.1 Annealing and sequencing reaction

For annealing reactions not requiring DMSO, a 10 µl volume of ddH₂O containing 2-3 µg of single stranded or denatured double stranded DNA was mixed with 2 µl of sequence specific oligonucleotide (10 ng/µl), 2 µl of annealing buffer (1 M Tris-Cl, pH 7.6, 100 mM MgCl₂, 160 mM DTT) and incubated at 65°C for 5 min, 37°C for 10 min and room temperature for at least 5 min. If DMSO was required for successful sequencing of GC rich

double stranded DNA, 3 μ l of DMSO was added to 7 μ l of denatured double stranded DNA (2-3 μ g) followed by 2 μ l of sequence specific oligonucleotide (10 ng/ μ l) and 2 μ l of annealing buffer. The reaction was incubated at 37°C for 20 min and then for at least 10 min at room temperature before proceeding to the sequencing reaction.

The primer was extended by adding 3 μ l of labeling mix (1.375 μ M each dCTP, dGTP and dTTP, 333.5 mM NaCl), 1 μ l [α -³⁵S] dATP and 2 μ l of T7 DNA polymerase (diluted 4 fold in 20 mM Tris-Cl, pH 7.5, 5 mM DTT, 100 μ g BSA/ml, 5% glycerol). The reaction was incubated at room temperature for 5 min and then 4.5 μ l aliquots of the reaction were added to 2.5 μ l of prewarmed (37°C) short-mix A, C, G, and T (Short-mix A: 40 mM Tris-Cl, pH 7.6, 50 mM NaCl and 840 μ M of dCTP, dGTP and dTTP; 93.5 μ M dATP and 14 μ M ddATP. Short-mix C: 40 mM Tris-Cl, pH 7.6, 50 mM NaCl and 840 μ M of dATP, dGTP and dTTP; 93.5 μ M dCTP and 14 μ M ddCTP. Short-mix G: 40 mM Tris-Cl, pH 7.6, 50 mM NaCl and 840 μ M of dATP, dCTP and dTTP; 93.5 μ M dGTP and 14 μ M ddGTP. Short-mix T: 40 mM Tris-Cl, pH 7.6, 50 mM NaCl and 840 μ M of dATP, dCTP and dGTP; 93.5 μ M dTTP and 14 μ M ddTTP). All four reactions were incubated at 37°C for 5 min then stopped with 5 μ l stop solution (0.3% bromophenol blue, 0.3% xylene cyanol FF, 10 mM EDTA, pH 7.5 and 97.5% deionized formamide).

2.7.2 Sequencing gel electrophoresis

Sequencing reactions were denatured at 100°C for 3 min and separated at 60 W for 2-4 hr in a 6% polyacrylamide/42% urea gel made with 1 x modified TBE (134 mM Tris, 44 mM Boric acid, 2.5 mM EDTA) using a model S2 BRL sequencing gel apparatus. Gels were transferred onto Gl002 gel lift paper (Schleicher and Schuell) and dried for 1 hr at 80°C. The

separated ^{35}S -labeled DNA fragments were detected by autoradiography for 12-24 hr at room temperature using Kodak X-Omat AR X-ray film.

2.8 Comparative molecular modeling

2.8.0 Multiple amino acid sequence alignments

The ClustalW1.7 program was used for multiple sequence alignments (72; 73). The multiple alignments contained as many related sequences as possible including the target sequence to be modeled and the sequence of at least one protein whose X-ray structure had been determined (reference structure). To improve the multiple alignments, sequences which increased the frequency of insertions and deletions were sequentially removed and the remaining sequences realigned. This minimized the number of insertions and deletions within the target sequence, and maximized alignments of the secondary structural elements as visualized using the graphics program O (74). The final alignment between the reference and target sequences was optimized manually by comparing initial 3-dimensional models of the target sequence to the reference structure using the program O (75).

2.8.1 Model building

Using the program Homology (Insight software, BioSym Technologies, Inc.), the aligned amino acid sequence of the target sequence was substituted onto the backbone atomic coordinates of the reference structure. Initial side-chain steric clashes were relieved using a rotamer library containing the most favorable residue side-chain conformations. Gaps were manually joined using the program O. Moreover, O was used to relieve any serious steric clashes within the model.

2.8.2 Energy minimization

In order to relieve any energetically unfavorable steric clashes and geometry that raise the overall free energy of the model (introduced when the amino acids of the target structure were substituted onto the backbone atomic coordinates of the reference structure), the model was brought to its lowest possible free energy state by repeatedly introducing small rotational corrections to the main-chain dihedral angles (ϕ, ψ) using X-PLOR (76; 77). The model was placed in a primitive lattice (P1 space group) and subjected to energy minimization using a conjugate gradient target function implemented in X-PLOR. The unit cell parameters were set sufficiently large to avoid intermolecular contacts during minimization. Minimization was performed with the van der Waals and electrostatic terms turned on. The model was minimized until convergence was reached, as judged by the rms of the energy gradient (average derivative < 0.1 kcal/mol/Å).

2.8.3 Quality assessment

In order to assess the overall stereochemical quality of the model, the program PROCHECK (78) was used to determine 1) the number and positions of any bad contacts (any atomic interaction less than 2.5 Å) within the model, and 2) The overall geometrical quality of the main-chain bond angles and lengths and planar groups within the model. The geometrical quality of the model was based on comparison values typical for a 2.0 Å resolution X-ray structure.

2.9 Site-directed mutagenesis

2.9.0 Annealing of the oligonucleotide to ssDNA

Site directed mutagenesis using ssDNA template was performed following instructions

from Bio-Rad based on the procedure of Kunkel *et al.* (79) but using T7 polymerase (80). Briefly, to anneal a complementary oligonucleotide primer containing the desired base changes to the ssDNA template, a 10 µl mixture of 200 ng ssDNA, 100 ng oligonucleotide, 20 mM Tris-Cl, pH 7.4, 2 mM MgCl₂ and 50 mM NaCl was brought to 70°C and slowly cooled (~1°C/ min) to 30°C before being placed on ice for 10 min.

2.9.1 Synthesis of complementary DNA strand

The annealed primer was extended by the addition of 1 µl of 10 x synthesis buffer (4 mM dNTP, 7.5 mM ATP, 175 mM Tris-Cl, pH 7.4, 37.5 mM MgCl₂, 215 mM dithiothreitol), 1 µl T4 DNA ligase (2-5 units) and 1 µl T7 DNA polymerase (diluted to 0.5 units/ µl with 20 mM KPO₄, pH 7.4, 1 mM DTT, 0.1 mM EDTA, 50% glycerol) to the annealing reaction. The reaction was placed on ice for 5 min then brought to 25°C for 5 min, 37°C for 30 min and then immediately separated on a 0.9% agarose gel as described in section 2.4.1. The double stranded DNA product was purified using a Geneclean II kit (section 2.4.3). The product (20-40%) was transformed into *E. coli* and colonies containing mutant plasmid constructs were identified by PCR amplification of the relevant region followed by restriction enzyme digestion.

2.9.2 Polymerase chain reaction (PCR)

The polymerase chain reaction (81) was used to amplify specific regions of plasmid DNA. Reactions (100 µl each) contained 10 mM Tris-Cl, pH 8.3, 50 mM KCl, 1.5 mM MgCl₂, 0.01% gelatin, 0.2 mM dNTP (dATP, dCTP, dGTP, dTTP from NEB), 10% glycerol, 100 ng sense primer, 100 ng anti-sense primer, 0.3 U Taq DNA polymerase (Gibco/BRL) and a portion of a transformed bacterial colony.

Reactions were overlaid with approximately 50 μ l mineral oil and incubated at 94°C for 10 min then immediately subjected to 32 PCR cycles in a Perkin Elmer Cetus DNA thermo cycler. One cycle included 90 sec of denaturation at 94°C, 1 min of annealing at 60°C and 90 sec of extension at 72°C.

2.9.3 Restriction enzyme analysis of PCR products

Site directed mutations were designed such that they created or destroyed a restriction enzyme site. Thus, when PCR products amplified from plasmid containing site-directed mutations were digested with the relevant restriction enzyme, the banding pattern was different from the PCR product amplified from the corresponding wild type plasmid and this difference could be visualized using polyacrylamide gel electrophoresis (section 2.4.2).

2.10 Expression and purification of pMAL-c2 derived fusion proteins

Purification of the pMAL-c2 derived fusion protein was based on instructions from NEB. *E. coli* transformed with pMAL-c2 based constructs were grown to an optical density of 0.5 (λ 600) at 37°C in 125 ml LBA broth. The cultures were cooled to 25°C and fusion protein expression was induced for 3.5 h with 0.3 mM IPTG.

The induced cultures were centrifuged at 3000 x g for 10 min. at 4°C and the resulting pellets were frozen at -20°C for 14-18 hr. Each pellet was resuspended in 4 ml ice cold TNE (20 mM Tris-Cl, pH 7.4, 200 mM NaCl, 1 mM EDTA, pH 7.4) containing 1 mM PMSF, 5 ug/ml aprotinin, 5 ug/ml leupeptin, 2 ug/ml pepstatin A and sonicated (6 x 15 sec.) at 100 W on ice. After centrifugation at 10000 x g for 20 min. at 4°C, the supernatant was brought to 9 ml with ice cold TNE then mixed with 1 ml of amylose resin (NEB) for 45 minutes at 4°C. The resin was washed 5 times with 10 ml ice cold TNE. To elute the fusion

protein, the resin was resuspended in 0.5 ml ice cold TNE containing 10 mM maltose and mixed at 4°C for 1.5 hr. After a 5 min. centrifugation at 3000 x g, the eluate containing the fusion protein (~400 µl) was removed from the resin by aspiration with a pipette and stored at 4°C.

2.11 Factor Xa digestion of purified pMAL-c2 derived fusion protein

To hydrolyze the peptide bond joining the recombinant protein to the MBP, fusion protein was incubated with 2% (w/w) factor Xa protease (NEB) for 18-24 hr in TNE containing 2 mM CaCl₂. The fusion protein and cleavage products were visualized using SDS-PAGE (section 2.13).

2.12 Preparation of *S. plicatus* cytosolic protein fraction

To obtain a crude *S. plicatus* cytosolic fraction for immunoblotting experiments, the cells were collected by centrifugation at 5000 x g for 10 min, then resuspended in 4 ml TNE containing 1 mM PMSF, 5 µg/ml aprotinin, 5 µg/ml leupeptin and 2 µg/ml pepstatin A. The suspension was homogenized with a Dounce homogenizer for 1 min. at 4°C and sonicated (6 x 15 sec.) at 100 Watts on ice using a Braun-sonic 1510 (B. Braun Melsungen AG). The lysate was centrifuged at 20000 x g for 20 min and the supernatant, considered to be the crude cytosolic fraction, was stored at -20°C.

2.13 SDS-PAGE

SDS-polyacrylamide gel electrophoresis was performed as described by Laemmli (82) using a Bio-Rad Mini-Protean™ II cell apparatus. Proteins were resolved using a 5% (w/v) polyacrylamide stacking gel and 7.5-10% (w/v) polyacrylamide separating gels. Gels were separated by electrophoresis at 100 V until the bromophenol blue was within 1 cm of the

bottom of the gel. For all SDS-PAGE experiments, broad range prestained protein molecular weight markers (NEB) were separated by electrophoresis in parallel with the test samples. If the resolved proteins were not to be transferred to nitrocellulose as described below, they were visualized by commassie blue R250 staining.

2.14 Western blot analysis

2.14.0 Generation of Antibodies Against SpHex-MBP Fusion Protein

Rabbit-anti-*SpHex*-MBP (anti-*SpHex*-MBP) fusion protein antibodies were raised at the National Biological Laboratories (Oakbank, MB). Briefly, purified *SpHex*-MBP fusion protein emulsified with Freund's complete adjuvant was injected subcutaneously at multiple sites. Two weeks later, the antigen was emulsified with Freund's incomplete adjuvant and injected subcutaneously at multiple sites. This step was repeated 3 weeks later. Serum was collected one week after the final injection.

2.14.1 Transfer of proteins resolved by SDS-PAGE onto nitrocellulose membrane

Transfer of SDS-PAGE resolved proteins onto nitrocellulose was performed according to Towbin *et al.* (83). Gels containing resolved proteins were transferred to nitrocellulose within one hour using a current of 200 mA at 4°C in 25 mM Tris-Cl, pH 8.3, 193 mM glycine and 20% (v/v) methanol. The nitrocellulose containing the resolved proteins was air dried and stored at 4°C.

2.14.2 Immunodetection of proteins bound to nitrocellulose

Empty protein binding sites on nitrocellulose blots containing SDS-PAGE resolved proteins were blocked by immersing the blot in TBST (10 mM Tris-Cl, pH 7.5, 100 mM NaCl, 0.1% Tween 20 (w/v)) containing 5% milk powder for 1 hr with shaking. Next, an

appropriately diluted primary antibody was added and allowed to incubate with the blot for an additional 1-2 hr with shaking. The blot was washed with TBST and incubated with appropriately diluted donkey-anti-rabbit-IgG-horseradish peroxidase conjugate (Amersham) for 1 h in TBST containing 5% milk powder. After washing with TBST, the antigen-antibody complex was detected by incubating the blot in enhanced chemiluminescence (ECL) reagent (Amersham) for 1 min, wrapped in plastic wrap, and exposed to Kodak X-Omat AR X-ray film for 1-5 min depending of the strength of the signal.

2.15 Analysis of β -N-acetylhexosaminidase activity

2.15.0 In-plate assay of β -N-acetylhexosaminidase activity using 4-MUG

The in-plate assay for detecting β -N-acetylhexosaminidase in *E. coli* colonies plated on LBA agar was adapted from the method of Robbins *et al.* (66). Briefly, *E. coli* colonies were transferred to nitrocellulose and placed colony side up onto chromatography paper soaked in 3 mM 4-MUG (Toronto Research Chemicals), 10 mM Na₂HPO₄, 6 mM citric acid and 0.3% BSA, pH 3.0 for 10 min at 37°C. The nitrocellulose disk was then placed on chromatography paper soaked in 0.1 M glycine-carbonate buffer, pH 10 for 10 min. Colonies expressing β -N-acetylhexosaminidase fluoresced under 366 nm UV light.

2.15.1 β -N-acetylhexosaminidase activity assay and pH profile analysis

Protein concentrations of samples containing β -N-acetylhexosaminidase were determined as described in section 2.3.2. The pH profile was determined by incubating test samples with 4-MUG as a substrate in 10 mM Na₂HPO₄, 6 mM citric acid, 0.3% BSA ranging from pH 2.0 - 9.0. After a 45 min. incubation at 25°C, each reaction was quenched with 0.1 M glycine-carbonate buffer, pH 10. The fluorescent product, 4-MU, was measured

(excitation 364 nm, emission 448 nm) with a Hitachi F-2000 fluorescence spectrophotometer using a standard curve ranging from 0-400 nM 4-MU (Toronto Research Chemicals). Kinetics were obtained in an identical manner at the pH optimum of the wild type enzyme using 4-MUG concentrations ranging from 0.033 - 5 mM. *SpHex* activity towards 4-MUGS was determined but no kinetic analysis was performed. The 4-MUGS concentration used was 3.5. mM. Specific activity (S. A.) was determined using the following equation (for 1 ng protein):

$$\text{S. A.} = \frac{\text{4-MU formed (nM)}}{45 \text{ min}} \times 10^3 \times 60 \text{ min} = \mu\text{mole 4-MU/hr/mg}$$

K_M and V_{MAX} values were determined using the direct linear plot method (84).

3 Results

3.0 Restriction enzyme mapping of Clone 36

Clone 36, encoding the *SpHex* gene, was previously identified by Robbins *et al.* (65) by screening a *Streptomyces plicatus* expression library using 4-methylumbelliferyl (4-MU) glycosides of *N*-acetylglucosamine oligosaccharides as a substrate for an in-plate assay (section 2.13.0). Clone 36 was found to have a 5.0 kb insert and expressed a 60 kDa fusion protein (65). Based on the mass of the expressed protein, the coding region for *SpHex* was predicted to be within a 1.8 kb region of the insert, adjacent to the β -galactosidase gene of the vector which is also under control of the *E. coli* promoter (coded further upstream in the vector). The entire 5.0 kb insert was restriction mapped and subcloned into pBluescript (pBS⁺) (Stratagene) as 7 smaller fragments (Fig. 4).

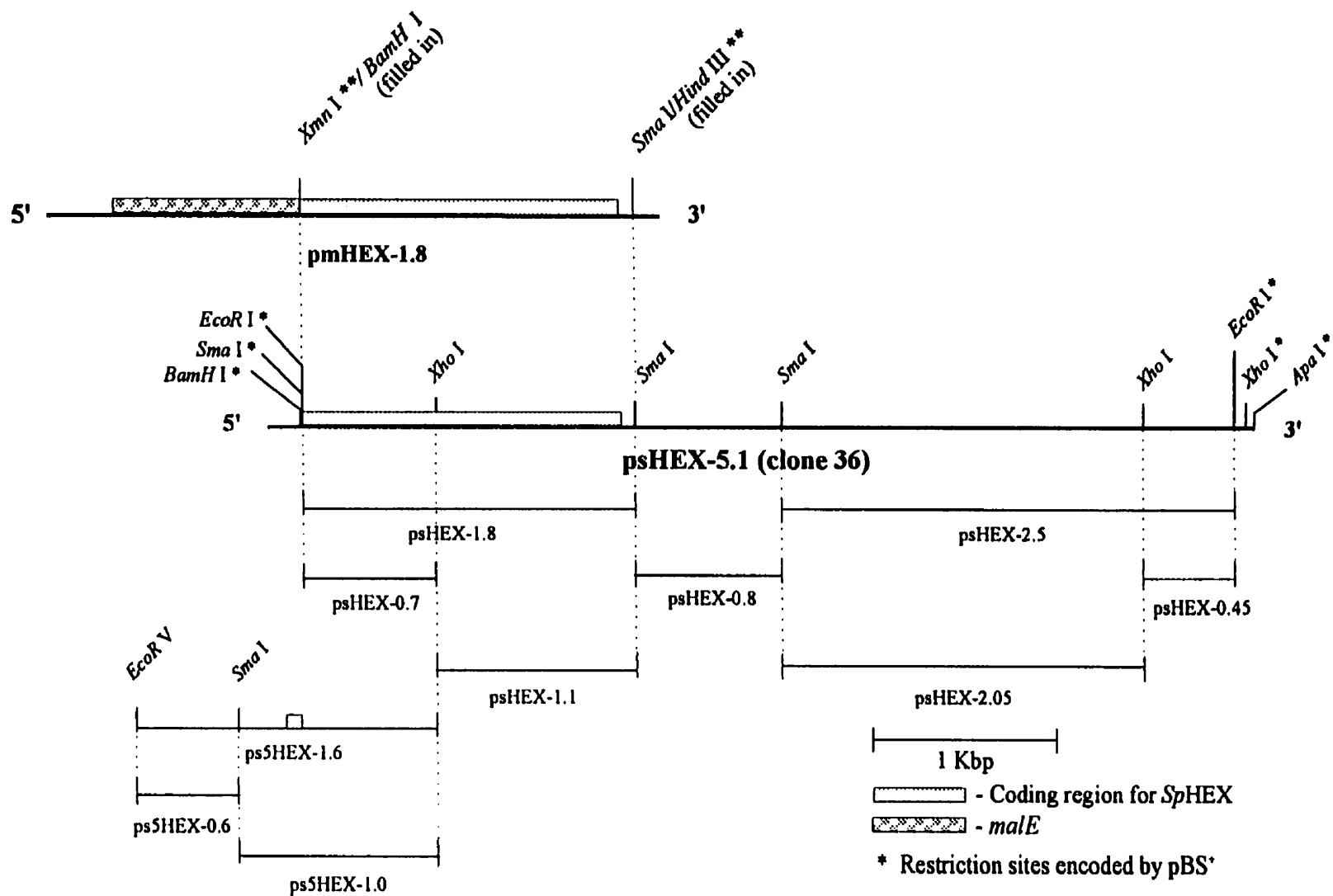
3.1 Sequencing the *SpHex* coding region from Clone 36

The first 1.8 kb of the 5'-end of Clone 36, believed to contain the coding region for *SpHex* (bounded by *EcoR* I (5') and *Sma* I (3') sites), was subcloned into pBS⁺ as 2 smaller fragments (Fig 4). The subclones were sequenced at the National Centers of Excellence core sequencing facility (Toronto, ON) by the dideoxy chain termination method as described in section 2.7. Sequencing was done on double-stranded DNA and covered both strands. To clarify the sequence in some areas, sequence-specific primers were used to sequence single-stranded DNA produced from the 1.8 kb *EcoR* I/*Sma* I clone (psHEX-1.8).

