Homology Modeling of Glycosyl Hydrolase Family 18 Enzymes and Proteins

Nathan N. Aronson,† Christopher J. Blanchard,‡ and Jeffry D. Madura*,‡

Department of Biochemistry and Molecular Biology, College of Medicine, and Department of Chemistry, College of Arts and Sciences, University of South Alabama, Mobile, Alabama 36688

Received June 17, 1997⊗

Using state-of-the-art homology modeling methods, three-dimensional coordinates for three family 18 glycosyl hydrolases were determined. The structures for Gp39, Brp39, and chitotriosidase were computer determined using the X-ray coordinates from *Sm*ChiA. The results of the modeling efforts are assessed, and comparison of the modeled structures to other known family 18 members is made.

INTRODUCTION

Chitin is a homopolysaccharide that consists of repeated *N*-acetyl-β-(1→4)-D-glucosamine (Figure 1). It is extremely abundant in nature, as it is a major structural component in the exoskeleton of insects and crustaceans. Chitin is also present in mollusca, coelenterata, nematodes, protozoa, and some fungi. Next to cellulose, chitin is the most abundant polysaccharide in the biosphere. Both polysaccharides virtually form "molecular boards". They are long and flat, and this conformation is likely why nature has put them to extensive architectural use. This shape allows chitin and cellulose chains to pack closely and form exceptionally stable bundles that are held together by interchain H-bonding. Annual biological production of chitin is estimated to be 2.5 billion tons.[3]

Hydrolysis of chitin must also be a major metabolic process throughout the biosphere in order to balance the intense level of its biosynthesis. Thus, massive amounts of the structural polysaccharide are deposited on the ocean floor annually, but there is relatively little chitin in marine sediments. This turnover of chitin is mainly accomplished by marine bacteria using three types of polysaccharide hydrolases (glycosidases).[1] Each type of chitinase has a substrate specificity for the *N*-acetyl-β-(1→4)-D-glucosamine bond, but they cleave at different locations along the chain (Figure 1); **[1] endochitinases** cleave internal bonds to create new substrate ends for two exochitinases: **[2] chitodextrinases** and **[3] *N*-acetyl-β-D-hexosaminidases**. Chitodextrinase processively hydrolyzes disaccharides from the nonreducing end of soluble chitin oligosaccharides, while hexosaminidase splits the chitin disaccharides as well as hydrolyzing monosaccharides from the nonreducing end of soluble oligomers. The set of three glycosidases efficiently converts chitin to free *N*-acetylglucosamine (GlcNAc) a biological source of both carbon and nitrogen especially important in the marine environment.

Glycosidases, the general class of enzymes that hydrolyze carbohydrate polymers, have been grouped into over 50 families based on their amino acid sequence homologies.[4−6] Chitinases fall into family 18 and 19, while exo-hexosaminidases such *N*-acetyl-β-D-hexosaminidases are in a separate class (family 20). We noted important sequence homologies between family 18 chitinases (mostly chitodextrinases) and two other types of related proteins (Table 1): [1] chitobiase, a lysosomal glycosidase that cleaves carbohydrates uniquely from the reducing-end, thereby removing either a single GlcNAc from oligosaccharides that are normally attached to asparagine-linked glycoproteins[7,8] or GlcNAc processively from chitin oligosaccharides and [2] chitinase-related and oviduct-specific proteins that no longer retain hydrolytic activity towards chitin or its fragments.

Glycosidases share important similarities in their reaction mechanisms, but there are variations that occur. Hen's egg white lysozyme was the first enzyme whose crystal structure was determined,[9] and it can act as a chitinase as well. Based on hen's egg white lysozyme, two mechanisms lead to either retention or inversion of the anomeric glycosyl bond being hydrolyzed.[10,11] However, the acid−base hydrolysis by lysozyme which employs two acidic amino acids within the substrate-binding groove of the active-site can vary for evolutionary forms of the enzyme that have only one acid group.[12]

One explanation for the similar reaction mechanism of many glycosidases is the discovery of a shared peptide folding pattern of the active-site domain for a large group of glycosidases including chitinase A from *S. marcescens* (*Sm*ChiA).[3] The general protein folding motif of these enzymes is the $(\beta/\alpha)_8$ TIM-barrel[13] which consists of a cylindrical, barrel-like framework made from eight internal parallel β-strands that are alternately connected by eight exterior α-helices. In the glycosidases this structure appropriately places in the substrate-binding groove two critical acidic residues at the correct distance above and below the sissle glycosidic bond. Like all TIM-barrel enzymes, these two catalytic residues are located just beyond the carboxyl-ends of two of the eight β-strands making the $(\beta/\alpha)_8$ barrel. Thus, evolution has led to a wide diversity of substrate specificity for hydrolysis of varied carbohydrate structures by using a few basic protein folds to provide a required disposition of identical catalytic residues.[11] For example, chitobiase is unique in that it hydrolyzes a reducing-end GlcNAc[7] while in contrast *Sm*ChiA processively cleaves disaccharides from the non-reducing ends of soluble chitin oligosaccharides,[3] a mechanism common to many exo-chitinases.

The chitinase-related proteins in Table 1 are especially intriguing relatives of family 18 chitinases. Homology analysis of sequences from all proteins in Table 1 indicates

* To whom correspondence is addressed.
† College of Medicine.
‡ College of Arts and Sciences.
⊗ Abstract published in *Advance ACS Abstracts,* November 1, 1997.