3.2 Cloning and sequencing the 5'-end of the *SpHex* gene

The predicted amino acid sequence of the psHEX-1.8 insert was found to contain 2 large open reading frames, one of which (frame +2) showed homology to other family 20

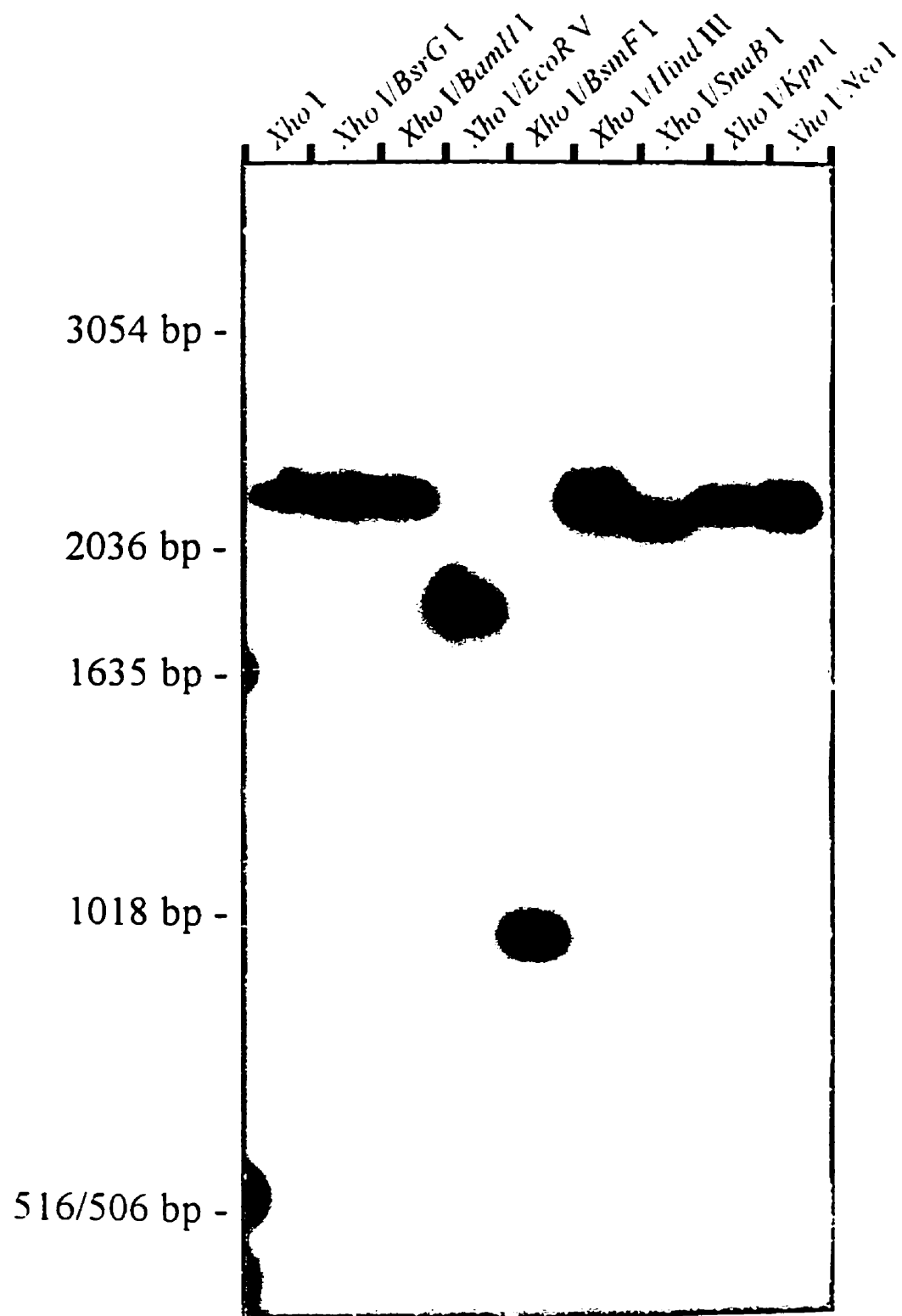
Figure 4 Restriction enzyme map and subclones of clone 36. Clone 36 was restriction enzyme mapped and subcloned into pBS⁺ as 7 smaller fragments to create the psHEX-[kb] constructs. A 1.6 kb (*EcoR* V/*Xho* I) fragment, containing the 5'-end of the *SpHex* gene, was cloned into pBS⁺ as a single fragment to create ps5HEX-1.6. The ps5HEX-1.6 construct was subsequently subcloned into pBS⁺ as two smaller fragments to create ps5HEX-1.0 and ps5HEX-0.7. In order to produce *SpHex*-MBP fusion protein, 1.8 kb of the 5'-end of clone 36 (*SpHex* coding region) was subcloned into pMAL-c2 (in-frame with *malE*) to create pmHEX-1.8.



β -N-acetylhexosaminidases in a ClustalW1.7 multiple sequence alignment. However, this open reading frame did not include an initiating Met codon, indicating that a portion of the 5'-end of the *SpHex* gene was missing from clone 36. To obtain the 5'-end of the *SpHex* gene, genomic DNA was isolated from *S. plicatus* as described in section 2.2.3. The genomic DNA was digested with combinations of *Xho* I and various restriction enzymes, and fragments that hybridized to a 600 bp *Asc* I/*Xho* I ³²P-labeled probe (from the 5'-end of clone psHEX-1.8) were identified by Southern blotting (Fig. 5) as described in section 2.5. There was only one *Xho* I site in the known *SpHex* gene sequence. Thus, using *Xho* I to cut out the 5'-end of the probe ensured that any genomic DNA fragment that had been cut with *Xho* I in combination with another restriction enzyme, that did not cut within the first 600 base pairs of the 5'-end of the known *SpHex* gene, would contain novel 5'-end sequence plus that which hybridized to the probe. Hence, a 1.7 kb *EcoR* V/*Xho* I fragment hybridized with the probe and was predicted to contain approximately 1.1 kb of sequence beyond the known 5' end of the gene. An area of an agarose gel containing this fragment was excised, purified using a GeneClean II kit (section 2.4.3), ligated into pBS⁻ (*EcoR* V/*Xho* I), and electroporated into *E. coli* (section 2.1.3).

The resulting colonies were screened by DNA/DNA hybridization (section 2.5.2) using the same DNA probe as described above. Ten positive colonies were identified, one of which (ps5HEX-1.7) was used to determine the sequence of the 5'-end of the *SpHex* gene (Fig. 1). A 700 bp fragment was removed from the 5'-end of the ps5HEX-1.7 insert by digestion with *Sma* I and *EcoR* V. The larger vector (4.2 kb), which now contained only 1.0 kb of DNA directly adjacent to the 5'-end of the known *SpHex* gene sequence, was purified

Figure 5 Southern blot autoradiograph of *S. plicatus* genomic DNA probed with a ³²P-labeled probe complimentary to the 5'-end of clone 36. In order to isolate a fragment of *S. plicatus* genomic DNA containing the 5'-end of the gene encoding SpHex, a Southern blot of *S. plicatus* genomic DNA (digested with *Xho* I and various other restriction enzymes and separated through a 0.7% agarose gel) was probed with a 600 bp (*Asc* I (5')/*Xho* I (3')) probe complimentary to the 5'-end of clone 36.



from the 700 bp insert using a GeneClean II kit and religated to create ps5HEX-1.0. Upon sequencing 284 bp beyond the 5'-end of the known *SpHex* gene sequence coded in the psHEX-1.0 insert (double stranded DNA, both strands) (section 2.7), the complete coding sequence (1683 bp; 561 aa, starting at +269 of Genbank accession number AF063001) of the *SpHex* gene could be determined (Fig 6).

3.3 Comparison of *SpHex* to EMBL/SWISS-PROT database

Comparison of the deduced amino acid sequence of *SpHex* with the SWISS-PROT database using MaxHom identified 14 β -*N*-acetylhexosaminidases and 1 chitobiase, all family 20 enzymes (Table I). The *SmChb* (reference structure) amino acid sequence, which has not been deposited in the SWISS-PROT database, was retrieved from the Brookhaven protein database (PDB) and compared to the *SpHex* by pairwise alignment using MaxHom.

3.4 Comparative Molecular Modeling of *SpHex*

Atomic coordinates for the X-ray structure of *Serratia marcescens* chitobiase (*SmChb*) were obtained from the Brookhaven National Laboratory protein database (PDB file: 1qbb) and used as the reference structure for molecular modeling. A pairwise alignment between *SmChb* and *SpHex* amino acid sequences using MaxHom revealed a sequence identity of 30%. When the alignment was restricted to *SmChb* domain 3, an α/β barrel structure which contains the enzyme active site on the C-terminal end, the sequence identity was 45%. Domain 3 is the largest of four *SmChb* domains. Alignment of *SpHex* residues Pro149 - Pro498 with domain 3 encompassed 350 residues (63%) of the total amino acid sequence of *SpHex*.

Although the multiple alignment of all 15 family 20 enzymes, including *SpHex*,

Figure 6 Nucleotide and deduced amino acid sequence of *SpHex*. The constructs psHEX-1.8 and ps5HEX-1.0 were sequenced as described in section 2.7. The predicted amino acid sequence used for comparative molecular modeling is underlined. Residues subjected to mutational analysis (Arg 162, Glu 314, Asp 246) are double underlined.

COOGGGTGTCAGCOGCGAGGOGTGOOGTGAAGACCAOGTTTOOGTOCT	-228
TCTGGATGATTTTCOGOCATGOGGCAATTGTTTCATOGCTCCOOGTGGGOGGTGATGGTGOGGGAGAGTGGTTOOG	-169
TGCOOGTCTCTCAAGCOCCAGCOGACCOCTOGATTGCACTCATOGCGAAGGGAGCOCCAGTGAACCTCATC	-85
GACGGCACACAGAAOGACTCCOOGCATCTCCOOGCTGCTGGGATOGCTGCTGCTGGTGGOGGOGGTGGGOGC	-1
ATG ACC ACC GGC GCC GCC OCG GAC OGG AAG GOG OOC GTC OGG OCG ACG OOC CTC	54
Met Thr Thr Gly Ala Ala Pro Asp Arg Lys Ala Pro Val Arg Pro Thr Pro Leu	18
GAC OGG GTG ATC OCG GCC OOC GOC TOG GTC GAC OOC GGC GGC GOG OCG TAC OGC	108
Asp Arg Val Ile Pro Ala Pro Ala Ser Val Asp Pro Gly Gly Ala Pro Tyr Arg	36
ATC ACG OGC GGC ACC CAC ATC OCG GTC GAC GAC TOG OCG GAG GCC OCG OCG GTC	162
Ile Thr Arg Gly Thr His Ile Arg Val Asp Asp Ser Arg Glu Ala Arg Arg Val	54
GGG CAC TAC CTC GOG GAC CTG CTG OGG OOC GCC ACC GGC TAC OGG CTC OOC GTC	216
Gly His Tyr Leu Ala Asp Leu Leu Arg Pro Ala Thr Gly Tyr Arg Leu Pro Val	72
ACC GCC CAC GGC CAC GGC GGC ATA OGA CTG OGC CTG GOG GGC GGC OCG TAC GGC	270
Thr Ala His Gly His Gly Gly Ile Arg Leu Arg Leu Ala Gly Gly Pro Tyr Gly	90
GAC GAG GGC TAC OGC CTC GAC AGC GGC OOC GCA GGC GTC ACC ATC ACC GCC OGC	324
Asp Glu Gly Tyr Arg Leu Asp Ser Gly Pro Ala Gly Val Thr Ile Thr Ala Arg	108
AAG GCC GCC GGA CTC TTC CAC GGC GTC CAG ACC CTG OGT CAG CTG CTG OOC OOC	378
Lys Ala Ala Gly Leu Phe His Gly Val Gln Thr Leu Arg Gln Leu Leu Pro Pro	126
GCC GTC GAG AAG GAC TOC GOG CAG OOC GGC OOC TGG CTG GTC GOG GGC GGC ACC	432
Ala Val Glu Lys Asp Ser Ala Gln Pro Gly Pro Trp Leu Val Ala Gly Gly Thr	144
ATA GAG GAC ACG OCG OCG TAC GCC TGG OGC AGC GOG ATG CTC GAC GTC TOC OCG	486
Ile Glu Asp Thr <u>Pro Arg Tyr Ala Trp Arg Ser Ala Met Leu Asp Val Ser Arg</u>	162
CAC TTC TTC GGC GTC GAC GAG GTC AAG OGC TAC ATC GAC OGC GTC GCC OGC TAC	540
<u>His Phe Phe Gly Val Asp Glu Val Lys Arg Tyr Ile Asp Arg Val Ala Arg Tyr</u>	180
AAG TAC AAC AAG CTC CAC CTG CAC CTC AGC GAC GAC CAG GGC TGG OGC ATC GCC	594
<u>Lys Tyr Asn Lys Leu His Leu His Leu Ser Asp Asp Gln Gly Trp Arg Ile Ala</u>	198
ATC GAC TOC TGG OCG OCG CTG GOG ACG TAC GGC GGC TOC ACC GAG GTC GGC GGC	648
<u>Ile Asp Ser Trp Pro Arg Leu Ala Thr Tyr Gly Gly Ser Thr Glu Val Gly Gly</u>	216
GGC OOC GGC GGC TAC TAC ACC AAG GOG GAG TAC AAG GAG ATC GTG OGT TAC GOG	702
<u>Gly Pro Gly Gly Tyr Tyr Thr Lys Ala Glu Tyr Lys Glu Ile Val Arg Tyr Ala</u>	234
GOC TOC OGC CAC CTC GAG GTG GTC OOC GAG ATC GAC ATG OCG GGC CAC ACC AAC	756
<u>Ala Ser Arg His Leu Glu Val Val Pro Glu Ile <u>Asp</u> Met Pro Gly His Thr Asn</u>	252
GCC GCC CTC GOC TOC TAC GOC GAG CTG AAC TGC GAC GGT GTG GOG OOC OCG CTG	810
<u>Ala Ala Leu Ala Ser Tyr Ala Glu Leu Asn Cys Asp Gly Val Ala Pro Pro Leu</u>	270
TAC ACC GGC ACC AAG GTC GGC TTC AGC TOG CTG TGC GTG GAC AAG GAC GTC ACC	864
<u>Tyr Thr Gly Thr Lys Val Gly Phe Ser Ser Leu Cys Val Asp Lys Asp Val Thr</u>	288
TAC GAC TTC GTG GAC GAC GTC ATC GGC GAA CTC GOC GOG CTC ACC OOC GGC OGC	918
<u>Tyr Asp Phe Val Asp Asp Val Ile Gly Glu Leu Ala Ala Leu Thr Pro Gly Arg</u>	306

TAC CTC CAC ATC GGC GGC GAC GAG GOG CAC TOC AOG COC AAG GOC GAC TTC GTG	972
<u>Tyr Leu His Ile Gly Gly Asp Glu Ala His Ser Thr Pro Lys Ala Asp Phe Val</u>	324
GCG TTC ATG AAG OGG GTG CAG COG ATC GTC GOC AAG TAC GGC AAG ACG GTC GTC	1026
<u>Ala Phe Met Lys Arg Val Gln Pro Ile Val Ala Lys Tyr Gly Lys Thr Val Val</u>	342
GGC TGG CAC CAG CTG GOC GGC GOG GAA COG GTC GAG GGC GOG CTC GTC CAG TAC	1080
<u>Gly Trp His Gln Leu Ala Gly Ala Glu Pro Val Glu Gly Ala Leu Val Gln Tyr</u>	360
TGG GGC CTG GAC OGC ACC GGC GAC GOC GAG AAG GOC GAG GTC GOC GAG GOC GOC	1134
<u>Trp Gly Leu Asp Arg Thr Gly Asp Ala Glu Lys Ala Glu Val Ala Glu Ala Ala</u>	378
AGG AAC GGC AOG GGG CTG ATC CTC TOC COG GOC GAC OGG AOG TAC CTG GAC ATG	1188
<u>Arg Asn Gly Thr Gly Leu Ile Leu Ser Pro Ala Asp Arg Thr Tyr Leu Asp Met</u>	396
AAG TAC ACC AAG GAC ACC COG CTG GGC CTG TOC TGG GOG GGC TAC GTC GAG GTG	1242
<u>Lys Tyr Thr Lys Asp Thr Pro Leu Gly Leu Ser Trp Ala Gly Tyr Val Glu Val</u>	414
CAG OGG TOC TAC GAC TGG GAC COG GOC GGC TAC CTG COG GGC GOC COG GOC GAC	1296
<u>Gln Arg Ser Tyr Asp Trp Asp Pro Ala Gly Tyr Leu Pro Gly Ala Pro Ala Asp</u>	432
GOG GTC OGC GGC GTC GAG GOC COG CTG TGG AOG GAG ACC CTG TOC GAC COG GAC	1350
<u>Ala Val Arg Gly Val Glu Ala Pro Leu Trp Thr Glu Thr Leu Ser Asp Pro Asp</u>	450
CAG CTG GAC TAC ATG GOC TTC COG AGG CTG COG GGC GTC GOG AAC TGG GCT GGT	1404
<u>Gln Leu Asp Tyr Met Ala Phe Pro Arg Leu Pro Gly Val Ala Asn Trp Ala Gly</u>	468
COG OGG OGT CCA OGC AOG ACT GGG ACA OCT ACA AGG TGC GGC TOG COG GCA GGC	1458
<u>Pro Arg Arg Pro Arg Thr Thr Gly Thr Pro Thr Arg Cys Gly Ser Pro Ala Gly</u>	486
COG GTA CTG GGA GGC GGC GGG GAT OGA CTT CTA COG CTC GOC GCA GGT GOC CTG	1512
<u>Pro Val Leu Gly Gly Gly Gly Asp Arg Leu Leu Pro Leu Ala Ala Gly Ala Leu</u>	504
GAC CTG ACC OGG AOG GCA OGA OGG GCA COC OGG AGG ACC GTC TOC GGG GTG COC	1566
<u>Asp Leu Thr Arg Thr Ala Arg Arg Ala Pro Arg Arg Thr Val Ser Gly Val Pro</u>	522
GCC OGT OGG TGG GTC AGA GOC COG OGA TGG GGT TCT TCA GGG TOC OGA TCA GCT	1620
<u>Ala Arg Arg Trp Val Arg Ala Pro Arg Trp Gly Ser Ser Gly Ser Arg Ser Ala</u>	540
GGA GCG CTC OGG AOG GGT OGG ACA GGT CCA CCA TCT GOC GGT TGT OGC GCA GCT	1674
<u>Gly Ala Leu Arg Thr Gly Arg Thr Gly Pro Pro Ser Ala Gly Cys Arg Ala Ala</u>	558
GGA GCC GGT TGAGGCAGGACAGOGOGAACTOGGGGOGGAACATGTGTACOGCTOGAACTTCTOGGCCAGC	1746
<u>Gly Ala Gly</u>	561
TCGGCACOGACTCCTGGTACTOOC	1771

Table I

Glycosyl hydrolases identified as having sequence similarity to S. plicatus β -N-acetylhexosaminidase†

Enzyme	Species	SWISS-PROT/PD accession code	% Sequence Identity	X-ray structure available	Ref.
Hex	<i>Porphyromonas gingivalis</i>	P49008*	31	N/A	85
Chitinase	<i>Serratia marcescens</i>	1QBB (PDB)*	30	Yes (1.9Å)	55
Chitinase	<i>Vibrio harveyi</i>	P13670	30	N/A	86
Hex	<i>Alteromonas sp.</i>	P49007	30	N/A	87
Hex	<i>Vibrio vulnificus</i>	Q04786	27	N/A	88
Hex	<i>Candida albicans</i> (yeast)	P43077*	27	N/A	89
Hex B, β -subunit	Cat	P49614*	26	N/A	90
Hex A, α -subunit	Mouse	P29416*	26	N/A	91
Hex B, β -subunit	Mouse	P20060*	25	N/A	92

Table I (continued)

Enzyme	Species	SWISS-PROT/PD	% Sequence	X-ray structure	Ref.
		accession code	Identity	available	
Hex A, α -subunit	Human	P06865/1QBC*	25	Homology model †	93
Hex B, β -subunit	Human	P07686/1QBD*	25	Homology model †	94
Hex	<i>Dictyostelium discoideum</i>	P13723*	24	N/A	95
Hex	<i>Entamoeba histolytica</i>	P49009*	24	N/A	96
Hex	<i>Bombyx mori</i> (silk moth)	P49010*	24	N/A	97

‡ All enzymes, except *S. marcescens* chitobiase, were identified in the SWISS-PROT database (release 34) using MaxHom (expert mode, n=25).

* Sequences chosen for alignment by ClustalW1.7 for use in comparative molecular modeling.

† Human HexA & B have been modeled by comparative molecular modeling and the coordinates deposited in the PDB (55).

revealed high conservation for specific residues predicted to be involved in catalysis, *SmChb* *Vibrio harveyi* chitobiase, *Alteromonas sp.* β -N-acetylhexosaminidase and *Vibrio vulnificus* β -N-acetylhexosaminidase contained many large insertions not present in the other family 20 members. Using the X-ray structure of *SmChb*, the insertions in this 4-member sub-family were found localized to loop regions that were absent in the other family members, including human Hex A and *SpHex*. The multiple alignment of the enzymes listed in Table I is shown in figure 7 and excludes the above mentioned sub-family of sequences - with the exception of the *SmChb*. This multiple sequence alignment indicated that *SpHex* residues Arg162 (human Hex A; α -Arg178, *SmChb*; Arg349), Glu314 (HexA; α -Glu323, *SmChb*; Glu540) and Asp246 (human HexA; α -Asp258, *SmChb*; Asp448) are conserved in all the species with the exception of *Porphyromonas gingivalis*. In this species, the functionally equivalent residue Glu, replaces Asp 246. Based on the *SmChb* structure, Arg162 and Glu314 of *SpHex* were predicted to be the substrate binding residue and catalytic acid residue respectively.

The pairwise alignment between *SmChb* and *SpHex*, shown in figure 8, was taken directly from the multiple sequence alignment shown in figure 7 and encompasses the sequence of *SmChb* that is responsible for composing the β/α -barrel of domain III. The pairwise alignment between *SpHex* and *SmChb* indicated a total of 9 deletions in the *SpHex* sequence and none in the *SmChb* sequence (fig 8). Deletions in the *SpHex* sequence localized primarily to surface loop structures of *SmChb*; however, some of the deletions partially overlapped onto secondary structural elements composing the α/β barrel. Thus, to improve the quality of the alignment shown in figure 8, deletions in the *SpHex* sequence that overlapped onto the secondary structural elements of the α/β -barrel were manually

Figure 7 ClustalW1.7 multiple sequence alignment of family 20 glycosyl hydrolases.

In order to obtain the best sequence alignment between *SpHex* (indicated as "Sequence 3") and *SmChb* (indicated as "Serratia"), these two sequences were multiply aligned with the family 20 enzymes indicated. Residues altered by site directed mutagenesis (*SpHex* Arg 162, Glu 314, Asp 246) are indicated by arrows.

Page 1.1

	1	15	30	45	60	75	90
Hex A Human α-subunit	---	---	---	MTSSRLWFSLLL	AAAFAG	---	RAQ
Hex A Mouse α-subunit	---	---	---	MAGCRLWVSLLL	AAALAC	---	LAT
Hex B Human β-subunit	MELCGLGLPRPPHLL	ALLLATLLAAMLALL	TQVALVQVAEAAARA	PSVSAK	---	---	PGP
Hex B Cat β-subunit	---	---	---	---	---	---	---
Hex B Mouse β-subunit	---	---	MPQSPRSAP	GLLLQLQALVSLVSLA	LVAPAR	---	LQP
Hex D. discoideum	---	---	---	MIKKIILFFAVLI	AIVIGQ	---	QPL
Hex E. histolytica	---	---	---	---	---	---	MSK
Hex C. albicans	---	---	M	VLDKMIIFHLLMLC	NVVHA	---	AKV
Hex B. mori	---	---	MW	LQAICIYTVFIIIGC	GIPTAA	---	BEH
Hex P. gingivalis	---	---	MK	RLTFGACICCLLSLM	ACSOKA	---	KQV
Hex S. plicatus	---	---	---	NSAPD	---	---	RKA
Chb S. marcescens	MNAFKLSALARLTAT	MGFLGCGMSAMADQQ	LYDQLSQLKLVKML	DNRAGENGVDCAALG	ADMASCRVLFTLSN	DQQAIDGKQWVIYPH	90

Page 2.1

	91	105	120	135	150	165	180
Hex A Human α-subunit	---	---	---	---	WPQN	FQTSQQR	---
Hex A Mouse α-subunit	---	---	---	---	WPQY	IQTYHRR	---
Hex B Human β-subunit	---	---	---	---	LPLS	VKMTPNL	---
Hex B Cat β-subunit	---	---	---	---	MPLS	VKTSPL	---
Hex B Mouse β-subunit	---	---	---	---	PPRS	VQMPPL	---
Hex D. discoideum	---	---	---	---	YFQQ	VSIGTCV	---
Hex E. histolytica	---	---	---	---	SILS	IEVTMS	---
Hex C. albicans	---	---	---	---	APQS	VTWENDT	---
Hex B. mori	---	---	---	---	WTCE	NNRCKIRNEPENKE	PVLSLEACKMFCDY
Hex P. gingivalis	---	---	---	---	YDKG	INIILP	---
Hex S. plicatus	---	---	---	---	TPLD	RVIPAP	---
Chb S. marcescens	SPRQTLRVNDNQFKI	AHLTGDLKLEPTAK	PSGFPAGKAVEIPV	AEYWQLFRNDPLFRW	YATSGDAKPKMLANT	DTENLDQFVAPFTGD	180

Page 3.1		181	195	196	210	211	225	226	240	241	255	256	270	
Hex A	Human α -subunit	-----YVLYP	NNFQ	PQYDVSSAAQP	GCSV-----			-LDEAPQRYRDL	LFG	SGSWPRPYLTQKRH-	-----	TLEKNVLVVS		98
Hex A	Mouse α -subunit	-----YTLYP	NNFQ	FRYHVSSAAQG	GCVV-----			-LDEAFRRYRN	LFG	SGSWPRPSFSNKQQ-	-----	TLGKNILVVS		98
Hex B	Human β -subunit	-----LHLAP	ENFY	ISHSPNSTAGP	SCTL-----			-LEEAFRRYHGY	IFG	FYKWHHEPAEPQAK-	-----	TQVQQLLVSI		131
Hex B	Cat β -subunit	-----LHLSR	DNFS	IGYGPSSTAGP	TCSL-----			-LQEAFFRYHEY	IFG	FDKRQRRPAKPNSA-	-----	IELQQLLVTV		73
Hex B	Mouse β -subunit	-----LYISA	EDFS	IDHSPNSTAGP	SCSL-----			-LQEAFFRRYNY	VFG	FYKRHHGPARFRAE-	-----	PQLQKLLVSI		110
Hex	<i>D. discoideum</i>	-----IPVAP	GSIL	IESNIESAT--	-----			-FSVSM DRYTN	LFFP	FSN-ESEPSSN----	-----	ESFLLSV		86
Hex	<i>E. histolytica</i>	-----YCINF	LKSN	FNHTISFPLQ-	-----			-QQTGLQDFK	VSLFK	EIDLPRITPSVS---	-----	SVITDVVEL		72
Hex	<i>C. albicans</i>	-----ATIIIN	PRLQ	ANTSCPLLED	AFVR-----			-TVSAIEKLK	WHFFP	IDDFNTANGKN----	-----	IKTSLVHI		97
Hex	<i>B. mori</i>	GLLWPK--PTIETNL	GNFL	SKINMNTIDIQ	ITKQ-----			GKSDDLLTAA	ADRFPK	TLVSSSVPKGFSAK-	-----	AAGKSVTVYL		135
Hex	<i>P. gingivalis</i>	-----MQLTE	SDDS	FEVDDKTTI--	-----			--CVSAEELK	PIAKL	LADKLRASADLSLQ-	-----	IEIGEEPSGN		96
Hex	<i>S. plicatus</i>	-----ASVDP	GGAP	YRITRGTTHIR-	-----			--VDSREARR	VGHY	LADLLRPATGYR---	-----	LPVTAHGH		74
Chb	<i>S. marcescens</i>	QWKRTKDDKNILMTP	ASRF	VSNADLQTLPA	GALRGKIVPTPMQVK			VHAQDADLRK	QVALD	LSTLVKPAADVVSQR		FALLGVPVQTN	GYPI	270

Page 4.1		271	285	286	300	301	315	316	330	331	345	346	360	
Hex A	Human α -subunit	VVTPGC-NQLPTLES	VENY	TLTINDDQ---	--CLLLSETV	WGALR	GLETFSQLVW	KSABG	-----	TFFINKTEI	EDFPRFP	PHRGLLLD	176	
Hex A	Mouse α -subunit	VVTAEC-NEFPNLES	VENY	TLTINDDQ---	--CLLASETV	WGALR	GLETFSQLVW	KSABG	-----	TFFINKTKI	KDFPRFP	PHRGVLLD	176	
Hex B	Human β -subunit	TLQSEC-DAFPNISS	DESY	TLLVKEPV---	--AVLKANRV	WGALR	GLETFSQLVY	QDSYG	-----	TFTINESTI	IDSPRFP	SHRGILID	209	
Hex B	Cat β -subunit	VLDSEC-DLFPNITS	DESY	TLLVKEPV---	--AFLKANRV	WGVL	GLETFSQLIY	QDSYG	-----	TFTVNESDI	IDSPRFP	PHRGILID	151	
Hex B	Mouse β -subunit	TLESEC-ESFPSSLSS	DETY	SLLVQEPV---	--AVLKANSV	WGALR	GLETFSQLVY	QDSFG	-----	TFTINESSI	ADSPRFP	PHRGILID	188	
Hex	<i>D. discoideum</i>	TIYSD--DETLLQIGI	DESY	SLSIEQGS---	--YQLKATNI	YGAMR	GLETFKQLIVY	NELE	N-----	SYSIVCVSI	SDSPRYP	PWRGFMVDS	164	
Hex	<i>E. histolytica</i>	SSSNP--MPKLQIGF	DESY	ILEVTTNS---	--ISIKAVTV	YGARH	AFETLLQLIRI	SSNK	F-----	VISQLPIKI	SDAPRFP	KWRGLMVD	150	
Hex	<i>C. albicans</i>	QVDDA--TVDLQLGV	NESY	TLKINTDG---	--INIHAATT	WGALH	GLVSLQQLI	IHTSED	K-----	YVVPLSVTI	SDFPNFK	HRLMIDS	175	
Hex	<i>B. mori</i>	VNENPY-IREFSLDM	DESY	EYISSTSSDK	VNATIRGNS	FFGVNR	GLETLSQLIVY	DDIR	NN----	LLIVRDVTI	KDRFPY	PYRGILLD	220	
Hex	<i>P. gingivalis</i>	AIYIGV-DTALPLK-	EEGY	MRLSDKRG---	--VSIIGKSA	HAFY	GMQTLQLL	PAEVES	SNEVLLPMT	VPVGEI	KDEPAPG	YRGFMLDV	179	
Hex	<i>S. plicatus</i>	GGIRLR-LAGGPYG-	DEGY	RLDSPGAG---	--VTITARKA	AGLFH	GVQTLRQL	LPFAVEK	DSAQFGP	WLVAGGTI	EDTPRYA	WRSAMLDV	157	
Chb	<i>S. marcescens</i>	KTDIQPGKPKGAMAV	SGAY	ELKIGKKE---	--AQVIGFDQ	AGVFPY	GLQSILSL	VPDSDGSG	-----	KIATLDA	SDAPRFP	YRGIFLDV	347	

Page 5.1

		361	375	376	390	391	405	406	420	421	435	436	450	
Hex A	Human α -subunit	SRHYLPSSILD ~~T~~ LD	VMAYNKLNVFHH ~~L~~ V	DD ~~S~~ FPYESTFPPEL	MRKGS-----	-----YNP-----	V	THIYTAQDVKEVIEY	245					
Hex A	Mouse α -subunit	SRHYLPSSILD ~~T~~ LD	VMAYNKFNVFHH ~~L~~ V	DDSSFPYESTFPPEL	TRKGS-----	-----FNP-----	V	THIYTAQDVKEVIEY	245					
Hex B	Human β -subunit	SRHYLPVKIILKTLD	AMAFNKFNVLHH ~~H~~ IV	DD ~~S~~ FPYQSITFPPEL	SNKGS-----	-----YS-----	L	SHVYTPNDVRM ~~V~~ IEY	277					
Hex B	Cat β -subunit	ARHFLPVKSILKTLD	AMAFNKFNVLHH ~~H~~ IV	DDQSPFYQSVTFPPEL	SNKGS-----	-----YS-----	L	SHVYTPNDVHTVIEY	219					
Hex B	Mouse β -subunit	SRHFLPVKTILKTLD	AMAFNKFNVLHH ~~H~~ IV	DDQSPFYQSTFPPEL	SNKGS-----	-----YS-----	L	SHVYTPNDVRM ~~V~~ LEY	256					
Hex	<i>D. discoideum</i>	ARHYIPKNMILHMID	SLGFSKFNTLHH ~~H~~ MV	DAVAFPVESTTYPDL	T-KGA-----	-----FS-----	P	SATFSHDDIQEVVAY	231					
Hex	<i>E. histolytica</i>	SRNPLSPLMPFKRIID	TLASVKANVLHIH ~~L~~ S	DAQTFVFESKKYPLL	HQKGM-----	-----YD-----	E	SPVLTQSFRLRLAQY	218					
Hex	<i>C. albicans</i>	GRNFLTVD ~~S~~ ILEQID	IMALSKMNSLHH ~~H~~ LA	DSQSWPV ~~A~~ LESYPHM	I-KDA-----	-----YS-----	N	DEVYSKNDLK ~~I~~ YIVDY	242					
Hex	<i>B. mori</i>	ARNFYSIDSIKRTID	AMAAVKLNTFHH ~~H~~ IT	DSQSFPLVLQKRPNL	SKLGA-----	-----YS-----	P	TKVYTKQDIREVVEY	288					
Hex	<i>P. gingivalis</i>	CRHFLSVEDIKKHID	IMAMPKINRFHH ~~L~~ T	EDQAWRIEIKKYPRL	TEVGS-----	-----TRTEGDGTQY		SGFYTQEQVRDIVQY	254					
Hex	<i>S. plicatus</i>	SRHFPGVDEVKRYID	RVARYKYNLHLH ~~L~~ S	DDQGWRI ~~A~~ IDSWPRL	ATYGG-----	-----STEVGGG--	P	GGYYTKAEYKEIVRY	230					
Chb	<i>S. marcescens</i>	ARNFHKKDAVLRLLD	QMAAYKLNK ~~F~~ HPHLS	DDEGWRIEIPGLPEL	TEVGGQRCHDLSETT	CLLPQYGGQGPD--	VY	GGFFSRQDYIDI ~~I~~ KY	435					



Page 6.1

		451	465	466	480	481	495	496	510	511	525	526	540	
Hex A	Human α -subunit	ARLRGIRVLA ~~E~~ FDTP	GHTLS-----W-	---	GPGIPG----	L	LTPCYSGSEPS----	-GTFGPVNPSLNNTY	EFMSTFFLEVSSVFP	313				
Hex A	Mouse α -subunit	ARLRGIRVLA ~~E~~ FDTP	GHTLS-----W-	---	GPGAPG----	L	LTPCYSGSHLS----	-GTFGPVNPSLNSTY	DFMSTLFLEISSVFP	313				
Hex B	Human β -subunit	ARLRGIRVLPE ~~F~~ DTP	GHTLS-----W-	---	GKGQKD----	L	LTPCYSRQNKL----	-DSFGPINPTLNNTY	SFLTTPFKRISEVFP	345				
Hex B	Cat β -subunit	ARLRGIRVIPE ~~F~~ DSP	GHTQS-----W-	---	GKGQKD----	L	LTPCYNEHKQS----	-GTFGPINPILNSTY	NFLSQFFKEVSMVFP	287				
Hex B	Mouse β -subunit	ARLRGIRVIPE ~~F~~ DTP	GHTQS-----W-	---	GKGQKN----	L	LTPCYNQKTKT----	-QVFGPVDPTVNTTY	AFPN ~~T~~ FFKEISSVFP	324				
Hex	<i>D. discoideum</i>	AKTYGIRVIPE ~~F~~ DIP	GHAAA-----W-	---	GIGYPE----	L	VATCPDYAAN----	-VNNIPLDISNPATF	TFIQNLFTETIAPLFI	298				
Hex	<i>E. histolytica</i>	GANRGVIVYGEIDTP	AHTAS-----W-	---	NLGYPG----	V	VANCWDYIVSTSMRY	GENVLSLNPANPNTF	PIIDALMKELSDTFG	291				
Hex	<i>C. albicans</i>	ARARGVRVIPEIDMP	GHRAG-----W-	---	KQVDP-----	T	IVECADAPWTDAAVE	-PPPGQLNIESEKTY	EVISNVYNELSDIFI	314				
Hex	<i>B. mori</i>	GLERGVRVLPE ~~F~~ DAP	AHVGE-----W-	---	QDTG-----		LTVCFKAEPWTKFCV	EPFCGQLNPTKEELY	DYLEDIYVEMAEAFE	359				
Hex	<i>P. gingivalis</i>	ASDHFITVIPMIEMP	GHAMA-----A-	---	LAAYP-----		QFRCFPREFKPRIIW	GVEQDVYCAGKDSVF	RFISDVIDEAPLFP	325				
Hex	<i>S. plicatus</i>	AASRHLEVVPIDMP	GHTNA-----	---	ALASY-----	A	BLNCDGVAPPLYTGT	KVGFSSLCVDKDVTY	DFVDDVIGELAA ~~L~~ TP	301				
Chb	<i>S. marcescens</i>	AQARQIEVIPEIDMP	AHARA ~~A~~ VVSMEARYK	KLHAAGKEQEANEFR	LVDPTDTSNTTSVQF	PNRQSYLNPCLDSSQ	RFV ~~D~~ DKVIGEIAQM ~~H~~ K	525						



Page 7.1

		541	555	556	570	571	585	586	600	601	615	616	630	
Hex A	Human α -subunit	-----DFYLHLGGDE	VD----	FTCWKSNP-	-----EI-Q-----		DFMRK-----	KG	FG-EDFKQLESFYIQ	TLLDIVSSYGKQY--				370
Hex A	Mouse α -subunit	-----DFYLHLGGDE	VD----	FTCWKSNP-	-----NI-Q-----		AFMKK-----	KG	F--TDFKQLESFYIQ	TLLDIVSDYDKQY--				369
Hex B	Human β -subunit	-----DQFIHLGGDE	VE----	FKCWESNP-	-----KI-Q-----		DFMRQ-----	KG	FG-TDFKKLESFYIQ	KVLDIATINKGS--				402
Hex B	Cat β -subunit	-----DHFVHLGGDE	VE----	FQCWESNP-	-----EI-Q-----		GFMKQ-----	KG	FG-KDFKKLESFYIQ	KLLGIVSTVKKGS--				344
Hex B	Mouse β -subunit	-----DQFIHLGGDE	VE----	FQCWASNP-	-----NI-Q-----		GFMKR-----	KG	FG-SDFRRLLESFYIK	KILEIISSLKKNS--				381
Hex	<i>D. discoideum</i>	-----DNYFHTGGDE	LV----	TGCWLEDP-	-----AI-A-----		NWMTK-----	MG	F--STTDAFQYFEN-	NLDVTMKSINRTK--				353
Hex	<i>E. histolytica</i>	-----TDYVHVGGDE	VW----	TSGWSKSKE	YS----DI-Q-----		KFMKS-----	KG	L--NSLTELEGYFNK	YAQEQVIHNGKHP--				350
Hex	<i>C. albicans</i>	-----DDVPFVGND	LQ----	EKCYSAQLL	PNN---TV-T-----		DLLKRYLKK--	ALPI	FN-KVNHRKLTMWDD	VLLSDVSADKIPSNI				383
Hex	<i>B. mori</i>	S----TDMPHMGDE	VS----	ERCWNSSE-	-----EI-Q-----		NFMIQN-----	RWN	LDKSSFLKLWNYFQK	NAQDRAYKAFGKR-L				421
Hex	<i>P. gingivalis</i>	-----GTYPHIGGDE	CP----	KDRWKACS-	-----LC-Q-----		KRMRD-----	NG	L--KDEHQLQSYFIK	QAEKVLQKHGKRL--				381
Hex	<i>S. plicatus</i>	-----GRYLHIGGDE	AH----	STPKADFV-			AFMKR-----		-----VQPIVAKY-	-----Q-----KTV--				338
Chb	<i>S. marcescens</i>	EAGQPIKTWHFGDE	AKNIRLGAGYTDKAK	PEPGKGIIDQSNEDK			PWAKSQVCQTMIKBG		K-VADMEHLPSYFGQ	EVSKLVKAHGIDR--				612