S0095-2338(97)00236-9 CCC: $14.00

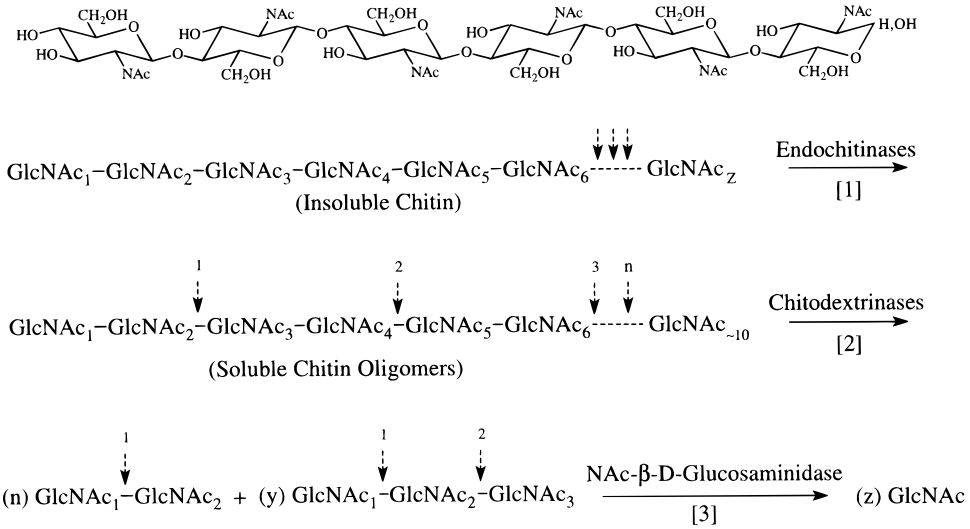


Figure 1. Degradation of chitin by bacterial enzymes.[1] A chitin hexasaccharide is depicted at the top. Each *N*-acetylglucosamine (GlcNAc) is linked β-1→4 and alternately rotated by 180°. [1] Endochitinase cleaves insoluble chitin randomly at internal bonds to release soluble oligomers of approximately 10 residues. [2] Chitodextrinase (e.g., *S. marcescens* chitinase A) progressively cleaves disaccharides from the nonreducing end of the soluble oligosaccharides with some remaining trisaccharide product ($n \gg y$). [3] N-Acetyl-β-D-glucosaminidase digests these di- and trisaccharides at their nonreducing ends to GlcNAc.

Table 1. Sequence Homologies in (I) Family 18 Chitinases; (II) Chitobiases; and (III) Chitinase-Related/Oviduct-Specific Proteins[a]