Page 8.1

		631	645	646	660	661	675	676	690	691	705	706	720	
Hex A	Human α -subunit	--VWQEVFDN----	-KVKIQPD	TIIQVWR	EDIP-VNYMKELELV		TKAGFRALLS---	AP	WYLN-----				-----RISYGPDWK	432
Hex A	Mouse α -subunit	--VWQEVFDN----	-KVKVRPD	TIIQVWR	REMP-VEYMLEMQDI		TRAGFRALLS---	AP	WYLN-----				-----RVKYGPDWK	431
Hex B	Human β -subunit	--IVWQEVFDD----	-KAKLAPGT	IVEVWK	DSA----YPEELSRV		TASGFPVILS---	AP	WYLD-----				-----LISYGQDWR	461
Hex B	Cat β -subunit	--IVWQEVFDD----	-HVKLLPGT	IVQVWK	NQV----YTEELREV		TAAGFPVILS---	AP	WYLD-----				-----WISYGQDWR	403
Hex B	Mouse β -subunit	--IVWQEVFDD----	-KVELQPGT	VVEVWK	SEH----YSYELKQV		TGSGFPAILS---	AP	WYLD-----				-----LISYGQDWK	440
Hex	<i>D. discoideum</i>	--ITWNDPIDY----	-GVQLNPET	LVQVWS	SGS-----DLQGI		VNSGYKALVS---	FA	WYLDKQNPDN----				-----NIHYEQDTWQ	417
Hex	<i>E. histolytica</i>	--VWEEVFK-----	-KGNADKNT	IIQVWD	DIR-----LLQQV		VNSGYKAIFS---	AG	FYLDKQMPLCNSYDS	STCVNTHSMVWWTNR				422
Hex	<i>C. albicans</i>	TLQVWHEISGV----	-KNLTSRGY	DVVVSL	SDFL-YLDCGNAGWV		TND--PRYVE---	TP	ENVDFNTG-----				-----QGSWCOPYK	450
Hex	<i>B. mori</i>	PLILWTSTLTIDYTHV	EKFLDKDEYIIQVWT	TGA-----DPQIQGL	LQKGYRLIMSN-YDA		LYFDCGFGAWVG---						-----SGNNWCSPYI	497
Hex	<i>P. gingivalis</i>	--IGWDEILEG----	---GLAPSAT	VMSWR	GED-----GGIAA		ANMNHVIMTPGSGG		LYLDHYQGDP----				-----TVEPVAIGGYAPLE	449
Hex	<i>S. plicatus</i>	--VGWHQLAG-----	--AEPVEGAL	VQYWG	LDRTGDAEKAEEVAEA		ARNGTGLILSP-ADR		TYLDMKYTKDT----				-----PLGLSWAGYV	409
Chb	<i>S. marcescens</i>	-MQAWQDGLKDAES-	SKAFATSRVGVNFWD	TLYWG--GFDSVNDW	ANKGYEVVVS-NPDY		VYMDFPYEVNPD---						-----ERGYWGTFRS	690

Page 9.1

		721	735	736	750	751	765	766	780	781	795	796	810	
Hex A	Human α -subunit	DFYVVEPLAFEG--T	PEQKAL-----				-----VIGGEAC	MWGEYVDNTN-LVPR	LWPRAGAVAERLWS-		---NKLTSD-----			492
Hex A	Mouse α -subunit	DMYKVEPLAFHG--T	PEQKAL-----				-----VIGGEAC	MWGEYVDSTN-LVPR	LWPRAGAVAERLWS-		---SNLTTN-----			491
Hex B	Human β -subunit	KYYKVEPLDFGG--T	QKQKQL-----				-----FIGGEAC	LWGEYVDATN-LTPR	LWPRASAVGERLWS-		---SKDVRD-----			521
Hex B	Cat β -subunit	NYYKVDPLHFDG--S	QEQQKL-----				-----VIGGEAC	LWGEFVDATN-LTPR	LWPRASAVGERLWS-		---PEDITS-----			463
Hex B	Mouse β -subunit	NYYKVEPLNFEG--S	EKQKQL-----				-----VIGGEAC	LWGEFVDATN-LTPR	LWPRASAVGERLWS-		---PKTVTD-----			500
Hex	<i>D. discoideum</i>	DFYAADPTNNIS--T	NAEN-----				-----IIGGEAT	MWAEQINQVN-WDVR	VWPRAIGIAERLWS-		---AQSVNS-----			475
Hex	<i>E. histolytica</i>	DMYDNDPVKSLS--S	SEKEN-----				-----VLGGEGC	SWGESTDEQN-FFDR	VFQRYSAIAERLWS-		---KESVVD-----			481
Hex	<i>C. albicans</i>	SYQRIYNFDF TANLT	ETEKNH-----				-----VLGREAA	LWSEQVDSTV-LTTK	IWPRTAALAE L TWS-		---GNKDSNGHHR--			516
Hex	<i>B. mori</i>	GGQKVYGN SPAV-MA	LSYRDQ-----				-----ILGGEVA	LWSEQSDPAT-LDGR	LWPRAAFAERMWA-		---EPSTAW-----			558
Hex	<i>P. gingivalis</i>	QVYAYNPLPKEL--P	ADKHRY-----				-----VLGAQAN	LWAEYLYTSERYDYQ	AYPRLLAVALTWT-		---PLAKKD-----			510
Hex	<i>S. plicatus</i>	EVQRSYDWDPAGYLP	GAPADA-----				-----VRGVEAP	LWTETLSDPDQLDYM	APFRLPGVAN--WA-		---GPRRPR-----			470
Chb	<i>S. marcescens</i>	DERKVFSPAPDNMPQ	NAETSVDRDGNHFNA			KSDKPWPGAYGLSAO	LWSETQRTDPQMEYM	IFPRALSVAERSWHR	AGWEDYRAGREYKG					780

Page 10.1

		811	825	826	840	841	855	856	870	871	885	886	900	
Hex A	Human α -subunit	-----LTFAY	ERLS-----	HFRCEL	LRRGVQA---QP---	LNVGFCEQEFEQT--	-----	-----	-----	-----	-----	-----	-----	529
Hex A	Mouse α -subunit	-----IDFAF	KRLS-----	HFRCEL	VRRGIQA---QP---	ISVGYCEQEFEQT--	-----	-----	-----	-----	-----	-----	-----	528
Hex B	Human β -subunit	-----MDDAY	DRLT-----	RHRCRM	VERGIAA---QP---	LYAGYCNHENM----	-----	-----	-----	-----	-----	-----	-----	556
Hex B	Cat β -subunit	-----VGNAY	NRLT-----	VHRCRM	VRRGISA---EP---	LFTGYCDYEYKT---	-----	-----	-----	-----	-----	-----	-----	499
Hex B	Mouse β -subunit	-----LENAY	KRLA-----	VHRCRM	VSRGIAA---QP---	LYTGycNYENKI----	-----	-----	-----	-----	-----	-----	-----	536
Hex	<i>D. discoideum</i>	-----VSLAL	PRIG-----	HFTCDL	SRRGIQS---GP---	LFPDYCPMQDDLVP-P	TMKPNTKLSKSEIKL	ILNK-----	-----	-----	-----	-----	-----	532
Hex	<i>E. histolytica</i>	-----KESHE	VRAN-----	YLRCLD	VRRDIMKGT-GP---	LYHSFCQLPKKEK-S	N-----	-----	-----	-----	-----	-----	-----	522
Hex	<i>C. albicans</i>	-----GYEFT	QRIL-----	NFREYL	VKLGYGV---SP---	LVPKYCLLNPHAC-D	LYKNPPVY-----	-----	-----	-----	-----	-----	-----	562
Hex	<i>B. mori</i>	-----QDAEH	RML-----	HVRERL	VRMGIQA---ES---	LEPEWCYQNQGLC-Y	G-----	-----	-----	-----	-----	-----	-----	596
Hex	<i>P. gingivalis</i>	-----FADFC	RRLD-----	NACVRL	DMHGinyHIPLP---	EQPGSSDFIAFT-D	KAKLTFTTSRPMKMV	YTLDETEPSLTSTPY						581
Hex	<i>S. plicatus</i>	-----TTGTP	TRCG-----	SPAGPV	LGGGGDR--LLP---	LAAGALDLTRTAR-R	APRRTVSGVPARRWV	RAPRWGS-SGSRGAG						538
Chb	<i>S. marcescens</i>	GETHFVDTOALEKDW	LRFANILGORELAKL	DKGGVAYRLPVPGAR	VAAGKLEANIALPGL	GIEYSTDGGKQWORY	DAKAKPAVSGEVQVR							870

[illegible][illegible]

	'age 13.1	1081	1095	1096	
Hex A	Human α -subunit	-----	-----	-----	529
Hex A	Mouse α -subunit	-----	-----	-----	528
Hex B	Human β -subunit	-----	-----	-----	556
Hex B	Cat β -subunit	-----	-----	-----	499
Hex B	Mouse β -subunit	-----	-----	-----	536
Hex	<i>D. discoldeum</i>	-----	-----	-----	532
Hex	<i>E. histolytica</i>	-----	-----	-----	522
Hex	<i>C. albicans</i>	-----	-----	-----	562
Hex	<i>B. mori</i>	-----	-----	-----	596
Hex	<i>P. gingivalis</i>	KGEKXPISDMLP	Q	-----	777
Hex	<i>S. plicatus</i>	-----	-----	-----	558
Chb	<i>S. marcescens</i>	-----	-----	-----	885

Figure 8 Amino acid sequence alignment *SpHex* and *SmChb*. The pairwise alignment was abstracted from the larger ClustalW1.7 multiple sequence alignment shown in figure 4 (section 2.8.0). Amino acid numbering is indicated down the right side of the figure. “*” indicate identical residues whereas “.” indicates similar residues. “†” indicate residues subjected to mutational analysis. Secondary structures are indicated along the top of alignment (α ; alpha helix, β ; beta sheet). Secondary structures responsible for forming the β/α barrel structure are numbered (β 1- β 8 and α 1- α 6, α 8; note there is no α 7 helix).

SpHex	EDTPRYAWRSAMLDV	SRHFFGVDEVKRYID	RVARYKYNKLHLHLS	DDQGWRIAIDSWPRL	205
SmChb	SDAPRFPYRGIFLDV	ARNFHKKDAVLRLLD	QMAAYKLNKFHFHLS	DDEGWRIEIPGLPEL	392
	* * * . * . . . * * *	* * * * *	. * * * * . * . . . * * *	* * * * * * * . * *	
SpHex	ATYGGSTEV-----	-----GGGPGG	YYTKAEYKEIVRYAA	SRHLEVVPIDMPGH	250
SmChb	TEVGGQRCHDLSETT	CLLPQYGGQPDVYGG	FFSRQDYIDI IKYAO	ARQIEVIPEIDMPAH	452
	** .	**	. . * . . . * *	* . * * . * * * * *	
SpHex	TNAALAS-----	-----Y AEL	NCDGVAPPLYTGTKV	GFSSLCVVDKDVTYDF	291
SmChb	ARAAVVSMEARYKKL	HAAGKEQEA NEFRLV	DPTDTSNTTSVQFFN	RQSYLNPCLDSSQRF	512
	* * . . *	.	.	* . * . * * . * *	
SpHex	VDDVIGELAAIT--	--PGRYLHIGGDEAH	S-----	-----	317
SmChb	VDKVIGELIAQMHEA	GQPIKTWHFGGDEAK	NIRLGAGYTDKAKPE	PGKGIIDQSNE DKPW	572
	** * * * . *	* . * . * * * .	.	.	
SpHex	-----	--TPKADFVAFMKRV	QPIVAKYGKTVVGWH	QLAGAEPVEG----A	356
SmChb	AKSQVCQTMKEGKV	ADMEHLPSYFGQEV	KL VKAHGIDRMQAWQ	DGLKDAESSKAFATS	632
	.	. . .	. . * . .	.	
SpHex	LVQYWGLDRTGDAE	KAEVAAEARNGTGLI	LSPADRTYLD MKYTK	D--TPLGLSWAGYVEV	414
SmChb	RVGVNFWDITLYWGG	FDSVNDWANKGYEVV	VSNPDYVYMDFPYEV	NPDERGYWGTFRSDE	692
	* . . * .	* . * * . .	* * . * * . *		
SpHex	QRSYDWD PAGYL--	-----	PGAPADAVRGVEAPL	WTETLSDPDQLDYMAF	457
SmChb	RKVFSFAPDNMPQN	AETSVDRDGNHFN AK	SDKPWPGAYGLSAQL	WSETQRTDPQMEYMIF	752
	. * . .	.	* . * . * *	* . * * * * . * * *	
SpHex	PRLPGVANWAGPRR	PR-----	-----TTGTP	TRCGSPAGPVLGGGGD	494
SmChb	PRALSAERSWHRA	GWEQDYRAGREYKGG	ETHFVDTQALEKDWL	RFANILGQRELAKLDK	812
	* * . . . * *	. *	. .	 *	
SpHex	RLLP	498			
SmChb	GGVA	817			

repositioned (1-5 residues in either the N or C-terminal direction) such that they mapped more precisely onto loop regions of *SmChb* domain III (described in section 2.8.0).

The pairwise alignment in figure 8 indicated that no amino acid insertions into the *SpHex* model were required, implying that there were no structural features within the modeled region of *SpHex* that were not present in the *SmChb* structure. After the amino acid sequence of *SpHex* had been substituted onto the α -carbon backbone of the equivalent region of the *SmChb* structure, all 9 deletions in the model (corresponding to the deletions in the pairwise alignment) were manually joined using the graphics program O. O allows the user to manipulate the three-dimensional conformation of the α -carbon backbone by rotating ψ , ϕ bond angles. Furthermore, O allows the user to rotate any of the bonds composing the amino acid side chains. Thus, all 9 deletions (none of which were found to create gaps greater than 10 Å) were joined together such that the *SpHex* model became one continuous polypeptide chain. In order to relieve any steric clashes (considered to be any atomic interaction of 2.5 Å or less) within the model, and to improve any unfavorable bond angles and/or bond lengths, the model was energy-minimized as described in section 2.8.2. The final energy-minimized model of *SpHex* is shown as a ribbon diagram in figure 9A, B.

A summary of the PROCHECK analysis before and after energy minimization is shown in Table II. Table II demonstrates that energy minimization dramatically improved the overall geometry and relaxed most of the bad contacts within the model. Of the 10 bad contacts remaining, none occur within the core of the model; only one contact, between the last two residues in the model, is less than 2.5 Å (Leu 497 C $\leftrightarrow$ Pro 498 N, 1.3 Å). A Ramachandran plot of the final energy minimized model (Fig 10) indicated that 94.3% of the

Figure 9 ***SpHex* model complexed with chitobiose.** The model contains residues Pro 146 - Pro495. *Panel A*, shows a ribbon diagram of the model looking down the axis of the α/β barrel. *Panel B*, is turned 90° on the Y-axis with respect to *panel A* and shows a side view of the barrel. Note that the substrate complexes with loop structures on the C-terminal end of the α/β barrel and secondary structures forming the α/β barrel are indicated: red cylinders; α -helix, green ribbons; β -sheets. *Panel C*, shows residues predicted to be involved in the catalytic mechanism of the β -N-acetylhexosaminidase. *Panel D*, depicts the predicted hydrophobic interactions occurring between the enzyme and substrate.

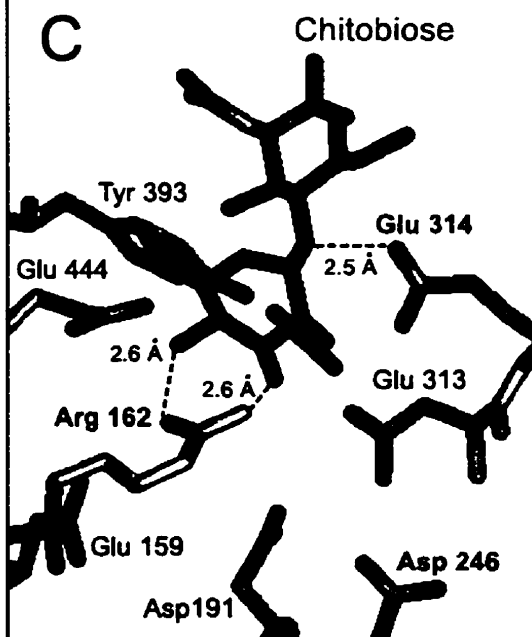
A



B



C



D

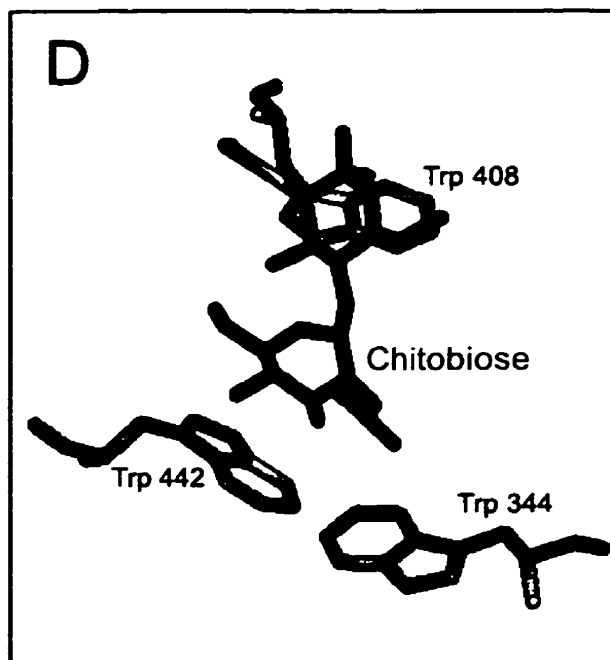


Table II

*PROCHECK analysis of the quality of the bond angle geometry within the SpHex model
before and after energy minimization using X-PLOR*

<u>Parameter measured</u>	<u>Before energy minimization</u>	<u>After energy minimization</u>
Number of bad contacts*	105	10
Main-chain bond lengths**	66.2% within normal limits 33.8% generously allowed 8% unacceptable	97.1% within normal limits 2.9% generously allowed
Main-chain bond angles**	56.5% within normal limit 43.5% generously allowed 21% unacceptable	71.7% within normal limits 28.3% generously allowed 9% unacceptable
Planar groups**	95.6% within normal limit 4.4% generously allowed	89.8% within normal limits 10.2% generously allowed 4% unacceptable

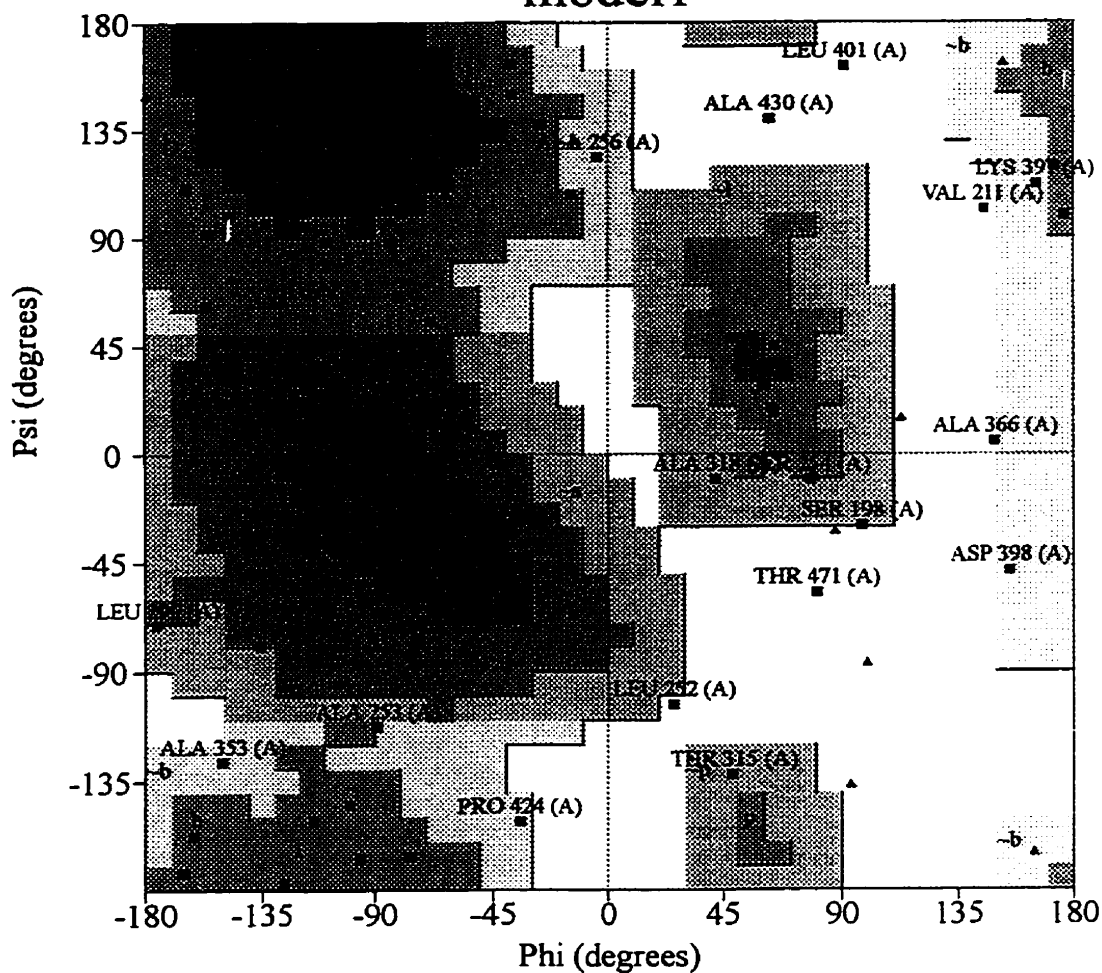
* any atomic interaction of 2.5 Å or less

** Bond lengths, angles and planar group geometry quality is based on the guidelines of acceptable geometry as defined by Engh and Huber (98).

Figure 10 **Ramachandran plot of energy minimized *SpHex* model.** The model was placed in a primitive lattice (P1 space group) and subjected to energy minimization using a conjugate gradient target function implemented in X-PLOR. The stereochemical quality of the energy minimized model was assessed using PROCHECK and output as a Ramachandran plot.

PROCHECK

Ramachandran Plot modell



Plot statistics

Residues in most favoured regions [A,B,L]	223	78.2%
Residues in additional allowed regions [a,b,l,p]	46	16.1%
Residues in generously allowed regions [-a,-b,-l,-p]	10	3.5%
Residues in disallowed regions	6	2.1%
Number of non-glycine and non-proline residues	285	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	39	
Number of proline residues	25	
Total number of residues	350	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

residues had ϕ , ψ torsion angles within core and allowed regions, 3.5% had ϕ , ψ angles within generously allowed regions and 2.1% (6 residues) had disallowed ϕ , ψ angles. The residues with disallowed ϕ , ψ angles (Val 214, Leu 255, Ala 369, Leu 404, Ala 433, Thr 474) all reside in loop structures on the outer surfaces of the model including locations that were manually joined together to compensate for deletions.

The modeled *SpHex* is similar in structure to *SmChb* domain III, forming an eight stranded α/β barrel where the seventh alpha helix of the barrel has been replaced by an extended loop structure (Fig 9A, B). As judged from the pairwise sequence alignment between *SpHex* and *SmChb*, the model predicts a significantly more compact α/β barrel structure for the *SpHex* as opposed to *SmChb* (Fig 11). Interestingly, it appears that some of the large loop structures present in domain III of *SmChb* are used to stabilize adjacent domains; however, because these domains do not appear to exist in the *SpHex* structure, the corresponding loops are also absent from the *SpHex* model. The most prevalent loop in *SmChb* (used to stabilize domain II) that is not present in the *SpHex* model is indicated by an arrow in figure 11. This compact α/β barrel structure is comparable to the theoretical model constructed for the human Hex A α -chain (Fig 12) (55).

Because the *SmChb* x-ray structure was solved complexed with chitobiose, the *SpHex* model could be built complexed with the same substrate (Fig. 9A, B). Of the natural substrates known to be hydrolyzed by family 20 enzymes, chitobiose is probably the most analogous to 4-MUG. It is therefore likely that chitobiose and 4-MUG complex with the *SpHex* in a similar fashion; thus, the kinetic data collected using 4-MUG as substrate (see below), should directly support the comparative modeling results.

Figure 11 Comparison of *SpHex* model with *SmChb* X-ray structure. *Panel A*, shows a ribbon diagram of the *SmChb* X-ray structure looking down the axis of the α/β barrel of domain III (cyan). *Panel B*, is turned 90° on the Y-axis with respect to *panel A* and shows a side view of the barrel of domain III. The predicted structure of *SpHex* is homologous to *SmChb* domain III. *Panel C* shows a ribbon diagram of the *SpHex* model (yellow) (looking down the axis of the α/β barrel) superimposed onto a α -carbon trace of *SmChb* domain III (cyan). *Panel D* is turned 90° on the Y-axis with respect to *panel C*. The largest loop structure of *SmChb* domain III (thought to provide structural support to domain I) not present in *SpHex* and is indicated with an arrow.

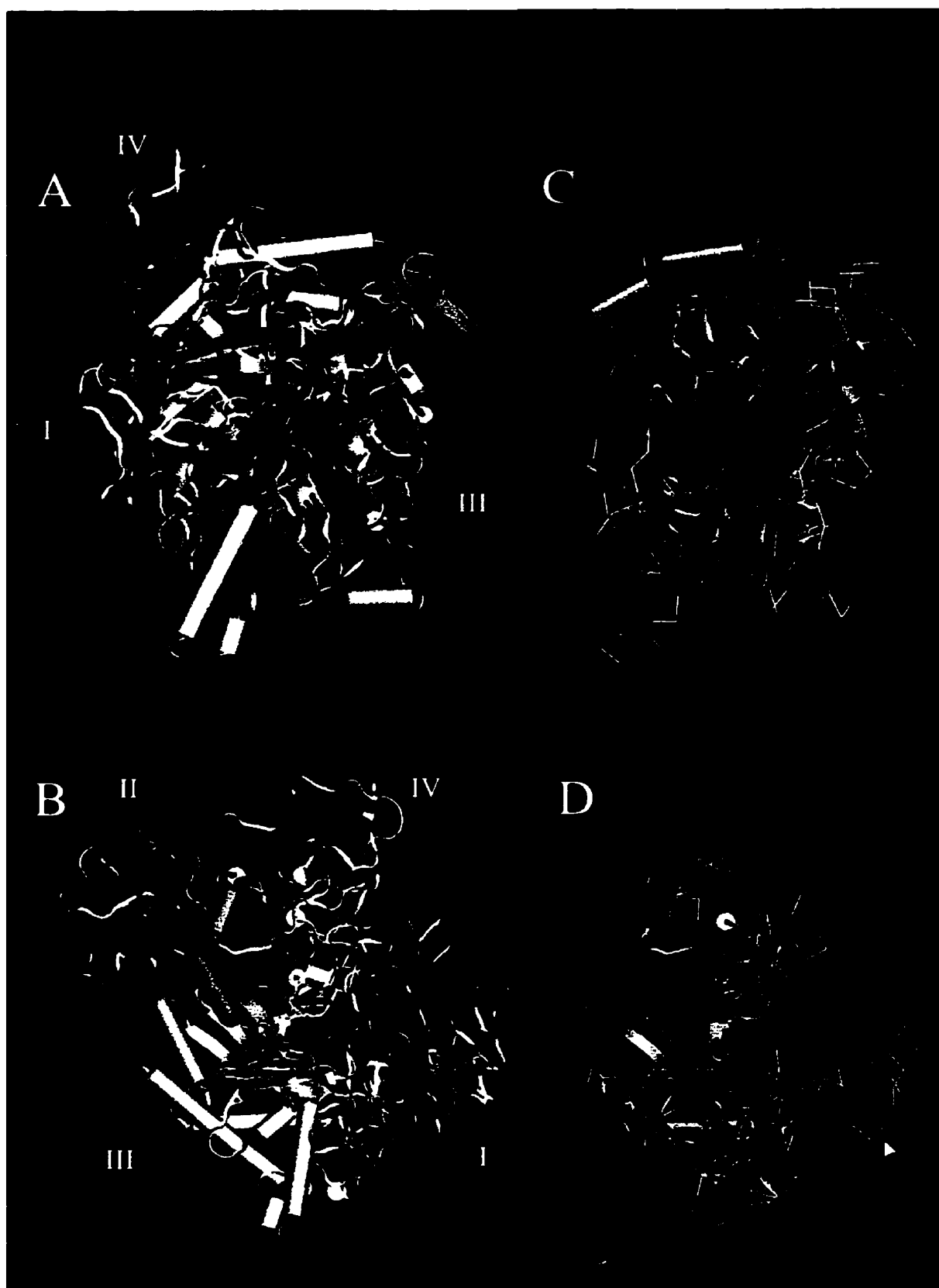


Figure 12 **Comparison of Human Hex A α -chain and the SpHex model.** *Panel A*, shows superimposed α -carbon traces of SpHex (yellow) and the human Hex A α -chain looking down the axis of the α/β barrel. *Panel B*, is turned 90° on the Y-axis with respect to *panel A* and shows a side view of the barrel.

A



B



As expected, the active site architectures of the *SpHex* model and the *SmChb* are highly conserved. The salient features, as identified in *SmChb* (55), include Glu314 (*SmChb*; Glu540), Arg 162 (*SmChb*; Arg349) (Fig. 9C) and three Trp residues Trp344, 442, 408 (Fig. 7d). The *SpHex* model indicates that the protonated carboxyl group of Glu314 is within hydrogen bonding distance (2.5 Å) of the oxygen in the glycosidic linkage of the substrate. Thus, Glu314 is perfectly situated to donate its proton to the glycosidic oxygen, thereby aiding in the hydrolysis of the substrate. The δ -guanido group of Arg162 appears to directly bind the substrate by forming two hydrogen bonds, each at a distance of 2.6 Å, to OH3 and OH4 of the non-reducing *N*-acetylglucosamine sugar of the chitobiose. As described for the *SmChb* X-ray structure, the active site model of *SpHex* appears to lack an appropriate residue that could act as the nucleophilic base in an acid/base reaction for this enzyme. As described for the *SmChb* catalytic mechanism (55), the nucleophile involved in the *SpHex* catalytic mechanism may be provided by the substrate itself.

A notable hydrogen bond is predicted to occur between the acetamido-O6 of the non-reducing sugar of the chitobiose substrate and the OH of the phenolic ring on Tyr393 of *SpHex* (*SmChb*; Tyr 683)(Fig 9C). This hydrogen bond would only occur when the chitobiose is in the energetically unfavorable 4-sofa conformation as described for the *SmChb* mechanism by Tews *et al.* (55). The twisting of chitobiose into a 4-sofa conformation is proposed to be essential for the substrate assisted catalytic mechanism for it allows the acetamido-O7 atom of the non-reducing *N*-acetylglucosamine to move within 3.0 Å of the C1 atom of the same sugar, an ideal position for the nucleophilic attack (55). This twisting of the chitobiose may be partially stabilized by hydrogen bonding to Tyr 393 in *SpHex*, as it is

by the equivalent residue (Tyr 683) in *SmChb*. To see potential for a similar hydrogen bonding interaction occurring in the *SpHex* model provides additional support for the catalytic mechanism of *SmChb* and helps validate the quality of the model.

The active site pocket of our model contains Trp residues that pack against the hydrophobic hexose rings of the chitobiose substrate (Fig. 9D). The indole rings of Trp 442 and Trp 344 appear to stabilize hydrophobic hexose ring of the non-reducible *N*-acetylhexosamine of the chitobiose substrate, exactly as is seen in the *SmChb* X-ray structure. Although Trp408 appears to be stabilizing the hexose ring of the reducible *N*-acetylglucosamine sugar of the substrate, the indole ring of this Trp is not parallel to the hexose ring of the sugar as is seen in the *SmChb* X-ray structure. The reason for this difference could not be explained.

3.5 Construction of the pMAL-c2-*SpHex* fusion protein vector

The construct pMAL-c2-*SpHex* (a gift from NEB) was made by subcloning the insert from clone 36 (*BamH* I/*Hind* III) into pMal-c2 (*Xmn* I/*Hind* III). The *BamH* I site was filled-in to facilitate subcloning and allow for correct translation of the maltose binding protein (MBP)-*SpHex* fusion protein. To remove 3.2 kb of extraneous DNA 3' to the *SpHex* coding region, pMAL-c2-*SpHex* was digested with *Sma* I and *Hind* III, the *Hind* III site was blunt-ended, and both the larger vector (7 kb) and a 1.8 kb (*Sma* I/*Sma* I) fragment containing the coding region, were gel purified using GeneClean (section 2.4.3) and ligated to generate pmHEX-1.8 (Fig 4). A plasmid with the fragment ligated in the correct orientation was identified using 4-methylumbelliferyl- β -*N*-acetylglucosaminide (4-MUG) (Toronto Research Chemicals Inc.) as substrate for an in-plate assay of *E. coli* colonies (section 2.13.0)

electroporated with the pmHEX-1.8 ligation reaction.

3.6 Site-directed Mutagenesis of the *SpHex* gene

Single-stranded DNA from psHEX-1.8 was isolated as described in section 2.6 and mutations were introduced as described in section 2.9. Phosphorylated oligonucleotides (ACGT Corporation, Toronto, ON) 5'-TCGGCGGCGACCAGGCGCACTCCAC-3', 5'-CACGAAGTCGTAGGTGACGT-3' and 5'-GTCCCCGAGATCAACATGCCGGGCC-3' were used to create the substitutions G1208C (Glu314Gln), G753A and G754T (Arg162His) and G1004A (Asp246Asn) respectively. Mutant plasmid constructs were identified by PCR amplification of the relevant region followed by restriction enzyme digestion as described in section 2.9.2 and 2.9.3 (base change G1208C created a *ScrFI* site, G753A and G754T combined destroyed a *MspI* site and G1004A destroyed a *TaqI* site). Fragments containing the mutation(s) were subcloned into pmHEX-1.8 (*AscI/BsrGI* fragment contained either G753A and G754T or G1004A, *BsrGI/AgeI* fragment contained G1208C) to create the mutant pmHEX-1.8 constructs. The subcloned regions were sequenced to confirm that only the desired base changes were introduced.

3.7 Purification of *SpHex*-MBP Fusion and Cleavage with Factor Xa Protease

Purification of the *SpHex*-MBP fusion protein, encoded by pmHEX-1.8, was performed as described in section 2.10. The pmHEX-1.8 constructs of wild type and mutant *SpHex*-MBP produced high-level fusion protein expression in *E. coli*, yielding approximately 0.5 mg/100 ml culture. The final protein concentrations for all purifications were brought to 1-2 mg/ml by eluting the fusion protein with 0.5 ml elution/storage buffer. Purified wild type and mutant *SpHex*-MBP fusion protein was stable for at least 48 hrs at room temperature and

at least 2 weeks at 4°C. *SpHex* displayed catalytic activity toward both 4-MUG and 4-MUGS; however, the enzyme was approximately 2600 fold more active toward 4-MUG than 4-MUGS (Fig 13). An SDS-PAGE gel of the purified proteins is shown in figure 14.

The purified *SpHex*-MBP fusion protein was digested with factor Xa protease as described in section 2.11. To assess the effects of the MBP on activity *SpHex*, digested *SpHex*-MBP fusion protein activity was measured as described in section 2.13.1. There was no difference in the catalytic activity between the undigested or digested enzyme (Fig 15).

3.8 Western blot analysis of recombinant and native *SpHex*

Western blot analysis using the anti-*SpHex*-MBP antibody indicated that factor Xa cleaved the *SpHex*-MBP fusion protein into a 55 kDa protein predicted to be recombinant *SpHex* and the 43 kDa MBP (fig 16; lane 3). Anti-*SpHex*-MBP antibody also detected a 55 kDa protein in the cytosol of *S. plicatus* (fig 16; lane 4), indicating that the recombinant and native *SpHex* are of the same approximate molecular mass. Two higher molecular weight proteins were also detected in the *S. plicatus* cytosol; however, their identity could not be determined and may be a result of nonspecific cross-reactivity of the antibody to other proteins in the fraction. Interestingly, *SpHex* was detected in the spent medium of *S. plicatus* cultures after approximately 5 days of growth, suggesting *S. plicatus* may secrete this enzyme into the media upon expression (Fig 16; lane 5).

3.9 Kinetic analysis of wild-type and mutant *SpHex*

Based upon multiple sequence alignments of family 20 enzymes, previous mutagenesis studies of human Hex A (48-50; 99) and the *SmChb* structure (55), we created the following mutations in *SpHex*: Arg162His, Asp246Asn and Glu314Gln. Since the kinetics of the *SpHex*-MBP fusion and factor Xa digested fusion protein were indistinguishable when

Figure 13 ***SpHex* catalytic activity towards 4-MUG and 4-MUGS.** To determine if 1) *SpHex* was catalytically active toward both 4-MUG and 4-MUGS and 2) the relationship between *SpHex* protein concentration and the rate of product formation, increasing concentrations purified *SpHex* were incubated at 25°C for 45 min in a 30 µl reaction containing either 3.5 mM 4-MUG or 4-MUGS in citrate/phosphate buffer, 0.3% BSA, pH 3.0 at protein (n=3).