ENZYME/PROTEIN	β3	β4	β5	β6	β7	β8	GenBank
I. Chitinases							
S. marcescens A	267KILPSIGGWT	308DGVDIDWE	358ELTSAISAG	383DHIFLMSYD	436KIVVGTAMYGRGWT	535GLFSW	Z36294
A. caviae	267KILPSVGGWT	308DGVDIDWE	358ELTSAISAG	383DHIFLMSYD	436KIVVGAAMYGRGWT	535GLFAW	U09139
A. sp. strain0-7	265KILPSVGGWT	306DGVDIDWE	357ELTSAIGAG	382DYIFAMTYD	451KLVMGVAMYGRGWE	564GLFGW	D13762
B. circulans	156KTIISVGGWT	197DGVDLDWE	247LLTIASGAS	272DWINIMTYD	330KLVLGVPFYGRGWD	429GAMFW	M57601
S. plicatus	334KILYSFGGWT	376DGIDLDWE	417LITAAVTAD	445DWYNVMTYD	499KLLIGIGFYGRGWT	586GAFFW	M82804
S. lividans	333KILYSFGGWT	376DGIDLDWE	417LITAAVTAD	445DWYNVMTYD	498KLLIGIGFYGRGWT	585GAFFW	D12647
S. thermoviolaceus	141KVLWSFGGWT	183DG--IDWE	220LVTAAVTAD	248DWYNVMTYD	302KLLLGIGFYGRGWT	389GAFVW	D14536
S. marcescens B	89RIMFSIGGWY	137DGVDIDWE	180QLTIAGAGG	207DYINLMTYD	284KIVMGVPFYGRAFK	399GVMFW	X15208
J. lividum	360KLFISLGGWS	417DGIDIDWE	467LLTAAVGAG	492DWINLMTYD	548KLLLGIPFYGRGWT	643GAFSW	U07025
R. oligosporus	105KVSLSIGGYT	146DGLDIDWE	191LLTVAVPCG	216DLFYLMAYD	262KLVVGMPLYGRGFC	351GAMFW	D87894
S. cerevisiae	175KVIMSIGGWS	216DGIDLDWE	263QLSIAAPAF	288DYWNMMTYD	336KLVLGMAAYGRSFH	442GGGFW	U28373
K. lactis	441KKIPSFGGWD	488DGIDLDWE	532TLSIAIPSS	557DYMVYMTYD	604KVFGGVANYGRSYK	703GTSLW	X07127
A. californica	257KILPSIGGWT	298DGVDIDWE	349ELTSAISAG	374GKIFLMSYD	427KIIVGVAMYGRGWT	524GLFAW	L22858
B. mori	113KILPSIGGWT	154DGVDVDWE	205ELTSAISAG	230DKIFLMTYD	283KIIVGVAMYGRGWT	374GLFAW	L33180
Phycodnaviridae PBCV-1	102NMHASIGGWS	145NSISLDWE	194MCTIAAPEK	215DEVHVMTYD	269KIFVGVAFYSRGFS	359GILVW	U42580
N. tabacum	92KTFLSIAGGR	134HGLDLDWE	176LLTAAVSYS	202NWINLMAYD	252KLVLGIPFYGYAWR	348GYFAW	X78325
C. thermocellum	115KTLISVGGWT	156DGVDIDWE	206LLTIAAPAG	231DFINIMTYD	288KINVGVPYYAAGWQ	381GIMIW	Z68924
B. malayi	99KVLLSYGGYN	141DGFDLDWE	181LLTAAVSAG	207DLLLFLMSYD	261KIIIGIPMYAQGWT	358GAFIW	M73689
B. pahangi	35KVLLSYGGYN	77DGFDLDWE	117LLTAAVSAG	143DLLLFLMSYD	197KIIIGIPMYAQGWT	294GAFIW	U59690
E. dispar	271KVLASIGGWN	318DGIDIDWE	365LLTIAAPAG	390DWINLMTYD	437KIMLGMAHYGRGWT	533GVMFW	U78318
E. histolytica	220KVLASIGGWN	267DGIDIDWE	314LLTIAAPAG	339DWINLMTYD	386KIMLGMAHYGRGWT	482GVMFW	U78319
E. invadens	226QVLASIGGWN	273DGIDIDWE	321LLTIAAPAG	346DWINLMTYD	393KMFVGMAHYGRGWT	489GAMIW	U78320
A. album	123KVMLSIGGWT	164DGIDIDWE	207LLSIAAPAG	232DYINLMAYD	285KIVLGMPIYGRSFQ	374GTMFW	X64104
T. harzianum	124KVMLSIGGWT	165DGIDVDWE	208LLSIAAPAG	233DYVNLMAYD	286KIVLGMPIYGRAFE	375GSMFW	L14614
T. hamatum	124KVMLSIGGWT	165DGIDVDWE	210LLSIAAPAG	235DYINLMAYD	290LIVLGMPIYWRSFE	379GSMFW	Z71415
A. viteae	98KILLSYGGYN	140DGFDLDWE	180LLTAAVSAG	206DLLFLMSYD	260KIIIGIPAYSRGWT	357GAFMW	L42010
A. nidulans	108KVLLSIGGWT	149DGIDIDWE	188QLTVAVPAG	213DFYNLMAYD	267KIILGMPIYGRAFQ	356GAMWW	D87063
O. volvulus	98KILLSTGGYN	140DGFDFDWE	180LLTAAVSAG	206DLLSLMSYD	260KIIIGIPTYGRGWT	357GAFIW	L42021
M. sexta	97KFMVAVGGWA	139DGLDLDWE	182ELTAAVPLA	208DAIHVMSYD	262KLVVGIPFYGRSFT	366GAMTW	P36362
Chelonus sp	96KIMVAVGGWN	138DGFDIDWE	177ILSAAVAAP	203DFINLMTYD	257KLVVGVPAYGKSFT	355GAMVW	U10422
K. zopfii	156KTIISVGGWT	197DGVDLDWE	247LLTIASGAS	272DWINIMTYD	330KLVLGVPFYGRGWD	429GAMFW	D63702
V. furnissii	267KILPSVGGWT	308DGVDIDWE	358ELTSAISAG	383DYIFLMSYD	436KVVVGAAMYGRGWT	535GLFAW	L42548
P. japonicus	116KTNIAVGGWA	158DGLDLDWE	201EITCAVPVA	227DAIHLMTYD	282KLVVGTPFYGRTYT	384GAMTW	D84250
C. elegans	102KLLYAVGGWE	143DGVDIDWE	191LISFAGAAG	217DFVNVMSYD	274KINMGVPFYGRFWK	374GVMIW	Z66524
Coccidioides immitis	123KTLLSIGGWT	164DGIDIDWE	207LLTIASPAG	232DFWNLMAYD	285KIFLGMPLYGRAFA	373GGGMW	L41663
Chitotriosidase (Human)	91KTLLAIGGWN	133DGLDLDWE	179LLSAAVOAG	205DFVNLMAYD	259KLILGMPTYGRSFT	354GAMVW	U29615
II. Chitobiase							
Human	105KGDVSLKDII	136DGINIDIE	175QVTFDVAWS	202DFLFVMSYD	246KLVMGVPWYGYDYT	350GIGMW	M95767
Rat	90KGDVALKDII	121DGINIDIE	160QVTFDVAWS	187DFLFVMSYD	231KLVMGIPWYGYDYI	335GIGMW	M95768
III. Chitinase-Related/Oviduct-Specific(OSG) Proteins							
Human gp39	91KTLLSVGGWN	133DGLDLAWL	173LLSGAVSAG	199DFISIMTYD	253KLVMGIPTFGRSFT	348GAMVW	M80927
Human YKL-39	90KILLSIGGYL	133DGLDVSWI	174LLTAGVSAG	200DFINLLSFD	256KVVMGIPTYGHSFT	351GAMIW	U49835
Mouse gp39	92KTLLAIGGWK	134DGLNLDWQ	180LLTSTGAGI	206DYIQVMTYD	260KLIVGFPAYGHTFI	357GAVVW	S27879
Mouse brp-39	92KTLLSVGGWK	134DGLDLAWL	174LLSAALSAG	200DFINLMTYD	254KLLMGIPTFGKSFT	349GAMVW	X93035
Mouse ECF-L	91KTLLAIGGWK	133DGLNLDWQ	179LLTSTGAGI	205DYIQVMTYD	259KLIVGFPAYGHTFI	356GAVVW	D87757
Pig gp38	91KTLLSVGGWN	133DGLDLAWI	173LLSGAVSAG	199DFISLLTYD	253KLVMGIPTFGRSFT	348GAMVW	U19900
Drosophila secreted gp	107KILLSVGGDK	158DGLDVAWQ	205QFTALLRDV	231NVNSSLFYD	301KINVGVATYGRPWK	395GIWSF	U13825
Human OSG	92KTLLSIGGWN	134DGLDLFFL	180LLSAAVSGV	206DFINVLSYD	255KLIMGIPTYGRTFR	351GAMVW	U09550
Baboon OSG	92KTLLSIGGWN	134DGLDLFFL	180LLSAAVSGV	206DFINVLSYD	255KLIMGIPTYGRTFR	351GAMVW	M59903
Mouse OSG	92KTLLSIGGWN	134DGLDLFFL	180LLSAAVSGI	206DFINVKSYD	255KLIMGFPTYGRNFY	351GAMVW	D32137
Hamster OSG	71KTLLSVGGWN	113DGLGLFFL	159LLSAAVSGI	185DFINVLSYD	234KLLMGFPAYGRTFH	330GAMVW	U15048
Bovine OSG	89KTLLSIGGWN	131DGLDLFFL	177LLSAAVSGD	203DFISVLSYD	252KLLMGLPTYGRTFH	348GAMVW	D16639
Ovine OSG	83KTLLSVGGWN	125DGLDLFFL	171LLSAAVSGD	197DFISVLSYD	246KLLMGLPTYGRTFH	342GAMVW	U17988

[a] The crystal structure of *S. marcescens* chitinase A has been determined to be a $(\beta/\alpha)_8$ barrel protein.[3] The sequences represent strands $\beta3-\beta8$ and a few adjacent amino acids. The sequences have been translated from GenBank (accession numbers indicated). The starting amino acid position of each protein segment is numbered based on the translational start methionine being residue 1.

that a common progenitor chitinase likely gave rise to these groups of proteins. However, all members of these chitinase-related groups are not active chitinases because they no longer possess the critical glutamic acid H^+ donor at the carboxyl end of the fourth β-strand (Table 1). A logical function for these chitinase-related and oviduct-specific proteins is they retain a binding activity for chitin-like carbohydrate polymers which is important for various

developmental or remodeling processes. For example, the oviduct-specific glycoprotein appears over a few days shortly after fertilization. This protein is expressed under the influence of estrogen and its proposed role is involvement in fertilization or early cell division by the fertilized egg or its implantation in the uterus.[14] Human Gp39 is a major 39kDa glycoprotein secreted by articular chondrocytes and synovial fibroblasts.[15] It is induced in stress situations such as rheumatoid arthritis. Mouse Brp39, a "breast regression protein", is induced in mammary epithelial cells a few days after weaning.[16] Mammary involution involves programmed cell death, and possibly Brp39 utilizes a chitin oligosaccharide binding ability while participating in the various signal transduction pathways that lead to apoptosis of the regressing cells. Besides the unique reducing-end chitobiase, a chitotriosidase[17−19] is the only other family 18 enzyme in humans that retains catalytic activity. This 39 kDa endochitinase is produced by macrophages and is found elevated in the plasma of patients with Gaucher disease.

The major goal of our research is to determine the structural biology of human lysosomal chitobiase and several of these chitinase-related proteins by theoretical computer modeling. The crystal structure of a chitinase from the bacteria *Serratia marcescens* (*Sm*ChiA) is known,[3] and this enzyme is a paradigm for the modeling experiments. Computer modeling of the proteins and enzymes in Table 1 should allow us to better predict their biological function and mechanism of reaction.

ALIGNMENT

All of the 52 family 18 chitinase and chitinase-related proteins in Table 1 have a larger mass than hevamine, concanavalin B, narbonin, endo F_1, and endo H, a separate group of this glycosidase family whose crystal structures have been determined and recently compared.[20] The number of amino acids comprising this latter group of smaller family 18 proteins is 273, 299, 290, 289, and 271, respectively. In contrast *S. marcescens* chitinase A has 406 residues in its chitinase domain (Table 1) which excludes an extra N-terminal domain not considered to be involved in its activity.[3] The major structural difference possessed by all family 18 proteins shown in Table 1 is an extra domain between the seventh and eighth β-strands. For all 52 proteins the number of residues making up this β7 and β8 insertion domain is in the very narrow range 89 ± 6 amino acids. As observed in the *Sm*ChiA crystal structure this domain consists of a five-stranded antiparallel β-sheet and two helices [one being α7 of the $(\beta/\alpha)_8$ barrel]. This region of the protein appears to make up one side of the substrate binding cleft.[20] In hevamine, concanavalin B, narbonin, endo F_1, and endo H the number of residues between β7 and β8 ranges between only 21 and 31 amino acids with this sequence comprising just α7 and a small amount of loop or random coil structure. The function and importance of the extra β7−β8 insertion domain in the 52 proteins listed in Table 1 needs to be characterized.

With few exceptions (e.g., *Aeromonas caviae* and *Aeromonas Sp*) most of the group of family 18 chitinase proteins in Table 1 differ from *Sm*ChiA at their N-terminal region by lacking about 40 amino acids in a loop structure between β2 and α2 (residues 199−240). This missing stretch of amino acids in the sequences of the three proteins that we have modeled against *Sm*ChiA (Gp39, Brp39, and chitotriosidase) prevented the *Gap* program from the Wisconsin Sequence Analysis Package (Genetic Computer Group, Madison WI) from aligning this segment correctly and required us to adjust the computer prediction by hand (Table 2). Thus β2 was easily recognized in all of these proteins by eye but was not aligned to *Sm*ChiA by the *Gap* program. The role of this extra loop structure in *Sm*ChiA and the other chitinases that possess it is not known.