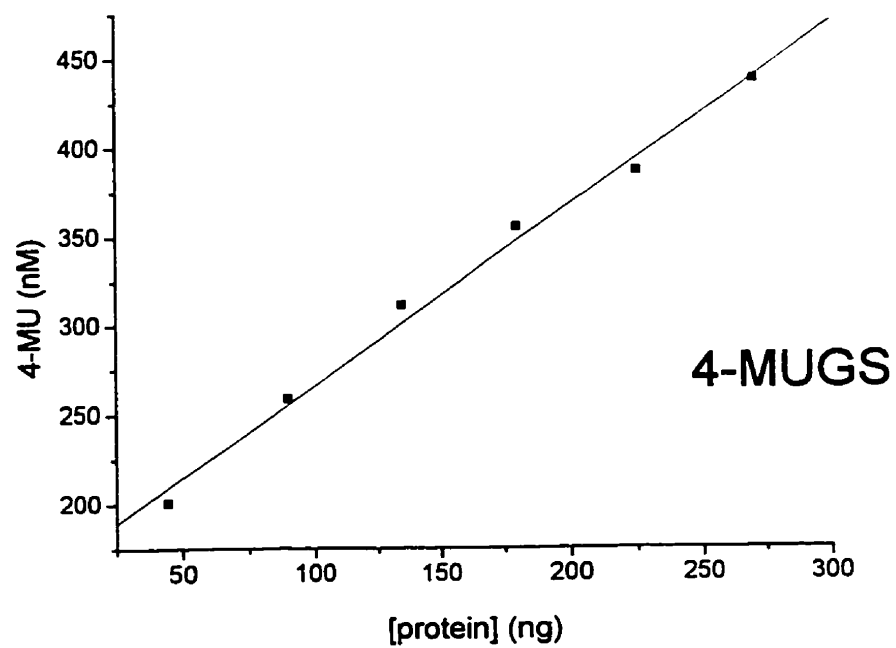
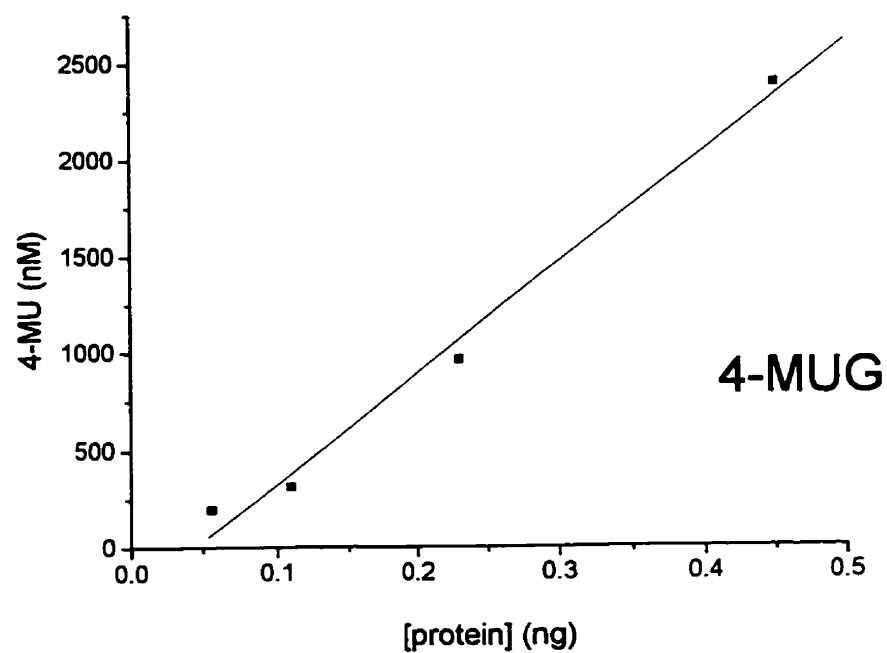


Figure 14 SDS-PAGE of purified wild type and mutant *SpHex*-MBP fusion protein. *SpHex*-MBP fusion protein was purified as described in section 2.10. To assess purity, purified and unpurified *SpHex*-MBP fusion protein (20 μ l) was separated in an 8.0% SDS-PAGE gel as described in section 2.13. *Lanes 1* and *2* consist of purified and unpurified wild type *SpHex*-MBP respectively. *Lanes 3* and *4* consist of purified and unpurified Asp246Asn mutated *SpHex*-MBP respectively. *Lanes 5* and *6* consist of purified and unpurified Arg162His mutated *SpHex*-MBP respectively. *Lanes 7* and *8* consist of purified and unpurified Glu314Gln mutated *SpHex*-MBP respectively. Mobility of prestained protein standards (NEB) are indicated on the left.

kDa

175 —

83 —

62 —

47.5 —

32.5 —

1

2

3

4

5

6

7

8

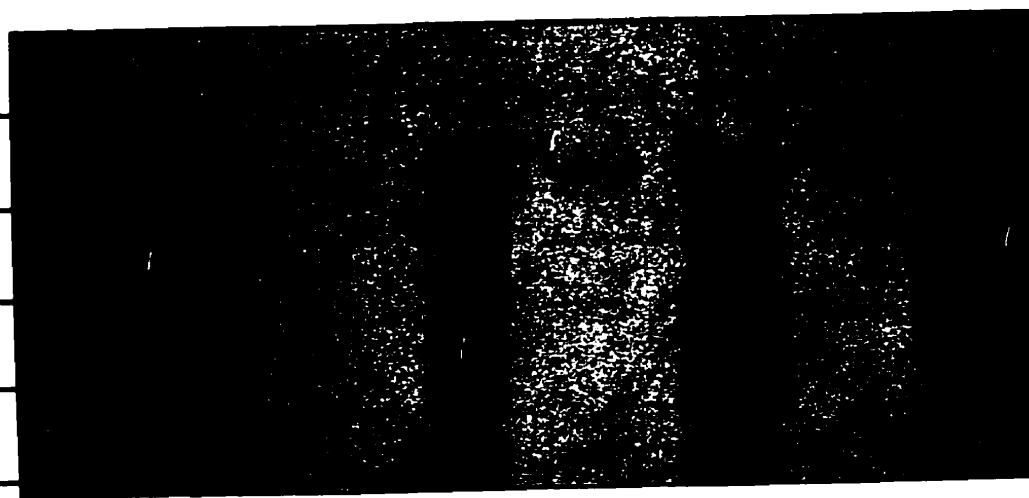


Figure 15 **Comparison of factor Xa digested and undigested *SpHex*-MBP fusion protein activity.** Kinetic analysis was performed as described in section 2.15.1 at pH 3.0 using 1 ng of protein in each reaction containing 3 mM 4-MUG as substrate.

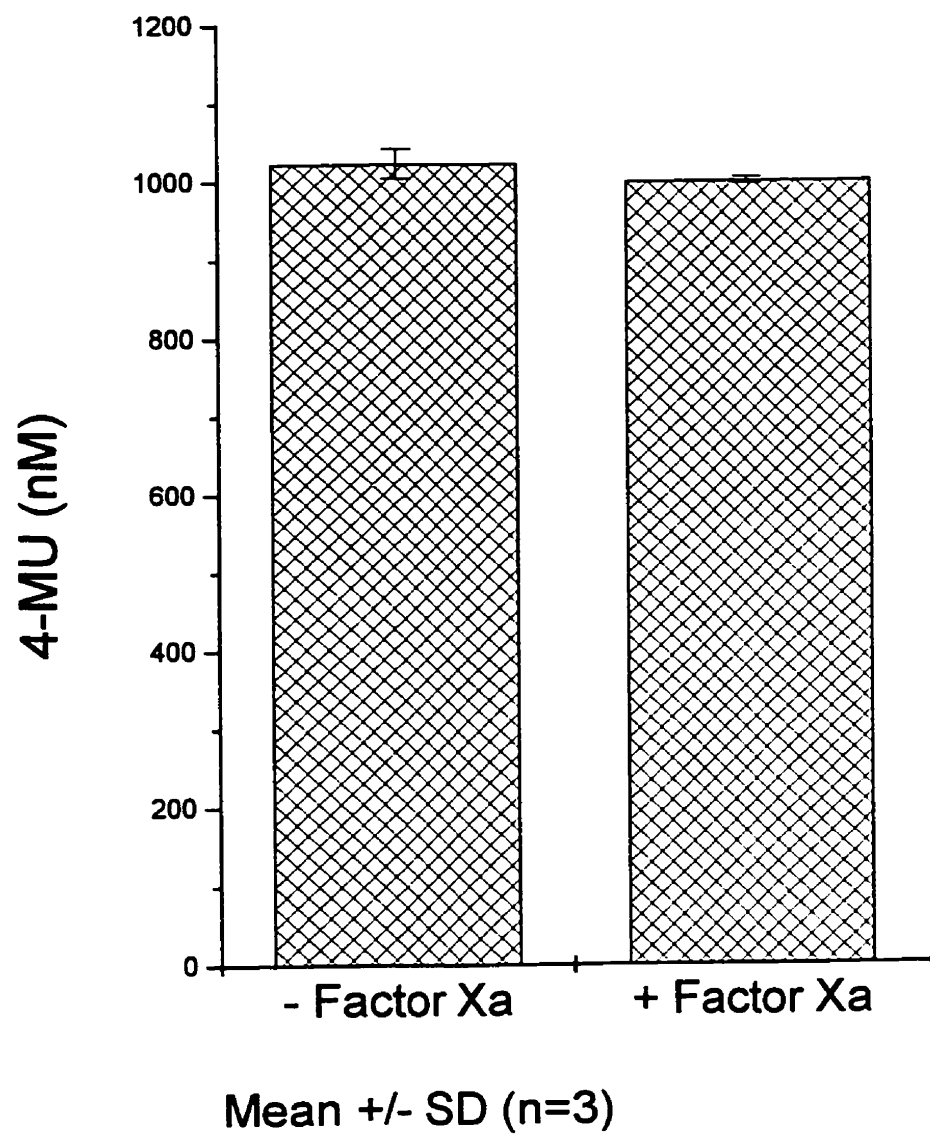
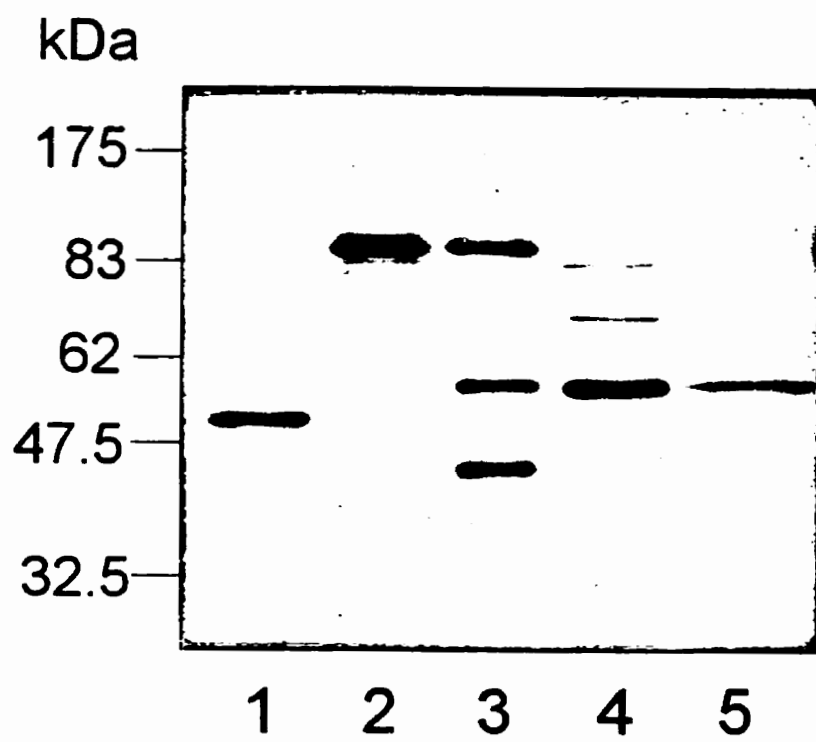


Figure 16 Comparison of recombinant and native *SpHex*. To directly compare molecular masses of recombinant and native *SpHex*, the MBP was removed from affinity purified *SpHex*-MBP fusion with factor Xa protease as described in section 2.11. A crude *S. plicatus* cytosolic fraction was prepared after growth in ISP medium for 5 days (section 2.12). Protein samples were resolved by 8.0% SDS-PAGE, transferred to nitrocellulose, and incubated with polyclonal anti-*SpHex*-MBP antibody (1:10000) as described in section 2.14.2. The antigen-antibody complex was visualized using the ECL protocol using a 1:10000 dilution of donkey-anti-rabbit-IgG-horseradish peroxidase conjugate. Mobility of pertained protein standards (NEB) are indicated on the left. *Lane 1*, 10 ng affinity purified MBP-*LacZ* α fusion (expressed from pMAL-c2 with no insert in the polylinker); *lane 2*, 10 ng affinity purified *SpHex*-MBP fusion protein; *lane 3*, 10 ng incomplete factor Xa digest of *SpHex*-MBP fusion protein; *lane 4*, 15 μ g *S. plicatus* crude cytosolic fraction; *lane 5*, 4 μ g ISP medium used to grow the *S. plicatus* culture.

IB: anti-SpHex-MBP



using 4-MUG as substrate, kinetic analysis was restricted to the *SpHex*-MBP fusion protein. The resulting kinetic analyses (Table III) are consistent with our predicted structure of *SpHex* (Fig 9c). The model of *SpHex* predicts that the Arg162His mutation would significantly affect the substrate binding capacity of the enzyme (Fig 9c). Indeed, this mutation had a pronounced affect on K_m , increasing this intrinsic rate constant 40 fold as compared to wild type *SpHex*. The proposed catalytic acid residue Glu314, when mutated (Glu314Gln), reduced V_{max} 296 fold, and the K_m of this mutant was reduced 7 fold, implying that Glu 314 is the proton donor in the catalytic reaction of *SpHex*.

In *SmChb*, Asp448 appears to indirectly hydrogen bond with the chitobiose substrate through a water molecule and Asp539. The equivalent residue in *SpHex* is Asp246. Comparative modeling suggests that Asp246 may fulfil a similar role in this enzyme. The mutation Asp246Asn results in a 2 fold decrease in V_{max} and a 2 fold increase in K_m . In *SpHex*, Asp246 most likely forms an indirect hydrogen bond with the substrate through Glu313.

3.10 pH Profile Analysis

The pH profiles of wild type and mutant *SpHex*-MBP fusion protein can be seen in figure 17. Hydrolysis of 4-MUG by *SpHex* was optimal under acidic conditions. The pH optimum of wild type *SpHex* was found to be 3.0. Although the pH optimum of this enzyme is quite broad, an environment more acidic than pH 3.0 resulted in a steady loss of activity. Interestingly, the Glu314Gln mutation changed the pH profile from that of a broad bell shaped curve to more a hyperbolic shaped pH curve with no distinctive peak in the pH profile. According to our model and Tews *et al.* (55), *SpHex* Glu 314 and *SmChb* Glu 540 must be protonated in order for catalysis to occur. This is reflected in the acidic pH optimum seen

Table III

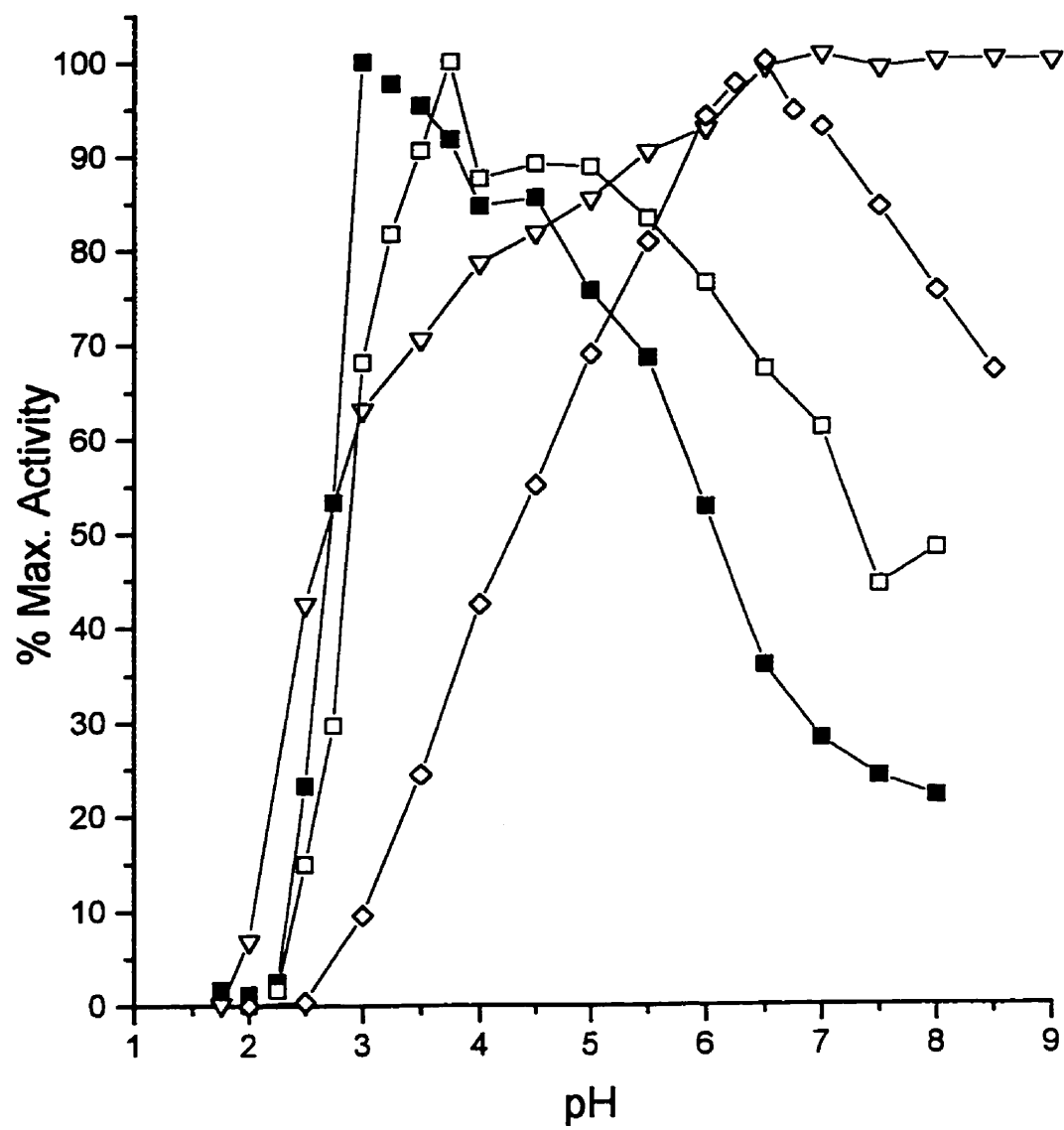
Kinetic studies of recombinant SpHex-MBP fusion protein

Recombinant wild type and mutant *SpHex-MBP* fusion protein activities were determined by incubating a reaction mix containing 1-75 ng of affinity purified *SpHex-MBP* fusion protein with 12 various concentrations of 4-MUG (0.033-5.0 mM) in citrate/phosphate buffer, 0.3% bovine serum albumin, pH 3.0, in 30 μ l at 25°C for 45 min. The K_m and V_{max} values were determined using the direct linear plot method (84).

Enzyme	K_m^*	V_{max}^*
	<i>mM</i>	$\mu\text{mole/min/mg}$
Wild type	0.14 ± 0.01	53.2 ± 2.1
Arg162His	5.67 ± 0.07	10.8 ± 0.72
Glu314Gln	0.02 ± 0.01	0.18 ± 0.01
Asp246Asn	0.25 ± 0.01	23.3 ± 0.16

* Mean $\pm$ S.D. (n = 3)

Figure 17 Analysis of pH profiles for wild type and mutant *SpHex*. Catalytic activity for wild type and mutant *SpHex* was measured in 10 mM Na₂HPO₄, 6 mM citric acid, 0.3% BSA ranging from pH 2.0 - 9.0 using 4-MUG as a substrate as described in section 2.15.1. After a 45 min. incubation at 25 °C, each reaction was quenched with 0.1 M glycine-carbonate buffer, pH 10 and the fluorescent product, 4-MU, was measured by fluorometry. Activity is indicated as an average percentage of maximal activity (n=3).



—■—WT
—▽—E314Q
—□—D246N
—◇—R162H

for wild-type *SpHex*. Gln does not undergo ionization; hence, the loss of a distinct peak at pH 3.0 suggests that the ionization state of Glu314 is responsible for the pH dependence seen for the wild-type enzyme.

The Asp246Asn mutation shifted the pH optimum from 3.0 to 3.75. This shift in pH optima contradicts results from human Hex A studies which showed that the analogous mutation in Hex A (Asp258His) resulted in a more acidic pH optima (50). The reason for the pH optimum shift could not be explained. Finally, the Arg162His mutation had a dramatic affect on the pH optimum for the *SpHex*, shifting it from pH 3.0 to pH 6.5. No clear explanation could be identified for this change; the drastic nature of this pH shift may suggest that the mutation has allowed for an alternative catalytic mechanism.

4 Discussion

4.0 Clone 36 and the *SpHex*-MBP fusion protein expression/purification system

Originally, the fusion protein encoded by Clone 36 was classified as a β -*N*-acetylhexosaminidase on the basis of its substrate specificity (65). In a search for enzymes related to human β -*N*-acetylhexosaminidase that might be useful for structure/function analysis, the complete nucleotide sequence (1683 bp) and deduced amino acid sequence (561 aa) of the enzyme encoded by Clone 36, *SpHex*, was determined. Comparison of the deduced amino acid sequence with the SWISS-PROT database and the Brookhaven protein database (PDB) (Table I) indicated that not only did *SpHex* have the substrate specificity of a β -*N*-acetylhexosaminidase, its primary amino acid sequence was most similar to those glycosyl hydrolases that are members of family 20 (chitobiases and β -*N*-acetylhexosaminidases) as classified by Henrissat (2; 3).