HOMOLOGY MODELING

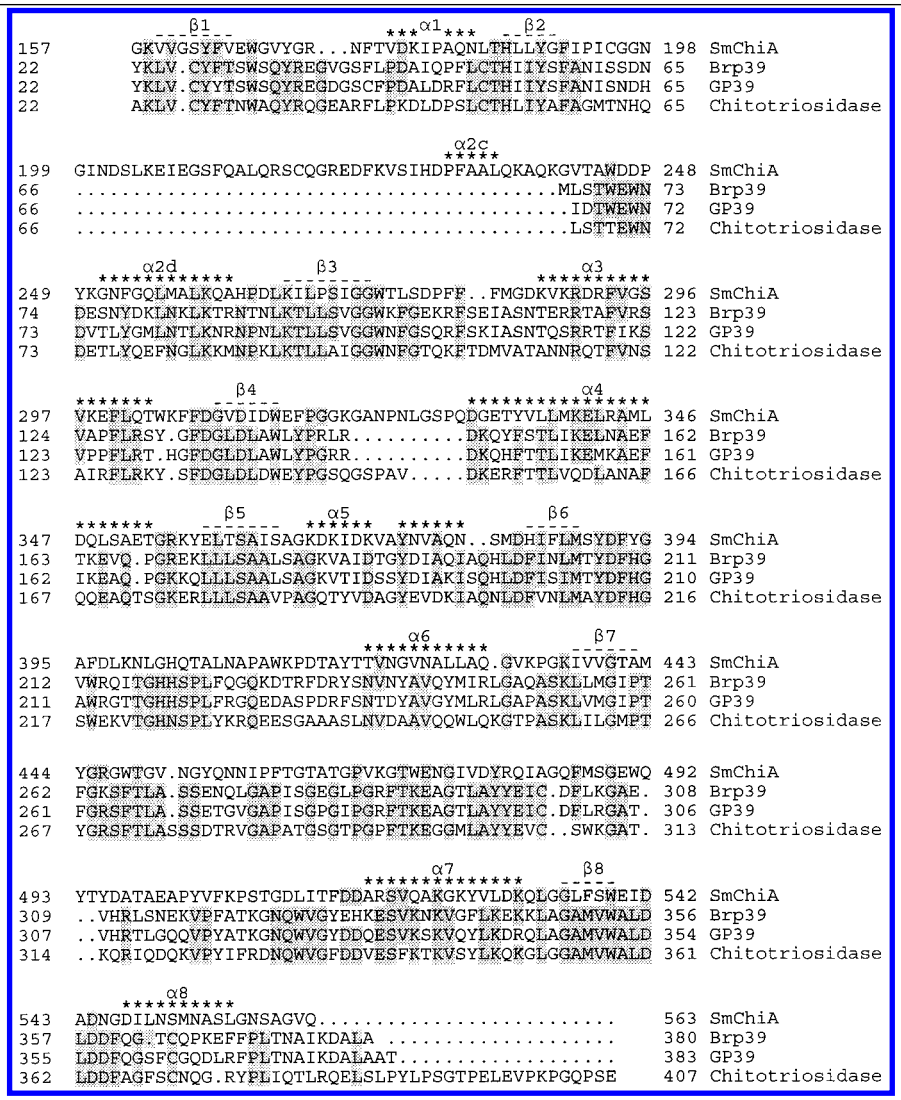
Once the alignment of Gp39, Brp39, and chitotriosidase with *Sm*ChiA had been completed, building of three-dimensional structures for these macromolecules was undertaken. All homology modeling was performed with the program LOOK which uses a method developed by M. Levitt[21] and is referred to as segment match modeling. The Look module SegMod takes the alignment data (Table 2) and automatically generates the three-dimensional structure using this segment match modeling. SegMod works by first breaking the target structure sequence into a set of short segments. Each segment is constructed by randomly choosing a residue which is missing at least one atom and then choosing a sufficient number of residues before and after the randomly chosen one to ensure that there are sufficient constraints to guide the matching process. Next a database of highly refined known protein X-ray structures is searched for matching segments which are then fitted onto the framework of the target structure. A matching database segment is chosen based on three criteria: The first is amino acid similarity (i.e., sequence), second is conformational similarity (i.e., structural), and third is compatibility with the target structure. In the conformational similarity evaluation the atomic coordinates of Cα coordinates for the reference structure are used, and in the compatibility evaluation van der Waals' interactions are checked. The reference coordinates are then used to update the target structure coordinates. This process is repeated until all unknown coordinates are determined. Ten independent models are generated using this process and then averaged to obtain one final structure. This final structure is energy minimized to relieve all strain which may arise during the coordinate building and averaging phases. All of the homology modeling and energy minimizations in this paper were performed using the LOOK 2.0 software from Molecular Applications Group (Palo Alto, CA).

DISCUSSION

The homology modeled structures obtained from model building are illustrated in Figures 2 and 3. Each of the macromolecules is represented as a Cα trace in both figures. Highlighted in color is the region in which there is a large "gap" of approximately 42 residues between *Sm*ChiA and the target proteins (see Table 2). The "active" site residues D315/E391 (based on *Sm*ChiA numbering) are drawn as a ball-and-stick. Figure 2 is a side view of the three macromolecules superimposed upon the reference structure *Sm*ChiA looking into the active site, while Figure 3 is an orthogonal view of Figure 2, looking from the top of the β-barrel. All of the structures are very similar, which they should be due to the high degree of sequence identity between them (see Table 3).

The first striking feature of Figure 2 is the lack of structural disruption due to the gap in the target structures as opposed

Table 2. Alignment of Protein Sequences of *S. marcescens* Chitinase A, Brp39, Gp39, and Chitotriosidase[a]



[a] The amino acid sequences of Brp39, Gp39, and chitotriosidase have been aligned to the sequence of *S. marcescens* chitinase A using the alignment program *Gap* from Genetics Computer Group, Inc. and improved by hand alignment of the region between the N-terminus and strand β3. Protein sequences are translations from GenBank (Z36294, X93035, M80927, and U29615, respectively) and are numbered from the initiation methionine as 1. The signal peptides are not shown, and only the chitinase catalytic domain of *Sm*ChiA is used as the model. Residues are shaded where three or more of the same amino acid occur at the same position. The secondary structures making the (β/α) framework of *Sm*ChiA as observed in its crystal structure[3] are indicated above its sequence: - - (β strands) and *** (α-helices).

to *Sm*ChiA. Starting from the end of β2 to the beginning of α2 there are approximately nine residues in the target proteins which are evidently a sufficient number of residues to join these two highly conserved regions and not cause significant structural disruption. This region in *Sm*ChiA forms part of the entrance to the active site. The model-built residues joining β2 and α2 for the target proteins partially maintain the entrance to the active site.

In both figures the two proposed active residues for all the modeled structures are represented as ball-and-sticks and fall into approximately the same locations as that found in *Sm*ChiA. It has been proposed that residue D313 of *Sm*ChiA is the proton donor to residue E315 and is therefore necessary for catalytic activity.[20,22] The residue D138 of chitotriosidase, corresponding to residue D313 of *Sm*ChiA, points away from the catalytic residue E140, and they are approximately 6 Å apart (Figure 4). Based on modeling efforts using Sculpt 2.5[23] we observed that these two groups can come within hydrogen bonding distance in order to exchange a hydrogen. In Brp39 and Gp39 the two residues in the equivalent position of D313/E315 of *Sm*ChiA are an A/L pair where L has replaced the catalytic residue E (Table 2). Both Brp39 and Gp39 are inactive proteins, while chitotriosidase is an active enzyme. Therefore it is not clear from a comparison of these three proteins whether this aspartic acid residue is necessary for catalytic activity. A point of caution is that this argument is based on a static X-ray picture. D138 of chitotriosidase points into the interior of the β-barrel and could possibly move close enough to E140 and participate in a proton shuttle between the two acids. Potential interaction between D138 and E140 could possibly be observed during a molecular dynamics calculation.

The currently accepted hypothesis on family 18 activity states that hydrolysis occurs via an anchimeric effect[20,22] (see: http://www.expasy.ch/cgi-bin/lists?glycosid.txt). In this mechanism, the substrate assists in the hydrolysis by forming a stabilizing oxazolinium intermediate. However we believe that both E315 and D391 are necessary for activity as suggested by Perrakis et al.[3] Therefore the loss of activity for Gp39 and Brp39 is due to the replacement of the glutamic

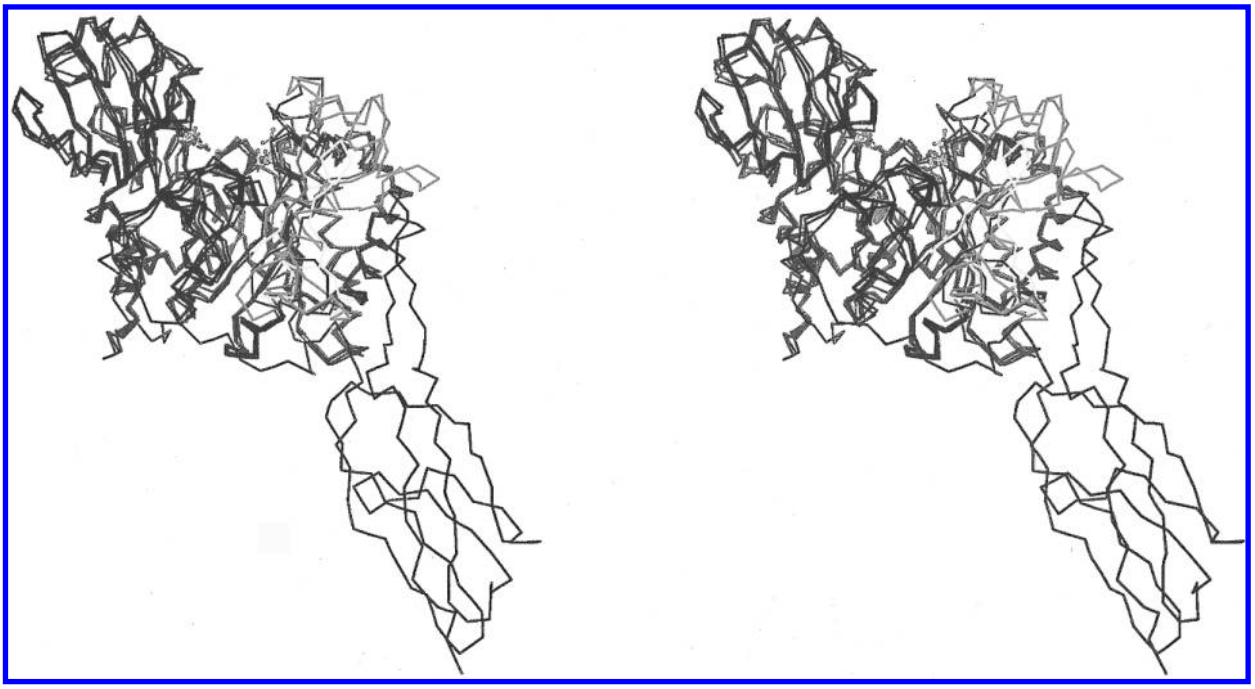


Figure 2. C_α stereoview of Gp39, Brp39, chitotriosidase superimposed with *Sm*ChiA. This is a side view of the TIM barrel. Residues 189−259 of *Sm*ChiA are colored red. Residues 56−81 of Gp39, Brp39, and chitotriosidase are colored yellow, green, and blue, respectively.

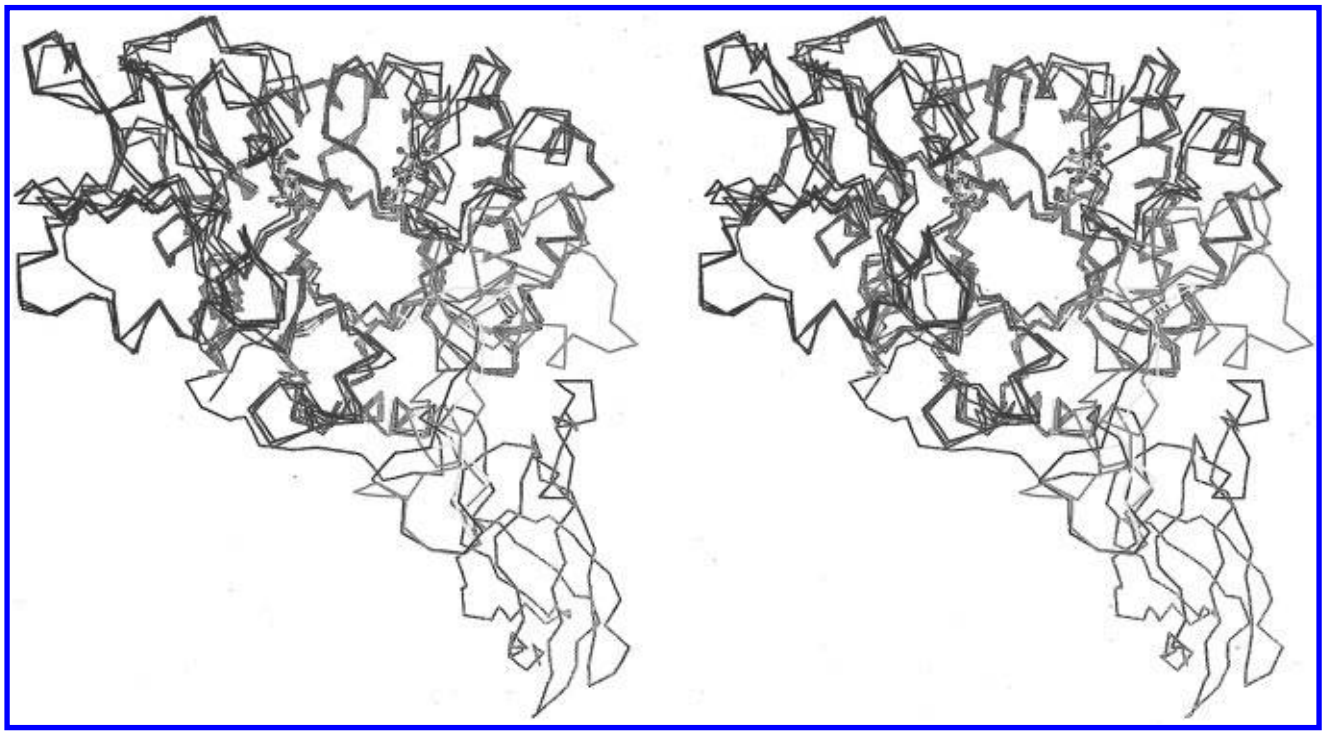


Figure 3. C_α stereoview of Gp39, Brp39, chitotriosidase superimposed with *Sm*ChiA. This is a top view of the TIM barrel. β1/α1 are at the six o'clock position while β5/α5 are near the 12 o'clock position. The β/α numbering proceeds counterclockwise. Residue coloring scheme is the same as that in Figure 2.

acid residue by a leucine residue. The precise mechanistic role of the β4 aspartate two residues in front of the proton-donor glutamate versus that of the conserved aspartate across the barrel at the end of β6 needs to be explained.

In *Sm*ChiA, narbonin, and concanavalin B, there are non-proline *cis*-peptide bonds.[20] In all three of the modeled structures non-proline *cis*-peptide bonds were not constructed. This is most likely due to a deficiency in LOOK.

The active site of *Sm*ChiA has a "wall" consisting of the β5/α5 loop which is part of a three-sided binding cleft. The β5/α5 loop in the three modeled proteins is present and presumably obstructs a bound substrate from extending beyond the +2 binding subsite.[24]

In a recent paper by Terwisscha et al.[20] they make several observations in their comparison of *Sm*ChiA with several other family 18 proteins. In particular they point out that phenylalanine at the end of β2 and the tryptophan at the end of β8 are near the front of the active site and may participate in substrate binding. These residues, which are conserved in all of the proteins modeled in this paper are found near the front of the active site and opposite each other. They also make note of the aspartic acid preceeding β4 and the lysine at the beginning of β3. These two residues, again highly conserved, form a salt bridge in the modeled proteins and may be there to assist in their folding. The conserved GG pair of β3 is close to the phenylalanine of β2 in all of

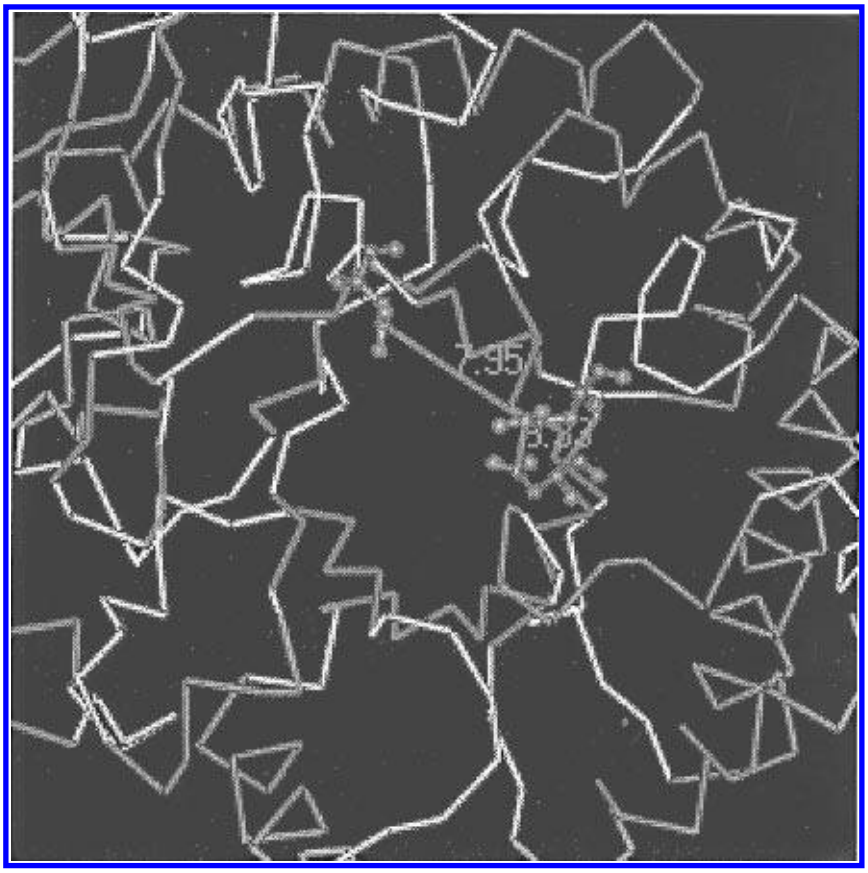


Figure 4. Ilustrated is the Cα trace of the active site for chitotriosidase looking down the $(\beta/\alpha)_8$ barrel. Residues D138, E140, and D213 are displayed as ball and stick models. D213 (lone residue on the left side of the active site) is 7.9 Å from residue E140.

Table 3. Percentage of Identical Amino Acid Residues in the Sequences of *Sm*ChiA, Brp39, Gp39, and Chitotriosidase[a]

	*Sm*ChiA	Chitotriosidase	Brp39	Gp39
***Sm*ChiA**	100	28.7	26.0	28.7
Chitotriosidase		100	51.7	54.2
Brp39			100	73.8
Gp39				100

[a] Identical residues were found using the sequence comparison program *Gap* from Genetics Computer, Group Inc.

our modeled structures. Terwisscha et al.[20] suggest this GG pair is necessary to maintain the β-barrel due to the crowding in this region. All the other observations made by Terwisscha et al.[20] are consistent with our modeled structures.

We have performed a stereochemical and structural analysis on our modeled structures using the Procheck Analysis program.[25] According to the Procheck analysis, our current model structures have approximately 79% of their residues in the most favored region of the Ramachandran plot. As stated near the bottom of the Ramachandran plot generated by Procheck "Based on analysis of 118 structures of resolution at least 2.0 Å and *R*-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favored regions".[3] Therefore these models are average at best. Further refinement is necessary before one can proceed to use these models in any binding studies. We are in the process of doing molecular mechanics and dynamics of these models to improve their quality. Once this has been accomplished we plan to bind substrates to these proteins and study both the binding and activity.

Overall the structures obtained from homology building appear to have similar features as other members of family 18 glycosyl hydrolases. These similarities and the results from the Procheck analysis provide some confidence that the structures built using *Sm*ChiA are reliable for further study until experimental coordinates become available.

REFERENCES AND NOTES

(1) Bassler, B. L.; Yu, C.; Lee, Y. C.; Roseman, S. Chitin Utilization by Marine Bacteria. *J. Biol. Chem.* **1991**, *226*, 24276−24286.
(2) Jeuniaux, C.; Voss-Foucart, M. F. Chitin Biomass and Production in the Marine Environment. *Biochem. Syst. Ecol.* **1991**, *19*, 347−356.
(3) Perrakis, A.; Tews, I.; Dauter, Z.; Oppenheim, A. B.; Chet, I.; Wilson, K. S.; Vorgias, C. E. Crystal Structure of a Bacterial Chitinase at 2.3Å Resolution. *Structure* **1994**, *12*, 1159−1180.
(4) Henrissat, B. A Classification of Glycosyl Hydrolases Based on Amino Acid Sequence Similarities. *Biochem. J.* **1991**, *280*, 309−316.
(5) Henrissat, B.; Bairoch, A. New Families in the Classification of Glycosyl Hydrolases Based on Amino Acid Sequence Similarities. *Biochem. J.* **1993**, *293*, 781−788.
(6) Henrissat, B.; Bairoch, A. Updating the Sequence Based Classification of Glycosyl Hydrolases. *Biochem. J.* **1996**, *316*, 695−696.
(7) Aronson, N. N. Jr.; Backes, M.; Kuranda, M. J. Rat Liver Chitobiase: Purification, Properties, and Role in the Lysosomal Degradation of Asn-Linked Glycoproteins. *Arch. Biochem. Biophys.* **1989**, *272*, 290−300.
(8) Fisher, K. J.; Aronson, N. N. Jr. Cloning and Expression of the cDNA Sequence Encoding the Lysosomal Glycosidase Di-*N*-Acetylchitobiase. *J. Biol. Chem.* **1992**, *267*, 19607−19616.
(9) Phillips, D. C. The Three-dimensional Structure of an Enzyme Molecule. *Sci. Amer.* **1966**, *215*, 78−90.
(10) Withers, S. G.; Aebersold, R. Approaches to Labelling and Identification of Active-Site Residues in Glycosidases. *Prot. Sci.* **1995**, *4*, 361−372.
(11) Davies, G.; Hennrisat, B. Structures and Mechanisms of Glycosyl Hydrolases. *Structure* **1995**, *3*, 852−859.
(12) Matsumura, I.; Kirsch, J. F. Is Aspartate 52 Essential for Catalysis by Chicken Egg White Lysozyme? The Role of a Natural Substrate-Assisted Hydrolysis. *Biochemistry* **1996**, *35*, 1881−1889.
(13) Reardon, D.; Farber G. K. The Structure and Evolution of α/β Barrel Proteins. *FASEB J.* **1995**, *9*, 497−503.

(14) DeSouza, M. M.; Murray, M. K. An Estrogen-Dependent Secretory Protein Which Shares Identity with Chitinases, is Expressed in a Temporally and Regionally Specific Manner in the Sheep Oviduct at the Time of Fertilization and Embryo Development. *Endocrinology* **1994**, *136*, 2485−2496.
(15) Hakala, B. E.; White, C.; Recklies, A. D. Human Cartilage Gp-39, a Major Secretory Product of Articular Chondrocytes and Synovial Cells, is a Mammalian Member of a Chitinase Protein Family. *J. Biol. Chem.* **1993**, *268*, 25803−25810.
(16) Morrison, B. W.; Leder, P. *Neu* and *Ras* Initiate Murine Mammary Tumors that Share Genetic Markers Generally Absent in c-*myc* and *int*-2-Initiated Tumors. *Oncogene* **1994**, *9*, 3417−3426.
(17) Hollak, C. E. M.; van Weely, S.; van Oers, M. H. J.; Aerts, J. M. F. G. Marked Elevation of Plasma Chitotriosidase Activity. *J. Clin. Invest.* **1994**, *93*, 1288−1292.
(18) Renkema, G. H.; Boot, R. G.; Muijsers, A. O.; Donker-Koopman, W. E.; Aerts, J. M. F. G. Purification and Characterization of Human Chitotriosidase, a Novel Member of the Chitinase Family of Proteins. *J. Biol. Chem.* **1995**, *270*, 2198−2202.
(19) Renkema, G. H.; Boot, R. G.; Strijland, A.; Donker-Koopman, W. E.; Van Den Berg, M.; Muijsers, A. O.; Aerts, J. M. F. G. Synthesis, Sorting and Processing into Distinct Isoforms of Human Macrophage Chitotriosidase. *Eur. J. Biochem.* **1997**, *244*, 279−285.
(20) Terwisscha van Scheltinga, A. C.; Hennig, M.; Dijkstra, B. W. The 1.8 Å Resolution Structure of Hevamine, a Plant Chitinase/Lysozyme, and Analysis of the Conserved Sequence and Structure Motifs of Glycosyl Hydrolase Family 18. *J. Mol. Biol.* **1996**, *262*, 243−257.
(21) Levitt, M. Accurate Modeling of Protein Conformation by Automatic Segment Matching. *J. Mol. Biol.* **1992**, *266*, 507−533.
(22) Terwisscha van Scheltinga, A. C.; Armand, S.; Kalk, K. H.; Henrissat, B.; and Dijkstra, B. W. Stereochemistry of Chitin Hydrolysis by a Plant Chitinase/Lysozyme and X-ray Structure of a Complex with Allosamidin: Evidence for Substrate Assisted Catalysis. *Biochemistry* **1995**, *34*, 15619−15623.
(23) Surles, M. C.; Richardson, J. S.; Richardson, D. C.; Brooks, F. P., Jr. Sculpting Proteins Interactively. *Prot. Sci.* **1994**, *3*, 198−210.
(24) Davies, G. J.; Wilson, K. S.; Hennrisat, B. Nomenclature for Sugar-binding Subsites in Glycosyl Hydrolases. *Biochem. J.* **1997**, *321*, 557−559.
(25) Lakowski, R. A.; MacArthur, M. W.; Ross, D. R.; Thorton, J. M. PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Cryst.* **1993**, *26*, 283−291.

CI970236V