The ease with which highly active, soluble *SpHex*-MBP fusion protein could be expressed and purified, made this enzyme an excellent family 20 enzyme to study structurally and functionally. Because *SpHex* originated from a bacteria, the need for complex post-translational modification was not required for normal function, making this enzyme compatible with overexpression in *E. coli*. Furthermore, the fusion of *SpHex* to MBP allowed for rapid, one-step affinity purification using amylose resin. Thus, this system resulted in a yield of highly pure, functional β -*N*-acetylhexosaminidase-MBP fusion protein (*SpHex*-MBP). The catalytic activity of the *SpHex* fusion did not appear to be affected by the presence of the MBP; therefore, the MBP did not have to be removed after purification of the fusion protein (Fig 14).

Assuming that the catalytic mechanism of family 20 enzymes is conserved, it was hypothesized that functional studies of *SpHex*, in combination with molecular modeling of its structure, could provide valuable information not only about its own catalytic mechanism, but of other family 20 enzymes as well, especially human β -*N*-acetylhexosaminidase.

4.1 Comparative molecular modeling of *SpHex*

Building a theoretical model of an enzyme by comparative molecular modeling can provide structural details of the enzyme that greatly enhance the understanding of results obtained from functional studies of the enzyme by site directed mutagenesis and other molecular biology techniques. Furthermore, comparative molecular modeling can provide insight into the structural relationships between enzymes, in this case, these structural relationships were studied within the *O*-glycosyl hydrolase family 20. Interestingly, comparative molecular modeling of *SpHex* demonstrated that even though *SpHex* and *SmChb* belong to the same family and therefore are thought to originate from a common ancestral gene (6), large structural differences between the two enzymes appear to have occurred during evolution. However, the elements secondary structure and key residues needed to preserve their common catalytic mechanism have remained unchanged (Figs 9c and 11).

The regions of highest amino acid sequence identity between *SpHex* and *SmChb* occurred within the sequence comprising the elements secondary structure of their β/α -barrels (Fig 8). This evolutionary conservation suggested that the structural elements of the β/α -barrel were crucial to the function of the enzymes. The major structural differences and lowest sequence identity between the two enzymes occurred within the loop structures that connect the conserved secondary elements of the β/α -barrel (Fig 11). Although the spatial

arrangement of the secondary elements of the β/α -barrels of both enzymes appeared to have been conserved, the lengths of the intervening loop structures were dramatically different. Some of the loops form entire domains within *SmChb* that were absent from the predicted *SpHex* structure.

Surprisingly, the key residues in the active site of *SpHex* (Glu 314 and Arg 162) and *SmChb* (Glu 540 and Arg349) were found positioned within two of the highly variable loop structures at the C-terminal end of the barrel. This is in keeping with the observation that the active site residues of enzymes with a β/α -barrel type fold are usually found within the highly variable loop structures at the C-terminal end of the β/α -barrel (100). Having the key catalytic residues positioned within these highly variable loop regions helps to explain the enormous variety of substrate specificities that various β/α -barrel type glycosyl hydrolases display. This also supports the theory that enzymes with a β/α -barrel type fold that may have completely different substrate specificities, have all evolved from a common ancestral gene (101; 102).

The modeling studies alone indicate that the catalytic mechanism of family 20 enzymes is conserved. These models indicate that *SpHex*, human β -N-acetylhexosaminidase (55) and *SmChb* all used the same catalytic acid residue (*SpHex* Glu 314, human α -Glu 323, β -Glu 355, *SmChb* Glu 540) and substrate binding residue (*SpHex* Arg162, human α -Arg 178, β -Arg 211, *SmChb* Arg 349), and that *SpHex* used substrate assisted catalysis just as predicted for *SmChb* (55). The x-ray crystallographic model of *SmChb* demonstrated that the enzyme can twist the chitobiose substrate into an energetically unfavorable conformation within the active site by using three strategically positioned Trp residues that pack against the

hydrophobic hexose rings of the substrate. Because *SmChb* is missing an appropriate residue to act as a nucleophile during catalysis, Tews *et al* suggested that the twisting of the substrate allows for the *N*-acetamido group of the non-reducing sugar to come close enough to the glycosidic linkage such that it could perform the role of the nucleophile as described for the retaining mechanism in figure 1 (55).

The three Trp residues used to distort the substrate in the *SmChb* active site were found conserved in the *SpHex* sequence and appeared to perform a similar function in the three-dimensional *SpHex* model. The x-ray crystallographic structure of the complex between *SmChb* and chitobiose depicted chitobiose in the twisted conformation and thus was thought to represent the 'pre-Michaelis' complex of chitobiose with *SmChb*. When this conformation of chitobiose was placed into the active site of *SpHex*, not only did the conserved Trp residues pack very well against hexose rings of the distorted substrate (Fig 9d), the conformation of the substrate appeared to be stabilized by hydrogen bonding with Arg 162 and Tyr 393 (Fig 9c). In addition to the *SpHex* active site being able to perfectly accommodate chitobiose in a conformation suited for the mechanism of substrate assisted catalysis, the *SpHex* model also demonstrated that the enzyme was lacking an appropriately situated acidic residue that could act as a nucleophile during catalysis, suggesting that *SpHex* must also rely on substrate assisted catalysis.

The theory of substrate assisted catalysis appears to extend to human β -*N*-acetylhexosaminidase as well, further substantiating the suggestion that the catalytic mechanism of the family 20 enzymes is conserved. When Liessem *et al.* (1995) photoaffinity labeled the β -subunits of human Hex B, the active-site directed affinity ligand reacted with

only one residue, Glu 355. This was unexpected because the retaining mechanism proposed to be acting in β -*N*-acetylhexosaminidases would require two acidic residues, one to act as a proton donor and one to act as the nucleophile (Fig 1). In fact, this observation was now in complete support of the substrate assisted catalytic mechanism as described above for *SmChb* and *SpHex*. Because Tews *et al.*, suggested that the *N*-acetamido group on the non-reducing end of the substrate could act as the nucleophile, there was no need for another acid residue within the enzyme active site. Hence, the photoaffinity labeling experiment by Liessem *et al.* appeared to provide physical evidence that the human β -*N*-acetylhexosaminidase used a substrate assisted catalytic mechanism, supporting mechanistic conservation within family 20.

In addition to providing structural evidence in support of the biochemical findings provided by functional analysis of *SpHex* by site directed mutagenesis, the molecular model of the enzyme may be used to provide insight into questions concerning substrate specificity. Comparison of the molecular models of the human β -*N*-acetylhexosaminidase α -subunit and *SpHex* indicates that there has been considerable tertiary conservation between the two proteins (Fig 12). Thus, given the global conservation in tertiary structure between the modeled sequences of the α -subunit of human Hex A and *SpHex*, further structure/function analysis of *SpHex* may help resolve the unanswered questions surrounding why only the α and not the β -subunit of human β -*N*-acetylhexosaminidase can hydrolyze negatively charged substrates such as 4-MUGS.

The structural similarities between the α and β -subunits of human β -*N*-acetylhexosaminidase has allowed for the creation of enzymatically active chimeric α/β -

subunit polypeptides (103; 104). These chimeric polypeptides have indicated that the region of the α -subunit that confers its ability to specifically bind negatively charged substrates corresponds to residues α 132-283 (104). Residues α 225-556 are involved in interacting with the GM2 activator protein (103). Both of the homologous regions in *SpHex* occur within sequence used to build the *SpHex* homology model, save the first 29 residues corresponding to human α -subunit region α 132-283. Thus the *SpHex* model contains the majority of the homologous regions that confer substrate specificity and, surprisingly, GM2 activator binding of the α -subunit of human β -N-acetylhexosaminidase. Hence, the *SpHex* expression/purification system, in combination with the *SpHex* molecular model, may provide a valuable tool for further investigation into how differences in these regions of sequence confer different substrate specificities for the α and β -subunits of human β -N-acetylhexosaminidase.

Although much information can be gained from comparative molecular modeling, there are limits to its usefulness. Because the sequence identity between *SpHex* and *SmChb* was below 30% outside of the sequence encoding the β/α -barrel structure, only the sequence composing the β/α -barrel structure of *SpHex* could be successfully modeled. Thus, 37% of the *SpHex* sequence could not be modeled due to poor sequence identity and thus poor alignment in these areas. This low amino acid sequence identity indicates that evolutionary conservation between *SpHex* and *SmChb* diminishes with increasing distance from the nucleotide sequence encoding the β/α -barrel and the active site residues of the enzymes.

Despite the fact that the amino acid sequences are different between *SpHex* and *SmChb*, the fold of the comparatively modeled protein, i.e. *SpHex*, will be identical to the

crystallographic model of *SmChb* except where the multiple sequence alignment algorithm predicts an amino acid insertion or deletion. Thus, because the sequence to be modeled must conform to the same fold as the sequence of the x-ray structure, comparative modeling is highly successful when sequence identity high (> 70%), for when sequence identity is low (~ 30%) modeled structures tend to contain many structural errors (75). The sequence identity between *SpHex* and *SmChb* in the modeled region (domain III) was 45%; therefore, the structural representation of *SpHex* should be fairly accurate. This is exemplified by the accurate positioning of the key catalytic residues (Glu 314, Arg 162) within the active site of the model (Fig 6 c).

4.2 Functional analysis of *SpHex* by site-directed mutagenesis

The first residues whose role in the catalytic mechanism of *SpHex* was investigated, were the homologous to residues that, when mutated in human β -*N*-acetylhexosaminidase, result in the human B1 variant biochemical phenotype. These residues were chosen because B1 variant mutations were believed to effect catalytically important residues within the active site of the human Hex A α -subunit without affecting its ability to be correctly processed (42). Thus, emulating a B1 mutation in *SpHex* would be expected to change a catalytically important residue.

The first human B1 variant mutation emulated in *SpHex* was α -Arg178His. The analogous substitution in *SpHex* for this human B1 variant mutation is Arg162His. Kinetic analysis of the Arg162His *SpHex* mutant indicated that although the V_{max} was only reduced 5 fold, the K_m was increased 40.5 fold (Table III), indicating that Arg162 was potentially involved in substrate binding. This was the first biochemical evidence indicating that this

conserved Arg residue (Hex A, α -Arg178, β -Arg211 and SmChb Arg349) was acting as a substrate binding residue in a family 20 enzyme. Likewise, the Arg162His mutation in *SpHex* provided direct biochemical support for the crystallographic model of *SmChb* and the *SpHex* homology model. Both the *SpHex* homology model and the *SmChb* crystallographic model indicated that Arg162 (*SpHex*) (Fig 9c) and Arg349 (*SmChb*) (55) form hydrogen bonds with the non-reducing end of the chitobiose substrate.

Contrary to the kinetic data collected from the Arg162His *SpHex* mutant and the physical data provided by the *SmChb* crystallographic model, it has been suggested that the corresponding Arg residues in the α and β -subunits of the human β -*N*-acetylhexosaminidase, α -Arg178 and β -Arg211 respectively, are not involved in substrate binding and fulfil another undetermined role in the active site of human β -*N*-acetylhexosaminidase. This suggestion was made by Brown *et al.* (48; 49) who demonstrated that the human β -subunit mutation Arg211Lys (as opposed to Arg211His) resulted in a mutant Hex B enzyme with measurable catalytic activity and no apparent change in K_m . Thus, it was concluded that Arg 211 of the β -subunit and, by inference, Arg178 of the α -subunit, were somehow involved in the active sites of both subunits but neither was involved in substrate binding.

If the catalytic mechanism of family 20 enzymes is conserved, the role of Arg 178 and Arg 211 in the active site of the α and β -subunits of human β -*N*-acetylhexosaminidase respectively, is that of substrate binding and the conclusions from the work of Brown *et al.* reflect the difficulties in obtaining accurate kinetic data from the human β -*N*-acetylhexosaminidase isoenzymes. Because of the inherent error involved in measuring the extremely low levels activity that were exhibited by the human Hex B containing the β -

Arg211Lys mutation, the mutation may have been conservative enough (resulting in the loss of only one hydrogen bond) not to produce a detectable change in the apparent K_m of the enzyme.

Our studies identified Glu 314 as the catalytic acid residue in *SpHex*. According to the *SpHex* model (Fig 6c) and the Glu314Gln substitution experiment (Table III), it appeared that Glu 314 was acting as the catalytic acid residue, for not only did the model predict that the proton of Glu 314 is with 2.5 Å of the glycosidic linkage of the substrate, mutation of Glu 314 to Gln reduced the V_{max} 270 fold without increasing the K_m . Furthermore, the change in the pH profile of *SpHex* after introduction of the Glu314Gln substitution, suggested that the ionization state of Glu 314 was responsible for creating the acidic pH profile of the wild type enzyme (Fig 17).

Wild type *SpHex* required an acid environment for maximal activity (pH 3.0), suggesting that the acidic environment kept Glu 314 protonated and ready for catalysis. When the Glu314Gln substitution was introduced, the residual catalytic activity that was displayed by *SpHex* insensitive to pH. Only when the pH was less than 3 did the activity of the Glu314Gln mutant change significantly, reducing the activity to almost nothing, probably a result of denaturation of the protein under such acidic conditions (Fig 17). Thus, it appeared that the ionization state of Glu 314 was responsible for the pH profile of the wild type enzyme and this provided further evidence that this residue was serving as the catalytic acid residue in the *SpHex* active site.

SpHex Glu 314 is invariant in all family 20 enzymes. Recent mutagenesis experiments of Glu 323 and Glu 355 in the α -subunit and β -subunits of human β -*N*-acetylhexosaminidase

respectively, further substantiated the suggestion that the conserved *SpHex* residue Glu 314 (*SmChb* 540) was acting as the catalytic acid residue within the family 20 enzymes. Changing either Glu 323 or Glu 355 to a residue that cannot undergo ionization specifically abolished the enzymatic activity of the α and β -subunits without affecting their ability to be correctly processed (58; 59). Thus, the structural and functional data collected from *SpHex* in conjunction with the crystallographic model of *SmChb*, the mutagenesis experiments of Glu 323 or Glu 355 in the α and β -subunits of human β -N-acetylhexosaminidase and the evidence for substrate assisted catalysis in human Hex B, suggests that the catalytic mechanism of family 20 enzymes is conserved.

When the human B1 variant mutation, α -Asp258His, was emulated in *SpHex* (Asp246His), the results indicated that mutation of Asp 246 to Asn did not dramatically effect the catalytic activity of *SpHex* (Table III). Thus, it was concluded that Asp 246 was not directly involved in catalysis and, as the *SpHex* model indicates (fig 9c), Asp 246 appears to indirectly bind substrate through Glu 313 and a potential water molecule (see section 3.9) exactly as is seen for its counterpart, Asp 448, in the *SmChb* x-ray structure.

The structural and functional studies of *SpHex* in conjunction with the *SmChb* crystallographic data and molecular modeling of the α and β -subunits of human β -N-acetylhexosaminidase by Tews *et al.* (55), demonstrate that bacterial enzymes belonging to family 20 O-glycosyl hydrolases can be used to provide insight into the structural and functional relationships of human β -N-acetylhexosaminidase.

5 REFERENCES

1. Horsch, M., Mayer, C., Sennhauser, U. and Rast, D.M. B-N-Acetylhexosaminidase: A target for the design of antifungal agents. *Pharmac.Ther.* 76:187-218, 1997.
2. Henrissat, B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem.J.* 280:309-316, 1991.
3. Henrissat, B. and Bairoch, A. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem.J.* 293:781-788, 1993.
4. Koshland, D.E. Stereochemistry and the mechanism of enzymatic reactions. *Biol.Rev.Camb.Philos.Soc.* 28:416-436, 1953.
5. Sinnott, M.L. Catalytic mechanisms of enzymatic glycosyl transfer. *Chem.rev.* 90:1171-1202, 1995.
6. Davies, G. and Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Curr.Biol.* 3:853-859, 1995.
7. Okada, S. and O'Brien, J.S. Tay-Sachs disease: generalized absence of a beta-D-N-acetylhexosaminidase component. *Science* 165:698-700, 1969.
8. Aronson, N.N., Jr. and Kuranda, M.J. Lysosomal degradation of Asn-linked glycoproteins. *FASEB J.* 3:2615-2622, 1989.

9. Gravel, R.A., Clarke, J.T., Kaback, M.M., Mahuran, D., Sandhoff, K. and Suzuki, K. The GM2 Gangliosidosis. In: *The Metabolic Basis of Inherited Disease*, edited by Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D. New York: McGraw-Hill, Inc., 1994,
10. Svennerholm, L. Gangliosides and synaptic transmission. *Adv.Exp.Med Biol.* 125:533-544, 1980.
11. Triggs-Raine, B.L., Feigenbaum, A.S., Natowicz, M., et al. Screening for carriers of Tay-Sachs disease among Ashkenazi Jews. *New Engl.J.Med.* 323:6-12, 1990.
12. Kaback, M., Lim-Steele, J., Dabholkar, D., Brown, D., Levy, N. and Zeiger, K. Tay-Sachs disease-carrier screening, prenatal diagnosis, and the molecular era. *Journal of the American Medical Association* 270:2307-2315, 1993.
13. Neufeld, E.F. Natural history and inherited disorders of a lysosomal enzyme, β -hexosaminidase. *J.Biol.Chem.* 264:10927-10930, 1989.
14. Proia, R.L. and Soravia, E. Organization of the gene encoding the human β -hexosaminidase α chain. *J.Biol.Chem.* 262:5677-5681, 1987.
15. Nakai, H., Byers, M.G., Nowak, N.J. and Shows, T.B. Assignment of β -hexosaminidase A α -subunit to human chromosomal region 15q23-q24. *Cytogenet.Cell Genet.* 56:164, 1991.
16. Proia, R.L. Gene encoding the human β -hexosaminidase β chain: extensive homology

of intron placement in the α - and β -chain genes. *Proc.Natl.Acad.Sci.USA* 85:1883-1887, 1988.

17. Fox, M.F., DuToit, D.L., Warnich, L. and Reteif, A.E. Regional localization of alpha-galactosidase (GLA) to Xpter—q22, hexsaminidase B (HEXB) to 5q13—qter, and arylsulfatase B (ARSB) to 5pter—q13. *Cytogenet.Cell Genet.* 38:45, 1984.
18. Little, L.E., Lau, M.M., Quon, D.V., Fowler, A.V. and Neufeld, E.F. Proteolytic processing of the alpha-chain of the lysosomal enzyme, beta-hexosaminidase, in normal human fibroblasts. *J Biol.Chem* 263:4288-4292, 1988.
19. Stirling, J., Leung, A., Gravel, R. and Mahuran, D.J. Localization of the pro-sequence within the total deduced primary structure of human β -hexosaminidase B. *FEBS Letters* 231:47-50, 1988.
20. Weitz, G. and Proia, R.L. Analysis of glycosylation and phosphorylation of the α -subunit of the lysosomal enzyme, β -hexosaminidase A, by site-directed mutagenesis. *J.Biol.Chem.* 267:10039-10044, 1992.
21. Sonderfeld-Fresko, S. and Proia, R.L. Analysis of the glycosylation and phosphorylation of the lysosomal enzyme, β -hexosaminidase B, by site-directed mutagenesis. *J.Biol.Chem.* 264:7692-7697, 1989.
22. Sonderfeld-Fresko, S. and Proia, R.L. Synthesis and assembly of a catalytically active

- lysosomal enzyme, β -hexosaminidase B, in a cell-free system. *J.Biol.Chem.* 263:601-607, 1988.
23. Kornfeld, S. Lysosomal enzyme targeting. *Biochem.Soc.Trans.* 18:367-374, 1990.
24. Kornfeld, S. Structure and function of the mannose-6-phosphate/insulin like growth factor II receptors. *Annu.Rev.Biochem.* 61:307-330, 1992.
25. Mahuran, D.J. β -Hexosaminidase: biosynthesis and processing of the normal enzyme, and identification of mutations causing Jewish Tay-Sachs disease. *Clin.Biochem.* 28:101-106, 1995.
26. Hasilik, A., Figura, K., Conzelmann, E., Nehrkorn, H. and Sandhoff, K. Lysosomal enzyme precursors in human fibroblasts. *Eur.J.Biochem.* 125:317-321, 1982.
27. Kytzia, H.J. and Sandhoff, K. Evidence for two different active sites on human β -hexosaminidase A. *J.Biol.Chem.* 260:7568-7572, 1985.
28. Bayleran, J., Hechtman, P. and Saray, W. Synthesis of 4-methylumbelliferyl- β -D-N-acetylglucosamine-6-sulfate and its use in classification of GM2 gangliosidosis genotypes. *Clin.Chim.Acta* 143:73-89, 1984.
29. Tallman, J.F., Brady, R.O., Navon, R. and Padeh, B. Ganglioside catabolism in hexosaminidase A-deficient adults. *Nature* 252:254-255, 1974.
30. Wenger, D.A., Okada, S. and O'Brien, J.S. Studies on the substrate specificity of

- hexosaminidase A and B from liver. *Arch.Biochem Biophys* 153:116-129, 1972.
31. Wiktorowicz, J.E., Awasthi, Y.C., Kurosky, A. and Srivastava, S.K. Purification and properties of human kidney-cortex hexosaminidases A and B. *Biochem J* 165:49-53, 1977.
32. Conzelmann, E. and Sandhoff, K. 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.USA* 75:3979-3983, 1978.
33. Kaback, M.M. Thermal fractionation of serum hexosaminidases: applications to heterozygote detection and diagnosis of Tay-Sachs disease. *Meth.Enzymol.* 28:862-867, 1972.
34. Kaback, M.M., Shapiro, L.J. and Hirsch, P. Tay-Sachs disease heterozygote detection: a quality control study. In: *Tay-Sachs disease: screening and prevention*, edited by Kaback, M.M., Rimoin, D.L. and O'Brien, J.L. New York: A.R. Liss, 1977, p. 267-279.
35. Li, Y.T., Mazzotta, M.Y., Wan, C.C., Orth, R. and Li, S.C. Hydrolysis of Tay-Sachs ganglioside by beta-hexosaminidase A of human liver and urine. *J Biol.Chem* 248:7512-7515, 1973.

36. Furst, W., Schubert, J., Machleidt, W., Meyer, H.E. and Sandhoff, K. The complete amino-acid sequences of human ganglioside GM2 activator protein and cerebroside sulfate activator protein. *Eur.J Biochem* 192:709-714, 1990.
37. Burg, J., Conzelmann, E., Sandhoff, K., Solomon, E. and Swallow, D.M. Mapping of the gene coding for the human GM2 activator protein to chromosome 5. *Ann.Hum.Genet.* 49:41-45, 1985.
38. Hama, Y., Li, Y.T. and Li, S.C. Interaction of GM2 activator protein with glycosphingolipids. *J Biol.Chem* 272:2828-2833, 1997.
39. Conzelmann, E. and Sandhoff, K. Partial enzyme deficiencies: residual activities and the development of neurological disorders. *Dev.Neurosci.* 6:58-71, 1983.
40. Church, W., Swenson, L., James, M.N. and Mahuran, D. Crystallization of human β -hexosaminidase B. *J.Mol.Biol.* 227:577-580, 1992.
41. Ohno, K. and Suzuki, K. Mutation in GM2-gangliosidosis B1 variant. *J.Neurochem* 50:316-318, 1988.
42. Suzuki, K. and Vanier, M.T. Biochemical and molecular aspects of late-onset GM2-gangliosidosis: B1 variant as a prototype. *Dev.Neurosci.* 13:288-294, 1991.
43. Dos, S., Tanaka, A., Sa, M., Ribeiro, M.G., Maia, M. and Suzuki, K. GM2 gangliosidosis B1 variant: analysis of β -hexosaminidase α gene mutations in 11 patients from a defined region in Portugal. *Am.J.Hum.Genet.* 49:886-890,

1991.

44. Goebel, H.H., Stolte, B., Kustermann-Kuhn, B. and Harzer, K. B1 variant of GM2 gangliosidosis in a 12-year-old patient. *Pediatr.Res.* 25:89-93, 1989.
45. Tanaka, A., Ohno, K. and Suzuki, K. GM2-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.* 156:1015-1018, 1988.
46. Tanaka, A., Ohno, K., Sandhoff, K., et al. GM2-gangliosidosis B1 variant: analysis of β -hexosaminidase α gene abnormalities in seven patients. *Am.J.Hum.Genet.* 46:329-339, 1990.
47. Triggs-Raine, B.L., Akerman, B.R., Clarke, J.T. and Gravel, R.A. Sequence of DNA flanking the exons of the *HEXA* gene and identification of mutations in Tay-Sachs disease. *Am.J.Hum.Genet.* 49:1041-1054, 1991.
48. Brown, C.A., Neote, K., Leung, A., Gravel, R.A. and Mahuran, D.J. 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.* 264:21705-21710, 1989.
49. Brown, C.A. and Mahuran, D.J. Active arginine residues in β -hexosaminidase. *J.Biol.Chem.* 266:15855-15862, 1991.

50. Bayleran, J., Hechtman, P., Kolodny, E. and Kaback, M. Tay-Sachs disease with hexosaminidase A: characterization of the defective enzyme in two patients. *Am.J.Hum.Genet.* 41:532-548, 1987.
51. Akli, S., Chelly, J., Lacorte, J.M., Poenaru, L. and Kahn, A. Seven novel Tay-Sachs mutations detected by chemical mismatch cleavage of PCR-amplified cDNA fragments. *Genomics* 11:124-134, 1991.
52. Tanaka, A., Punnett, H.H. and Suzuki, K. A new point mutation in the β -hexosaminidase α subunit gene responsible for infantile Tay-Sachs disease in a non-Jewish Caucasian patient (a Kpn mutant). *Am.J.Hum.Genet.* 47:567-574, 1990.
53. Liessem, B., Glombitza, G.J., Knoll, F., et al. Photoaffinity labeling of human lysosomal β -hexosaminidase B. *J.Biol.Chem.* 270:23693-23699, 1995.
54. Tse, R., Vavougios, G., Hou, Y. and Mahuran, D.J. Identification of an active acidic residue in the catalytic site of β -hexosaminidase. *Biochem.* 35:7599-7607, 1996.
55. Tews, I., Perrakis, A., Oppenheim, A., Dauter, Z., Wilson, K.S. and Vorgias, C.E. Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease. *Nat.Struct.Biol.* 3:638-648, 1996.
56. Tews, I., Vincentelli, R. and Vorgias, C.E. N-acetylglucosaminidase (chitobiase) from

Serratia marcescens: Gene sequence, and protein production and purification in *Escherichia coli*. *Gene* 170:63-67, 1996.

57. Drouillard, S., Armand, S., Davies, G., Vorgias, C.E. and Henrissat, B. *Serratia marcescens* chitobiase is a retaining glycosidase utilizing substrate acetamido group participation. *Biochem.J.* 328:945-949, 1997.
58. Fernandes, M.J., Yew, S., Leclerc, D., et al. Identification of candidate active site residues in lysosomal β -hexosaminidase A. *J.Biol.Chem.* 272:814-820, 1997.
59. Pennybacker, M., Schuette, C.G., Liessem, B., et al. Evidence for the involvement of Glu355 in the catalytic action of human β -hexosaminidase B. *J.Biol.Chem.* 272:8002-8006, 1997.
60. Cao, Z.M., Petroulakis, E., Salo, T. and Triggs-Raine, B.L. Transient expression of β -hexosaminidase pseudodeficiency mutations, R247W and R249W, allow them to be readily differentiated from disease-causing mutations. *J.Biol.Chem.* 272:14975-14982, 1997.
61. Brown, C.A. and Mahuran, D.J. β -Hexosaminidase isozymes from cells cotransfected with α and β cDNA constructs: analysis of the α -subunit missense mutation associated with the adult form of Tay-Sachs disease. *Am.J.Hum.Genet.* 53:497-508, 1993.
62. Trop, I., Kaplan, F., Brown, C., Mahuran, D. and Hechtman, P. A glycine250 to

- aspartate substitution in the α -subunit of hexosaminidase A causes juvenile-onset Tay-Sachs disease in a Lebanese-canadian Family. *Hum.Mutat.* 1:35-39, 1992.
63. Prezant, T.R. Expression of human lysosomal β -hexosaminidase in yeast vacuoles. *Biochem.Biophys.Res.Comm.* 170:383-390, 1990.
64. Boose, J.A., Tiff, C.J., Proia, R.L. and Myerowitz, R. Synthesis of a human lysosomal enzyme, β -hexosaminidase B, using the baculovirus expression system. *Prot.Express.Purif.* 1:111-120, 1990.
65. Robbins, P.W., Overbye, K., Albright, C., Benfield, B. and Pero, J. Cloning and high-level expression of chitinase-encoding gene of *Streptomyces plicatus*. *Gene* 111:69-76, 1992.
66. Robbins, P.W., Albright, C. and Benfield, B. Cloning and expression of a *Streptomyces plicatus* chitinase (chitinase-63) in *Escherichia coli*. *J.Biol.Chem.* 263:443-447, 1988.
67. Sambrook, J., Fritsch, E.F. and Maniatis, T. Molecular Cloning : a Laboratory Manual, 2nd Ed., Cold Spring Harbor, NY, 1989.
68. Nisen, P., Medford, R., Mansour, J., Purucker, M., Skalka, A. and Shapiro, L. Cell-cycle-associated rearrangement of inverted repeat DNA sequences. *Proc.Natl.Acad.Sci.U.S.A.* 76:6240-6244, 1979.

69. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal.Biochem* 72:248-254, 1976.
70. Carter, P. in *Directed Mutagenesis: A practical Approach*, New York:Oxford University Press, 1991. pp. 1-25.
71. Sanger, F., Nicklen, S. and Coulson, A.R. DNA sequencing with chain terminating inhibitors. *Proc.Natl.Acad.Sci.USA* 74:5463-5467, 1977.
72. Higgins, D.G., Bleasby, A.J. and Fuchs, R. CLUSTAL V: improved software for multiple sequence alignment. *Comput.Appl.Biosci.* 8:189-191, 1992.
73. Thompson, J.D., Higgins, D.G. and Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic.Acids.Res.* 22:4673-4680, 1994.
74. Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.A.* 47:110-119, 1991.
75. Mosimann, S., Meleshko, R. and James, M.N. A critical assessment of comparative molecular modeling of tertiary structures of proteins. *Proteins* 23:301-317, 1995.

76. Powell, M.J.D. Restrart procedures for the conjugate gradient method. *Mathematical Programming* 12:241-254, 1977.
77. Brunger, A.T. *XPLOR: a system for X-ray Crstallography and NMR*, New Haven, CT.:Yale University Press, 1993.
78. Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. PROCHECK: A program to check the stereochemical quality of protein structures. *J.Appl.Crystallogr.* 26:283-291, 1993.
79. Kunkel, T.A., Roberts, J.D. and Zakour, R.A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* 154:367-382, 1987.
80. Bebenek, K. and Kunkel, T.A. The use of native T7 DNA polymerase for site-directed mutagenesis. *Nucleic.Acids.Res.* 17:54081989.
81. Saiki, R.K. Amplification of genomic DNA. In: *PCR Protocols*, edited by Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. San Diego: Academic Press, INC, 1990, p. 21-27.
82. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680-685, 1970.
83. Towbin, H., Staechelin, T. and Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications.

84. Henderson, P.J. Statistical analysis of enzyme kinetic data. In: *Enzyme Assays-A practical approach*, edited by Eisenthal, R. and Danson, J. New York: Oxford University Press, 1993, p. 277-313.
85. Lovatt, A. and Roberts, I.S. Cloning and expression in *Escherichia coli* of the *nahA* gene from *Porphyromonas gingivalis* indicates that beta-N-acetylhexosaminidase is an outer-membrane-associated lipoprotein. *Microbiology*. 140:3399-3406, 1994.
86. Soto-Gil, R.W. and Zyskind, J.W. N,N'-diacetylchitobiase of *Vibrio harveyi*. *J.Biol.Chem.* 264:14778-14783, 1989.
87. Tsujibo, H., Fujimoto, K., Tanno, H., et al. Molecular cloning of the gene which encodes beta-N-acetylglucosaminidase from a marine bacterium, *Alteromonas* sp. strain O-7. *Appl.Environ.Microbiol.* 61:804-806, 1995.
88. Somerville, C.C. and Colwell, R.R. Sequence analysis of the β -N-acetylhexosaminidase gene of *Vibrio vulnificus*: evidence for a common evolutionary origin of hexosaminidases. *Proc.Natl.Acad.Sci.USA* 90:6751-6755, 1993.
89. Cannon, R.D., Niimi, K., Jenkinson, H.F. and Shepherd, M.G. Molecular cloning and expression of the *Candida albicans* β -N-acetylglucosaminidase (*HEX1*) gene.

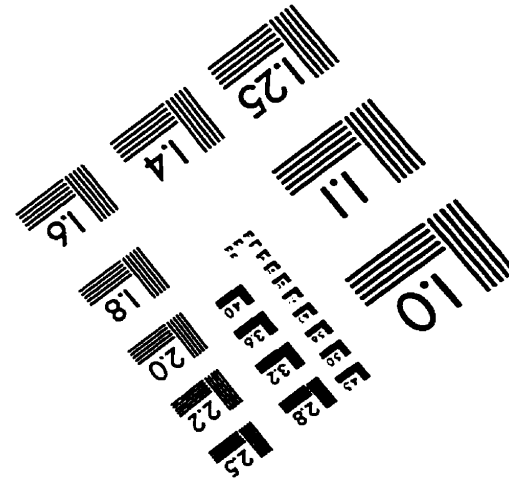
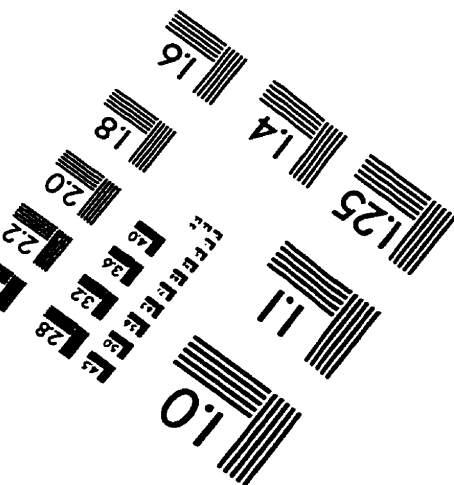
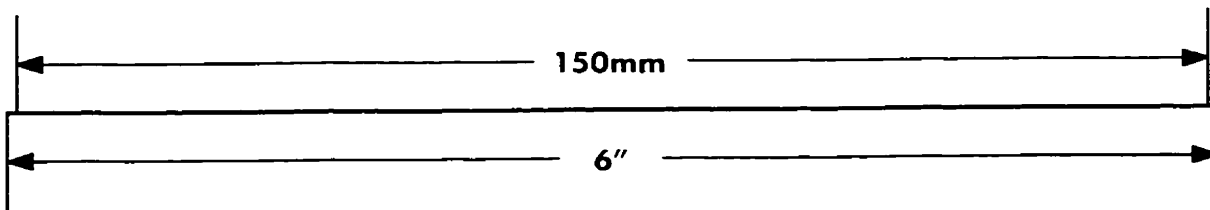
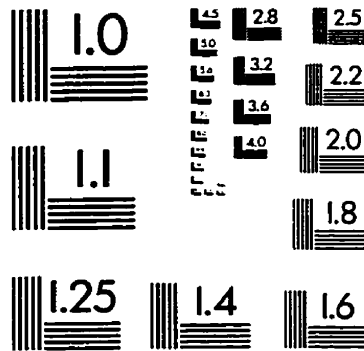
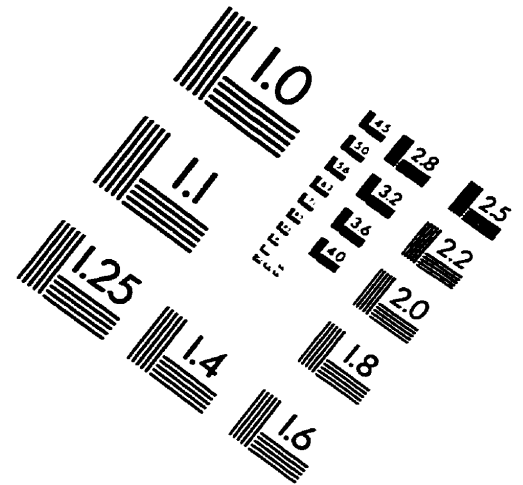
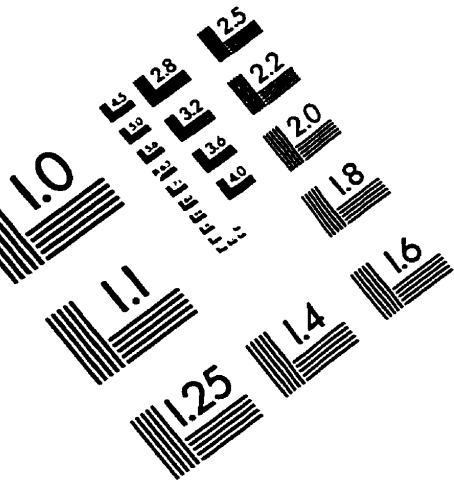
J.Bacteriol. 176:2640-2647, 1994.

90. Muldoon, L.L., Neuwelt, E.A., Pagel, M.A. and Weiss, D.L. Characterization of the molecular defect in a feline model for type II GM2-gangliosidosis (Sandhoff disease). *Am J Pathol* 144:1109-1118, 1994.
91. Beccari, T., Hoade, J., Orlacchio, A. and Stirling, J.L. Cloning and sequence analysis of a cDNA encoding the α -subunit of mouse β -N-acetylhexosaminidase and comparison with the human enzyme. *Biochem.J.* 285:593-596, 1992.
92. Bapat, B., Ethier, M., Neote, K., Mahuran, D. and Gravel, R.A. Cloning and sequence analysis of a cDNA encoding the β -subunit of mouse β -hexosaminidase. *FEBS Letters* 237:191-195, 1988.
93. Myerowitz, R., Piekarz, R., Neufeld, E.F., Shows, T.B. and Suzuki, K. Human β -hexosaminidase α chain: coding sequence and homology with the β chain. *Proc.Natl.Acad.Sci. USA* 82:7830-7834, 1985.
94. O'Dowd, B.F., Quan, F., Willard, H.F., et al. Isolation of cDNA clones coding for the β subunit of human β -hexosaminidase. *Proc.Natl.Acad.Sci. USA* 82:1184-1188, 1985.
95. Graham, T.R., Zassenhaus, H.P. and Kaplan, A. Molecular cloning of the cDNA which encodes β -N-acetylhexosaminidase A from *Dictyostelium discoideum*. *J.Biol.Chem.* 263:16823-16829, 1988.

96. Beanan, M.J. and Bailey, G.B. The primary structure of an *Entamoeba histolytica* beta-hexosaminidase A subunit. *J Eukaryot. Microbiol.* 42:632-636, 1995.
97. Nagamatsu, Y., Yanagisawa, I., Kimoto, M., Okamoto, E. and Koga, D. Purification of a chitooligosaccharidolytic β -N-acetylglucosaminidase from *Bombyx mori* larvae during metamorphosis and the nucleotide sequence of its cDNA. *Bioscience Biotechnology and Biochemistry* 59:219-225, 1995.
98. Engh, R.A. and Huber, R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr.A.* A47:392-400, 1991.
99. Fernandes, M., Kaplan, F., Natowicz, M., et al. A new Tay-Sachs disease B1 allele in exon 7 in two compound heterozygotes each with a second novel mutation. *Hum.Molec.Genet.* 1:759-761, 1992.
100. Farber, G. and Petsko, G.A. The evolution of the alpha/beta-barrel enzymes. *Trends in Biochemical sciences* 15:228-234, 1990.
101. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.-P. and Davies, G. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc.Natl.Acad.Sci.U.S.A.* 92:7090-7094, 1995.
102. Reardon, D. and Farber, G.K. The structure and evolution of alpha/beta barrel proteins. *FASEB J* 9:497-503, 1995.

103. Pennybacker, M., Liessem, B., Moczał, H., Tiff, C.J., Sandhoff, K. and Proia, R.L.
Identification of domains in human β -hexosaminidase that determine substrate specificity. *J.Biol.Chem.* 271:17377-17382, 1996.
104. Tse, R., Wu, Y.J., Vavougios, G., Hou, Y., Hinek, A. and Mahuran, D.J.
Identification of functional domains within the alpha and beta subunits of B-hexosaminidase A through the expression of alpha-beta fusion proteins. *Biochem.* 35:10894-10903, 1996.

IMAGE EVALUATION TEST TARGET (QA-3)



APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved