X-ray Structure of Papaya Chitinase Reveals the Substrate Binding Mode of Glycosyl Hydrolase Family 19 Chitinases[†,‡]

Joëlle Huet,[§] Prakash Rucktooa,[||] Bernard Clantin,[||] Mohamed Azarkan,[⊥] Yvan Looze,[§] Vincent Villeret,[||] and René Wintjens*[,§]

Service de Chimie Générale (CP: 206/4), Institut de Pharmacie, Université Libre de Bruxelles (ULB), Campus de la Plaine, Boulevard du Triomphe, B-1050 Brussels, Belgium, CNRS-UMR 8161, Institut de Biologie de Lille, Institut Pasteur de Lille, Université de Lille 1, Université de Lille 2, BP 447, 1 rue du Professeur Calmette, F-59021 Lille, France, and Protein Chemistry Unit, Faculty of Medicine, Université Libre de Bruxelles (ULB), Campus Erasme (CP 609), 808 route de Lennik, B-1070 Brussels, Belgium

Received April 14, 2008; Revised Manuscript Received June 10, 2008

ABSTRACT: The crystal structure of a chitinase from *Carica papaya* has been solved by the molecular replacement method and is reported to a resolution of 1.5 Å. This enzyme belongs to family 19 of the glycosyl hydrolases. Crystals have been obtained in the presence of *N*-acetyl-D-glucosamine (GlcNAc) in the crystallization solution and two well-defined GlcNAc molecules have been identified in the catalytic cleft of the enzyme, at subsites −2 and +1. These GlcNAc moieties bind to the protein via an extensive network of interactions which also involves many hydrogen bonds mediated by water molecules, underlying their role in the catalytic mechanism. A complex of the enzyme with a tetra-GlcNAc molecule has been elaborated, using the experimental interactions observed for the bound GlcNAc saccharides. This model allows to define four major substrate interacting regions in the enzyme, comprising residues located around the catalytic Glu67 (His66 and Thr69), the short segment E89−R90 containing the second catalytic residue Glu89, the region 120−124 (residues Ser120, Trp121, Tyr123, and Asn124), and the α-helical segment 198−202 (residues Ile198, Asn199, Gly201, and Leu202). Water molecules from the crystal structure were introduced during the modeling procedure, allowing to pinpoint several additional residues involved in ligand binding that were not previously reported in studies of poly-GlcNAc/family 19 chitinase complexes. This work underlines the role played by water-mediated hydrogen bonding in substrate binding as well as in the catalytic mechanism of the GH family 19 chitinases. Finally, a new sequence motif for family 19 chitinases has been identified between residues Tyr111 and Tyr125.

Higher plants produce a myriad of defense proteins that allow their protection against stresses and pathogenic attacks. These so-called pathogenesis-related proteins encompass a large variety of functions with putative protective roles (*1*). The most abundant class of these proteins contains lytic enzymes such as β-1,3 glucanases and chitinases.

Chitinases, like lysozymes, hydrolyze the β-1,4 glycosidic bonds which are often found in structural polysaccharides of cell walls. Chitinases cleave preferentially the glycosidic bonds between GlcNAc[1] residues and differ from lysozymes, which hydrolyze the bonds between *N*-acetyl-D-muramic acid and GlcNAc. Chitin, the natural substrate of chitinases, is a linear β-1,4-linked homopolymer of GlcNAc ($(GlcNAc)_n$), found in the cuticle of insect shells, in the outer shell of crustaceans, and in the cell walls of many fungi. Chitin is believed to be the second most abundant polysaccharide on earth next to cellulose.

Chitin-degrading enzymes have shown various potential applications in agricultural, biological, and environmental fields, such as their use against chitin-containing pathogens, in the biological control of soil-borne fungal diseases, or, also, in case of chitinases with chitosanase activity, in preparation of low molecular weight chitosan oligomers for clinical applications as wound healer, blood anticoagulants, and hemostatic materials.

Chitinases are classified into the glycosyl hydrolase (GH) family 18 or 19 (*2, 3*). These two families do not share

[†] J.H., R.W., and Y.L. gratefully acknowledge the Communauté Française de Belgique (ARC) for its financial support. R.W. is a Research Associate at the National Fund for Scientific Research (Belgium). V.V. is supported by an Action Thématique et Incitative sur Programme from the CNRS and by the Région Nord-Pas de Calais through the Contrat Plan Etat-Région and Fonds Européen de développement Régional programs.

[‡] The refined coordinates and structure factors of papaya chitinase have been deposited in the RCSB Protein Data Bank with accession code RCSB047093 (pdb id: 3cql). Cartesian coordinates of the $(GlcNAc)_4$/papaya chitinase complex model can be obtained by request to authors.

* To whom correspondence should be addressed. Telephone: +32255556299. Fax: +3225556782. E-mail: rene.wintjens@ulb.ac.be.

[§] Service de Chimie Générale, Institut de Pharmacie, Université Libre de Bruxelles.

[||] Institut de Biologie de Lille, Institut Pasteur de Lille, Université de Lille.

[⊥] Protein Chemistry Unit, Faculty of Medicine, Université Libre de Bruxelles.

[1] Abbreviations: GH, glycosyl hydrolase; GlcNAc, *N*-acetyl-D-glucosamine; HEWL, hen egg white lysozyme; NAM, *N*-acetylmuramic acid; $(GlcNAc)_n$, β-1,4-linked oligosaccharide of GlcNAc with a polymerization degree of *n*; $(GlcNAc)^X$, GlcNAc molecule bound at the substrate binding subsite X; rms, root mean square.

10.1021/bi800655u CCC: $40.75
Published on Web 07/18/2008

sequence or structural similarities and act using different catalytic mechanisms. GH family 18 chitinases are found in a wide diversity of organisms. Their catalytic domains are defined by an eight (α/β) barrel (*4*) and use a substrate-assisted double-displacement mechanism, which leads to the retention of the anomeric carbon C1 in its initial configuration (*5, 6*). GH family 19 chitinases are mainly found in plants but have also been recently discovered in bacteria. Their 3D structures show a high α-helical content and have a topology similar to lysozymes from goose, phage, and hen (*7–9*). The catalytic mechanism of family 19 chitinases involves a single displacement which leads to an inversion of the configuration of the anomeric carbon (*10*).

There is so far no structure of an enzyme in complex with substrates or ligands for GH from family 19, unlike the enzymes from family 18, for which many complexes have been reported (*11*). Only theoretical models were built, mostly based on the complex between HEWL and a trisaccharide substrate (*7, 11–13*). However, HEWL (GH family 22) is known to hydrolyze its substrate with retention of the anomeric configuration (*10*) like GH family 18 chitinases and, as such, is not ideally suited for modeling studies of GH family 19 catalysis. Moreover, the binding cleft of HEWL appears more extended on one side when compared to that of the barley chitinase, so far the most studied chitinase of GH family 19 (*11*), an observation which underlies differences in enzymatic mechanisms. HEWL predominantly hydrolyzes a molecule of $(GlcNAc)_6$ into $(GlcNAc)_4$ and $(GlcNAc)_2$ (*14*), whereas the barley chitinase produces in equal amount $(GlcNAc)_3$ and $(GlcNAc)_4$/$(GlcNAc)_2$ (*15, 16*). Finally, all reported models of family 19 chitinase complexes with $(GlcNAc)_n$ do not include water molecules in the substrate binding cleft, leading to an incomplete view of the presumed catalytic mechanism (*7, 11–13*).

We report here the crystal structure of a complex of a *Carica papaya* chitinase with GlcNAc. Two GlcNAc molecules are found in the active site, allowing to model a complex of this GH family 19 chitinase with $(GlcNAc)_4$, in which water molecules are incorporated.

EXPERIMENTAL PROCEDURES

X-ray Diffraction Data Collection. Protein purification and crystallization procedures were reported previously (*17, 18*). Well-diffracting crystals of the papaya chitinase were obtained by the vapor diffusion method in hanging drops with the following conditions: Li_2SO_4 (0.17 M), Tris-HCl (0.085 M) pH 8.5, PEG 4000 (25.5%), glycerol (15%), and in the presence of a large excess of GlcNAc (*18*). X-ray diffraction data were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) on beamline BM30-A. Diffraction data were processed with the XDS suite of programs (*19*). The crystals belong to space group $P2_1$, with cell dimensions $a = 44.49$ Å, $b = 69.00$ Å, $c = 76.81$ Å, and $\beta = 94.97°$ and diffract to at least 1.5 Å resolution. Crystal data and data collection statistics are summarized in Table 1.

Structure Determination and Refinement. Analysis of the unit-cell content suggested the presence of two protein molecules in the asymmetric unit, consistent with a solvent content of 42.78% (*20*). The structure was solved by molecular replacement using the CCP4 suite of programs (*21*). For the molecular replacement procedure, a model was elaborated using MODELER (*22*), starting from the structure of barley chitinase (pdb code: 2baa) (*7*). This model was subsequently used in the MOLREP program (*23*).

Table 1: X-ray Data Collection, Refinement, and Model Statistics

	papaya chitinase
Crystal Parameters	
space group	$P2_1$
unit cell (Å)	$a = 44.49$, $b = 69.0$, $c = 76.81$; $\beta = 94.97°$
no. of protein molecules in asymmetric unit	2 monomers
Data Collection[a]	
source/beamline	ESRF/BM30A
wavelength (Å)	0.978872
resolution (Å)	20.0−1.50 (1.59−1.50)
no. of reflections	301176 (45870)
no. of unique reflections	71589 (10910)
completeness (%)	95.2 (90.6)
R_{merge}[b]	6.4 (39.6)
mean $I/\sigma(I)$	17.94 (4.23)
Refinement Statistics	
resolution (Å)	19.68−1.50
R_{factor}[c] (%)	16.1
R_{free}[d] (%)	18.8
Model Statistics	
total no. of protein atoms	3744
no. of GlcNAc molecules	6
no. of water molecules	403
no. of other atoms	26
$rmsd_{bond}$ (Å)	0.008
$rmsd_{angle}$ (deg)	1.17
$\langle B\text{-factor}\rangle_{\text{all protein atoms}}$ (Å^2)	11.74
$\langle B\text{-factor}\rangle_{\text{protein main chain atoms}}$ (Å^2)	10.54
$\langle B\text{-factor}\rangle_{\text{GlcNAc}}$ (Å^2)	13.98
$\langle B\text{-factor}\rangle_{\text{water}}$ (Å^2)	24.48
Ramachandran distribution	
most favorable regions (%)	87.0
in additional regions (%)	12.5
in disallowed regions (%)	0.5

[a] Statistics for the highest resolution shell in parentheses. [b] $R_{merge} = \sum_{hkl}\sum_i|I_{hkl,i} - \langle I_{hkl}\rangle|/\sum_{hkl}\sum_i\langle I_{hkl}\rangle$, where $I_{hkl,i}$ is the ith observed intensity of reflection hkl and $\langle I_{hkl}\rangle$ is the mean intensity for all observations i of reflection hkl. [c] $R_{factor} = \sum||F_o| - |F_c||/\sum|F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively. [d] R_{free} was calculated from a random selection of 5.02% of the data.

The structure was progressively rebuilt and fitted, using both $2F - F_c$ and $F_o - F_c$ maps in coot (*24*). The loop from residue 88 to residue 109 was completely reconstructed by hand. Final refinement cycles were performed with restrained constraints in REFMAC5 (*25*). Water molecules were added using the "find waters" procedure implemented in REFMAC5 (*25*). Finally, the quality of the structure was evaluated using PROCHECK (*26*). Refinement statistics are given in Table 1.

Structural Analysis and Comparisons. Secondary structures were defined with DSSP (*27*). Intermolecular interactions were analyzed using LigPlot (*28*) and HBplus (*29*).

The papaya chitinase structure was compared to the other available chitinase structures, using a modified version of the SoFi algorithms (*30*), which extend the alignment to allow that all residues, except those corresponding to insertions and deletions, are superimposed. The X-ray structures included in this comparison are the chitinases from rice (pdb code 2dkv, *31*), barley (2baa, *7*), jack bean (1dxj,

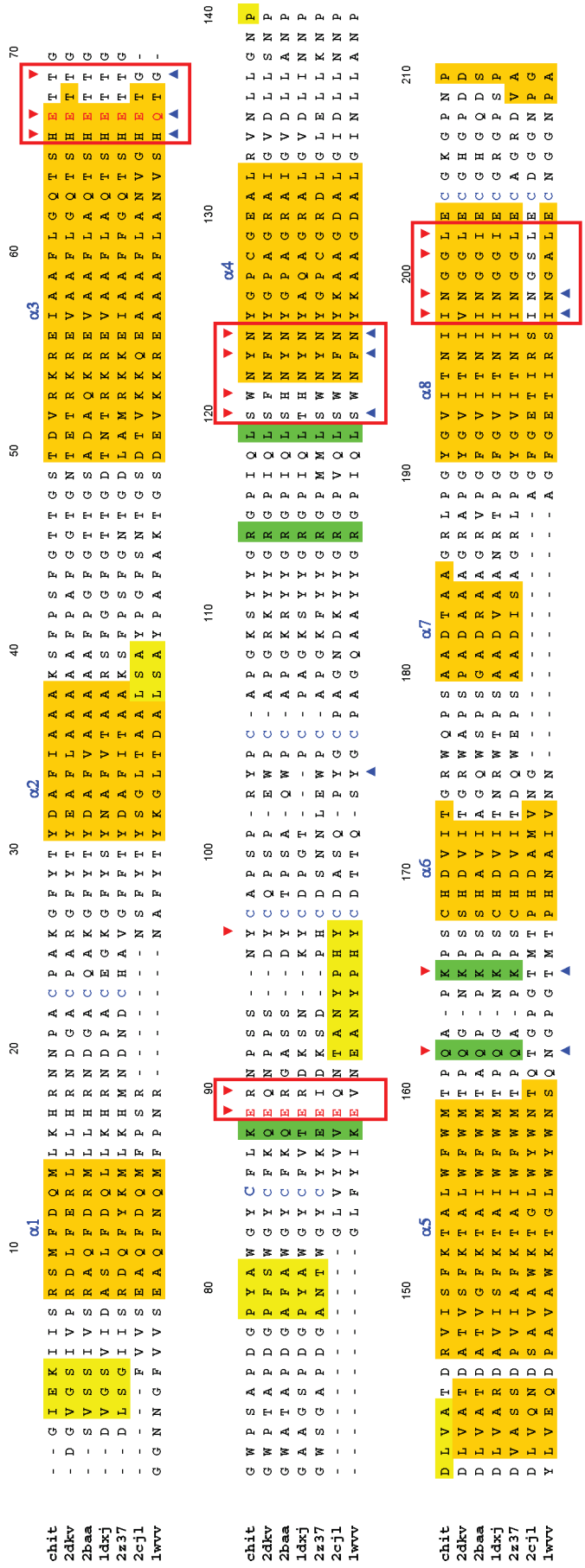

FIGURE 1: Structure-based multiple sequence alignment. Sequences for all GH family 19 chitinases with known structure were aligned onto the papaya chitinase sequence using a modified version of the SoFi program (30). The first row contains the residue numbering according to the papaya chitinase sequence. The proteins are labeled by their PDB entry code, except the papaya chitinase, which is listed first. Protein names are listed in the Experimental Procedures section. Secondary structures, defined with DSSP (27), are colored according to the code indicated at the bottom of the figure. Catalytic residues are in red and cysteine residues involved in disulfide bonds in blue. Residues found in intermolecular interactions in the papaya chitinase/$(GlcNAc)_4$ complex model, either directly or water-mediated, are shown by red arrows above the papaya chitinase. Residues predicted to interact with saccharide molecules in theoretical models of the barley chitinase/$(GlcNAc)_6$ complex (11, 12) are also indicated by blue arrows below the protein sequences. Main binding regions with GlcNAc molecules are boxed in red.

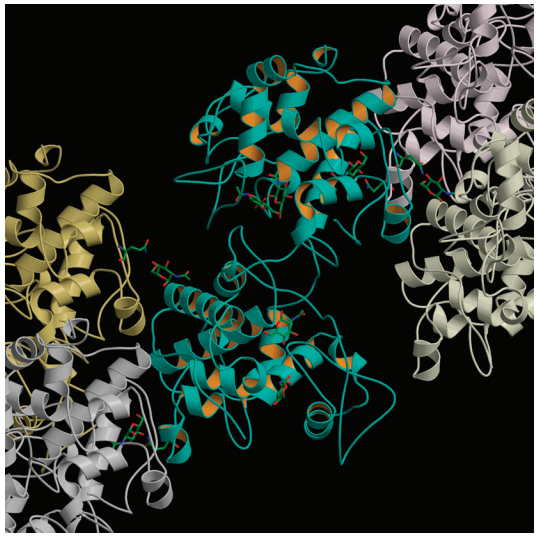

FIGURE 2: Crystal packing of papaya chitinase. The two protein molecules found in the asymmetric unit are drawn in turquoise-colored ribbons with four other symmetric molecules, which are shown in pale-colored ribbons. The GlcNAc molecules and also Gln177, which actively participate to the crystal packing, are illustrated. The two GlcNAc molecules found in the catalytic cleft are also shown. The figure was generated using MolScript and Raster3D.

32), leaf mustard (2z37, *33*), *Streptomyces coelicolor* (2cjl, *34*), and *Streptomyces griseus* (1wvv, *35*). Conservations of amino acids in GH family 19 chitinases were evaluated using an alignment of 370 sequences. This multiple alignment was obtained using clustalX (*36*) and subsequently optimized manually. This alignment is given in the Supporting Information section (Figure S1).

Steady-State Fluorescence Spectroscopy. Fluorescence spectra were recorded using a Perkin-Elmer LS 55 fluorometer, and measurements were made in the concentration range where the emission was linear regarding the fluorophore concentrations. The emission and excitation bandwidths were 5.0 nm, and the temperature of the solution in the cell was maintained at 25.0 ± 0.1 °C.

The quenching experiments were achieved in 50 mM sodium acetate buffer at pH 5.0. Solutions were prepared from the buffer containing 200 mM GlcNAc and a stock solution of protein in water. Measurements were carried out using a protein concentration of 77 μg/mL and concentrations of GlcNAc of 50 and 100 mM. Fluorescence intensities were measured at 340 nm using an excitation wavelength of 295 nm for selective excitation of Trp residues. Fluorescence contribution from GlcNAc was subtracted from all the spectra.

Molecular Modeling of the (GlcNAc)$_4$−Papaya Chitinase Complex. By convention, the polysaccharide units are named from the nonreducing to the reducing end. The GlcNac molecules in the crystal structure of the papaya enzyme are thus named (GlcNAc)^C and (GlcNAc)^E. The initial conformation of the tetra-*N*-acetyl-D-glucosamine moiety was taken from the hevamine complex structure (pdb entry 1kr0, *37*). It was superimposed to the GlcNAc molecules found in the papaya chitinase crystal structure, in such a way that the six cycle atoms of the first and third units of the (GlcNAc)$_4$ were superimposed respectively to (GlcNAc)^C and (GlcNAc)^E of the papaya chitinase. Thus the (GlcNAc)$_4$ molecule occupies subsites C, D, E, and F according to HEWL binding convention (*38*). A conjugate gradient minimization was then applied using X-plor (*39*), in two cycles of 200 minimization steps with harmonic restraint of 10 kcal/mol on the sugar cycle atoms of GlcNAc units at positions C and E. The positions of all protein backbone atoms were kept fixed during the minimization. A total of 189 water molecules were kept in the initial model, and their positions were also energetically optimized, including the nine water molecules found in the active site. The rms deviation of protein side chain atoms between the initial structure and the minimized model is 0.29 Å, and 1.01 Å for the (GlcNAc)$_4$ molecule and 0.35 Å for the 189 solvent molecules.

RESULTS

Structure of Papaya Chitinase. The crystal structure of the papaya chitinase was determined by the molecular replacement method using a model elaborated with MODELER (*22*), starting from the structure of barley chitinase (pdb entry 2baa (*7*); sequence identity of 73%). The crystal contains two molecules in the asymmetric unit, with a Matthews coefficient (V_M) value of 2.15 Å^3/Da (*20*), which corresponds to a solvent content of 42.78%. The final model of papaya chitinase was refined to a resolution of 1.5 Å. All atoms from the 243 residues were well defined in electron density, and thereby at that resolution, uncertainties inherent to sequencing by mass spectrometry (*17*) were elucidated. An isoleucine residue was thus identified instead of a leucine at positions 35, 57, 117, 149, 172, 194, 197, 198, 216, and 225, and an isoleucine was replaced by a leucine at position 155 (consequently, the SwissProt database entry P85084 has been modified accordingly). With the exception of Trp121 located in the active site, all non-glycine and non-proline residues were in the allowed regions of the Ramachandran plot (87.0% of residues in most favored regions and 12.5% in additionally allowed regions). Loop 88−109 was not visible in the initial electron density map and was manually built during the refinement process.

Like other family 19 chitinases, the papaya enzyme is made up of nine α-helices comprising 6−18 residues, thus representing an overall 40% of α-helix content. Three 3_{10} helices are also found at positions 2−4, 79−81, and 140−144 (Figure 1). Three disulfide bonds which are often conserved in plant chitinases are also present in the papaya chitinase, between Cys23 and Cys85, Cys97 and Cys105, and Cys204 to Cys236. Five residues are found within isolated β-bridges, forming a short β-hairpin (Gln162 and Lys165) and the smallest possible antiparallel β-sheet (Lys88, Arg114, Leu119). The catalytic cleft goes across the central portion of the protein and involves the β-sheet, the N-terminal part of helix α4, and the C-terminal part of helices α3 and α8 (Figure 1).

Six crystal structures of GH family 19 chitinases are currently available in the Protein Data Bank, four coming from plants and two from bacteria (they are listed in the Experimental Procedures section). The papaya enzyme presents a sequence identity ranging from 33% to 76% and is structurally similar to the other family 19 chitinases, with an overall rms deviation ranging from 0.85 to 2.07 Å. The

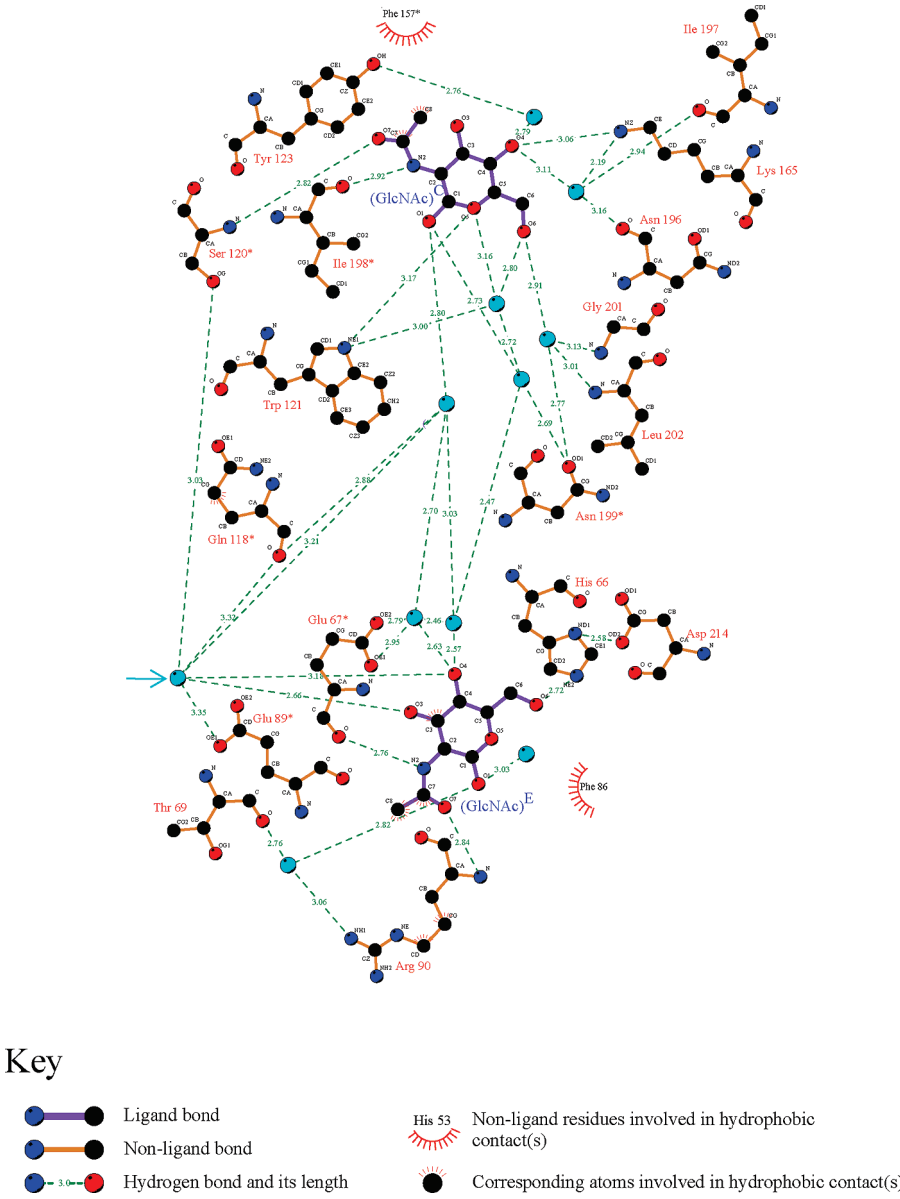

FIGURE 3: Schematic representation of the interactions between the enzyme and GlcNAc molecules. This picture was obtained using LigPlot (28). The atoms involved in hydrogen bonds (with distances) or hydrophobic contacts are depicted. The legend is indicated below. Residues well conserved (more than 85%) in the 370 sequences of the GH family 19 chitinases used in this study are marked by an asterisk. The water molecule proposed to actively participate in the enzymatic mechanism is shown by a light blue arrow.

structure of the papaya chitinase most resembles that of the mustard chitinase (2z37, *33*). Secondary structures among GH family 19 chitinases are rather well conserved (Figure 1) with only minor differences. For instance, the papaya chitinase has a 3_{10} helix in region 140−144 instead of an α-helix in other chitinases.

N-Acetyl-D-glucosamine Binding Sites in Papaya Chitinase X-ray Structure. The two monomers in the asymmetric unit are each complexed with three GlcNAc moieties. Two are located in the catalytic cleft, and the third one contributes to the crystal packing, essentially by hydrogen bonds between its O4 atom and the Oϵ1 of the Gln177 residue from an adjacent asymmetric unit (Figure 2). These intermolecular contacts seem important in obtaining well-diffracting crystals, as crystallization attempts in the absence of GlcNAc always resulted in poorly diffracting crystals (*18*).

The two GlcNAc molecules in the catalytic cleft are bound to subsites equivalent to subsites C and E in HEWL (according to the nomenclature early defined by Blake et al. (38)). These molecules are well defined, with average temperature factors of 9.9 and 5.9 Å^2 at subsites C $(GlcNAc)^C$ and E $(GlcNAc)^E$, respectively. These subsites correspond respectively to positions −2 and +1 in relation to the glycosidic bond cleavage (*40*). The so-called A, B, C, D, E, and F (*38*) binding subsites of HEWL must be adapted into B, C, D, E, F, and G in GH family 19 chitinases, as a result of a shift toward the reducing end side (*12, 15*) of the binding cleft in these enzymes.

The $(GlcNAc)^C$ unit binds to the papaya enzyme via an extensive network of interactions, some of them implicating water molecules (Figure 3). Four direct hydrogen bonds are observed between $(GlcNAc)^C$ and residues Ser120, Trp121, Lys165, and Ile198. Five hydrogen bonds mediated by water molecules are also observed, bridging $(GlcNAc)^C$ with residues Gln118, Tyr123, Asn199, Ile197, and Leu202 (Figure 3). In addition, a hydrophobic stacking is found

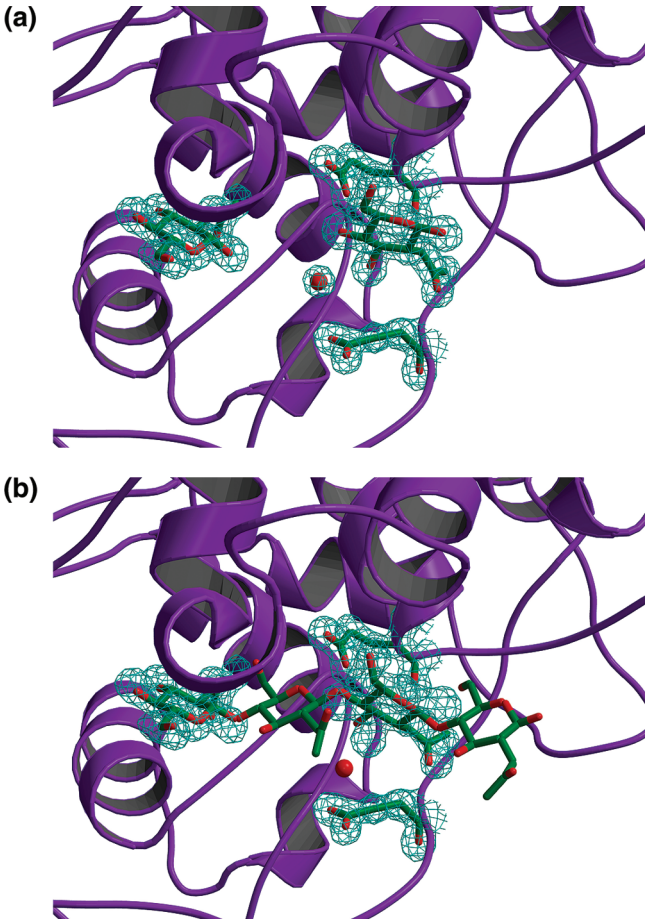

FIGURE 4: Close-up view of the GlcNAc binding site in the catalytic cleft: (a) observed GlcNAc molecules in the crystal structure of papaya chitinase; (b) complex model of $(GlcNAc)_4$/papaya chitinase. The protein was shown in mallow ribbons. In both pictures, the observed $2F_o - F_c$ electron density map was contoured at the 1.0σ level for the two observed GlcNAc molecules, the two catalytic glutamates, and the nucleophilic attacking water molecule. The picture was generated with BobScript.

between the sugar acetyl group of $(GlcNAc)^C$ and the side chain of Phe157. The second saccharide molecule found in the catalytic cleft, $(GlcNAc)^E$, interacts with the enzyme through eight hydrogen bonds, four of them involving water molecules. These hydrogen bonds involve the following protein residues: Arg90, Glu67, His66, Glu89, Gln118, and Ser120 (Figure 3).

The acetyl group of $(GlcNAc)^E$ is in hydrophobic contact with the side chain of Phe86. Within the asymmetric unit, loop 72−78 of one monomer penetrates into the substrate binding cleft of the other monomer at its reducing end side. Consequently, the $(GlcNAc)^E$ molecule of the second monomer interacts with residues Trp72 and Ser74 of the adjacent monomer (Figure 3). These additional interactions participate to the crystal packing, underlining the role played by GlcNAc molecules for the crystallization of papaya chitinase.

Binding Experiments. The interactions between the papaya chitinase and GlcNAc molecules were further characterized by fluorescence spectroscopy. The enzyme has one tryptophan residue (Trp121) located within its substrate binding cleft. Conformational changes may alter its emission spectra and thereby provide information about substrate binding. By setting the excitation wavelength at 295 nm, the enzyme has a strong fluorescence emission band at 340 nm, while GlcNAc has a fluorescence emission at 385 nm. Addition of increasing amounts of GlcNAc to papaya chitinase results in a linear decreasing of its fluorescence intensity, while no shift can be observed in the maximum emission wavelength (Figure S2 in Supporting Information). This behavior suggests a fluorescence quenching without large conformational changes in the protein structure. Unfortunately, the very low association constant (estimated value about 2−4 M^{-1}) between the enzyme and GlcNAc molecules did not allow us to perform a Stern−Volmer analysis, and thus we could not estimate the thermodynamic parameters of binding. Similar uncertainties have been reported regarding fluorescence spectroscopy experiments of GlcNAc binding to lysozyme (*41*).

Catalytic Binding Mode of Family 19 Chitinases. The binding of two GlcNAc molecules at specific subsites in the catalytic site of the papaya enzyme (Figure 4a) has allowed us to model a complex of this chitinase with $(GlcNAc)_4$. The $(GlcNAc)_4$ complex was elaborated using as a starting model the coordinates of the $(GlcNAc)_4$ molecule observed in the complex with hevamine (pdb code 1kr0, *37*). It has been positioned in the papaya chitinase structure in such a way as to conserve the positions of $(GlcNAc)^C$ and $(GlcNac)^E$, and the model was subsequently minimized with harmonic restraints on the sugar cycle atoms at these positions. These GlcNAc moieties showed thus only very small conformational changes upon energy minimization (Figure 4b). The carbohydrate cycle of $(GlcNAc)^E$ undergoes only a small flip due to the steric presence of the adjacent $(GlcNAc)^D$. Also, side chain atoms of the two catalytic glutamate (Glu67 and Glu89) and the catalytic water were only marginally displaced (Figure 4b). Consequently, experimentally observed interactions in the catalytic site of the enzyme in the crystal structure were preserved in the $(GlcNAc)_4$ complex.

The $(GlcNAc)_4$ molecule interacts through an extensive network of interactions, which also involves many hydrogen bonds mediated by water molecules (Figure 5). Interacting residues are mainly clustered into four regions along the amino acid sequence of papaya chitinase (Figure 1), and most of them were previously predicted to be involved in hexa-GlcNAc binding interactions (*7, 11, 12*). Nevertheless, five amino acids (Tyr96, Glu89, Gly201, Leu202, and Trp121) are now unambiguously identified as interacting residues by the inclusion of water molecules into the docking procedure.

The C-terminal part of helix α3 binds to the saccharide part of $(GlcNAc)^E$ and $(GlcNAc)^F$ via residues His66, Glu67, and Thr69. The backbone carbonyl of Glu67 forms a hydrogen bond with the *N*-acetyl amide of $(GlcNAc)^E$ whereas the carboxylate group of Glu67 forms a hydrogen bond with the β-1,4-glycosidic oxygen linking $(GlcNAc)^D$ and $(GlcNAc)^E$. The HO6 proton of $(GlcNAc)^E$ interacts with the side chain of His66. $(GlcNAc)^F$ is bound only by one hydrogen bond to the backbone amide of Thr69. Residues Glu89 and Arg90 form the second binding region (Figure 1). They are located into the very long flexible loop 88−114 that interacts with $(GlcNAc)^E$ via its *N*-acetyl carbonyl and HO3 groups. Interestingly, the carboxylate of Glu89 interacts with the HO3 group of $(GlcNAc)^E$ through the water molecule which was expected to play a role in the single-displacement inverting mechanism of family 19 chitinases (showed as orange dots in Figure 5) (*11, 32*).

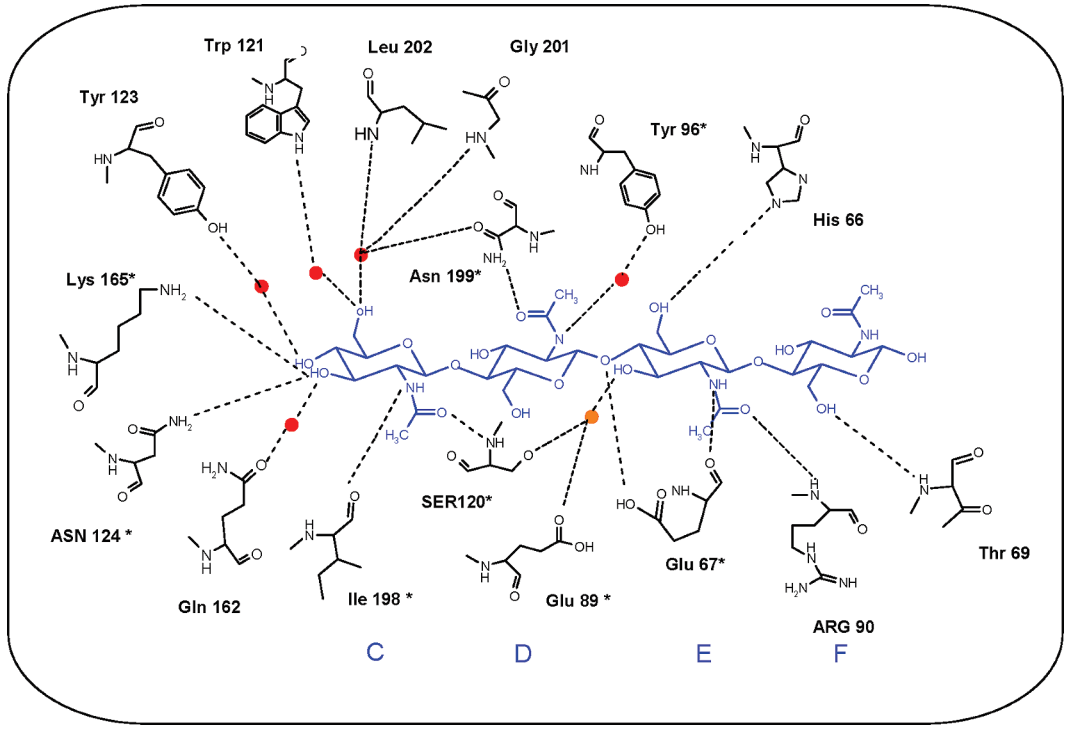

FIGURE 5: Schematic diagram of interactions in the model of the $(GlcNAc)_4$/papaya chitinase complex. $(GlcNAc)_4$ is drawn in blue, and water molecules are showed as red dots. The catalytic water molecule is in orange dots. Hydrogen bonds are represented in broken lines. Asterisks indicate residues highly conserved (>85%) in the 370 aligned sequences of GH family 19 chitinases.

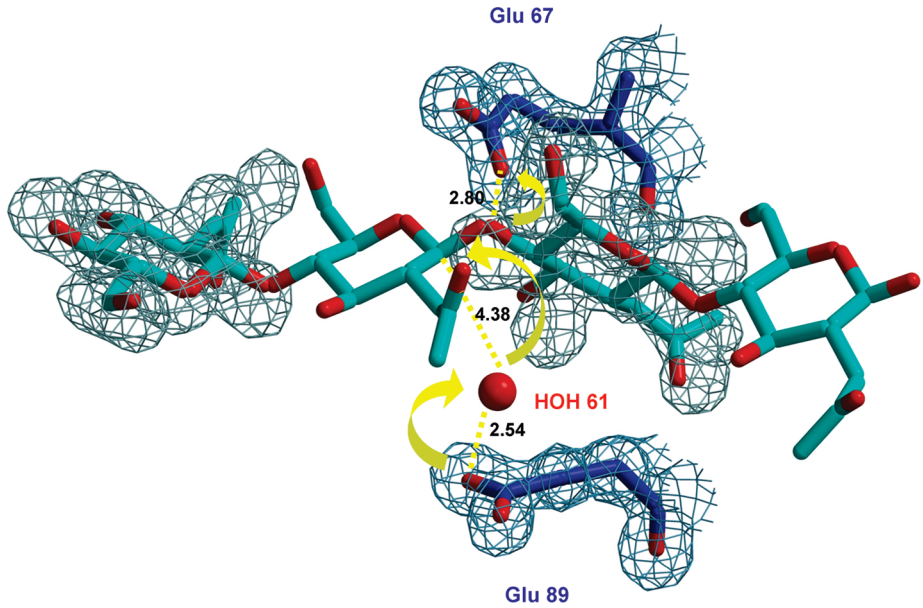

FIGURE 6: Catalytic mechanism proposed for GH family 19 chitinases inferred from the model of the chitinase complex. The electron density of the two GlcNAc observed in the crystal structure is depicted in gray. $(GlcNAc)_4$ is represented in light blue, and residues involved in the catalytic center are in dark blue. The figure was generated using BobScript.

The third binding region encompasses residues Ser120, Trp121, Tyr123, and Asn124, located in the N-terminal part of helix α4. The side chain of Ser120, like the catalytic residue Glu89, forms a water-mediated hydrogen bond with the OH3 group of $(GlcNAc)^E$. The last binding region is situated in the N-terminal part of helix α8 and involves residues Ile198, Asn199, Gly201, and Leu202, which again mainly interact with the sugar part of $(GlcNAc)^C$. Interestingly, the side chain amino group of Asn199 forms a hydrogen bond with the *N*-acetyl carbonyl of $(GlcNAc)^D$ whereas its side chain carbonyl group forms a water-mediated hydrogen bond with HO6 of $(GlcNAc)^C$.

Besides these four major interacting regions, three other residues also interact with the $(GlcNAc)_4$ molecule. The side chain hydroxyl of the highly conserved Tyr96 forms a water-mediated hydrogen bond with the *N*-acetyl amide of $(GlcNAc)^D$. The OH4 of the $(GlcNAc)^C$ sugar forms a hydrogen bond with the side chain amine of Lys165, and finally, the side chain of Gln162 forms a water-mediated hydrogen bond with the OH3 of $(GlcNAc)^C$ (Figure 5).

A New Sequence Motif in the GH Family 19 Chitinases. The main chain of Trp121 displays a characteristic structure with backbone torsion angles in a left-handed conformation. This conformation is also observed for the equivalent residue

in all available structures of GH family 19 chitinases. Trp121 also forms an aromatic cluster with Tyr96 and Tyr123, which is also conserved in all chitinase structures. A highly conserved motif, which characterizes GH family 19 chitinases, could be defined between Tyr111 and Tyr125 as ^{111}Y-[FHY]-G-R-G-[AP]-x-Q-[IL]-[ST]-[FHYW]-[HN]-[FY]-N-Y^{125}, where x represents any amino acid. Among the 370 studied sequences of chitinases, 337 of them (91%) contain this motif. A survey of the sequences in the TrEMBL data bank showed that the motif is very specific to GH family 19 chitinases. In the same way, a preliminary survey of the protein structures using the DaliLite server (http://ekhidna.biocenter.helsinski.fi/dali_server/) seems also to indicate a certain specific feature of this 15-residue-long structural motif. Finally, it is worth mentioning that this new motif is associated with the substrate binding site, contrary to the two other sequence patterns found in the PROSITE database for the GH family 19 chitinases (PS00773 and PS00774).

DISCUSSION

Three GlcNAc molecules bound to each chitinase monomer were found in the crystal structure, one involved in crystal packing and the two others located at subsites C and E in the catalytic site. Studies of free energy changes upon ligand binding in family 19 chitinases predicted that subsites C and F are the strongest binding positions (*13, 16*). Our results suggest indeed that the strongest binding site is $(GlcNAc)^C$ but followed by $(GlcNAc)^E$ instead of $(GlcNAc)^F$.

The 1.5 Å structure of the papaya chitinase reveals the presence of a water molecule (HOH61) in the active site, hydrogen bonded to the side chain carboxylate of Glu89, to the hydroxyl group of Ser120, and to O3 and O4 atoms of the $(GlcNAc)^E$ sugar moiety (Figure 3). It was proposed that this water molecule could be activated by Glu89 for the nucleophilic attack on the C1 atom of $(GlcNAc)^D$, thereby leading to a product with an inverted anomeric configuration (*7, 11*). In our study, the steric hindrance caused by the presence of the two saccharides, $(GlcNAc)^C$ and $(GlcNAc)^E$, probably prevents the additional binding of a $(GlcNAc)^D$ moiety. However, the distance between this water molecule and the O4 atom of $(GlcNAc)^E$, which corresponds formally to atom O1 of a $(GlcNAc)_n$ polysaccharide, is only 3.32 Å (Figure 3). Thus, our experimental observations support a mechanistic model of reaction involving the assistance of a water molecule.

It was also speculated that the very long loop 88−114 which contains the catalytic Glu89 can come closer to the second catalytic residue Glu67 in order to participate to the catalysis (*12, 33*). In the papaya chitinase crystal structure, this loop appears very flexible (average *B* factors of 16.5 Å^2), an observation which may support a potential conformational change upon substrate binding. Also, in the papaya chitinase the side chains of Glu67 and Glu89 are only 7.6 Å apart whereas there are more than 9 Å in the other chitinase that belong to family 19.

The model of papaya chitinase complexed with a tetra-GlcNAc substrate reveals four major binding regions (Figure 1) within the active site, which involve the residues around Glu67, the short segment around Glu89, the sequence region 120−124, and the α-helical segment 198−202. Site-directed mutagenesis studies have previously demonstrated the crucial role played by residues Glu67, Glu89, Tyr123, and Asn124 in the enzyme activity (*42, 43*). Our studies show that residues His66, Arg90, Tyr96, Ser120, Trp121, and Asn199, which are highly conserved in family 19 chitinases, are also involved in substrate binding.

Theoretical models of complexes between $(GlcNAc)_6$ substrate and family 19 chitinases have been previously described (*7, 11–13*). They were all elaborated from the complex between HEWL and a polysaccharide, without taking into account the presence of water molecules in the binding process. The tetrasaccharide/papaya chitinase complex, elaborated using experimental data at subsites C and E, provides additional information on substrate binding. The inclusion, during the docking procedure, of 189 water molecules observed in the crystal structure has allowed to show that Tyr96, Trp121, Gly201, and Leu202 are involved in additional binding interactions. Our results are in agreement with the following model for the catalytic mechanism (Figure 6) (*12*): the substrate binds in a conformation favoring the protonation by Glu67 of the anomeric oxygen linking sugar D and E; Asn199 can interact with the *N*-acetyl group of sugar D, preventing the formation of an oxazoline ion intermediate; Glu89, located in the very long loop 88−114, and Ser120 are coordinated to a water molecule that may be activated for the nucleophilic attack (Figure 6).

Finally, the papaya chitinase is highly resistant to proteolysis. Considering its chitosanase activity (*17*), this property could represent an advantage for its industrial use in the production of chitosan oligomers. Indeed, the enzyme is purified without any polypeptidic cleavage from a source containing up to 1 mM fully active cysteine proteinases (*17*). The crystal structure of papaya chitinase does not provide the clues to explain such property, excepting the relative high occurrence of proline residues in loop regions (*17*) that may be relevant in protecting this enzyme against proteolytic degradation, as already reported for several other proteins (*44–46*). Revealing the mechanisms underlying such protection will require further investigations.

ACKNOWLEDGMENT

We thank the staff of ESRF (Grenoble) for making the BM30A beamline available to us.

SUPPORTING INFORMATION AVAILABLE

Figure S1, multiple alignment of 370 sequences of GH family 19 chitinases, and Figure S2, effect of GlcNAc on the fluorescence spectrum of papaya chitinase. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Edreva, A. (2005) Pathogenesis-related proteins: research progress in the last 15 years. *Gen. Appl. Plant Physiol. 31*, 105–124.
2. Henrissat, B. (1991) A classification of glycosyl hydrolase based on amino acid sequence similarities. *Biochem. J. 280*, 309–316.
3. Henrissat, B., and Bairoch, A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J. 293*, 781–788.
4. Davies, G., and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure 3*, 853–859.

5. Tews, I., Terwisscha van Scheltinga, A. C., Perrakis, A., Wilson, K. S., and Dijkstra, B. W. (1997) Substrate-assisted catalysis unifies two families of chitinolytic enzymes. *J. Am. Chem. Soc. 119*, 7954–7959.
6. Synstad, B., Gaseidnes, S., Van Aalten, D. M. F., Vriend, G., Nielsen, J. E., and Eijsink, V. G. H. (2004) Mutational and computational analysis of the role of conserved residues in the active site of a family 18 chitinase. *Eur. J. Biochem. 271*, 253–262.
7. Hart, P. J., Pfluger, H. D., Monzingo, A. F., Hollis, T., and Robertus, J. D. (1995) The refined crystal structure of an endochitinase from *Hordeum vulgare* L. seeds at 1.8 Å resolution. *J. Mol. Biol. 248*, 402–413.
8. Holm, L., and Sander, C. (1994) Structural similarity of plant chitinase and lysozymes from animals and phage. *FEBS Lett. 340*, 129–132.
9. Monzingo, A. F., Marcotte, E. M., Hart, P. J., and Robertus, J. D. (1996) Chitinases, chitosanases, and lysozymes can be divided into prokaryotic and eukaryotic families sharing a conserved core. *Nat. Struct. Biol. 3*, 133–140.
10. Iseli, B., Armand, S., Boller, T., Neuhaus, J.-M., and Henrissat, B. (1996) Plant chitinases use two different hydrolytic mechanisms. *FEBS Lett. 382*, 186–188.
11. Fukamizo, T. (2000) Chitinolytic enzymes: catalysis, substrate binding, and their application. *Curr. Protein Pept. Sci. 1*, 105–124.
12. Brameld, K. A., and Goddard, W. A., III (1998) The role of enzyme distorsion in the single displacement mechanism of family 19 chitinases. *Proc. Natl. Acad. Sci. U.S.A. 95*, 4276–4281.
13. Sasaki, C., Itoh, Y., Takehara, H., Kuhara, S., and Fukamizo, T. (2003) Family 19 chitinase from rice (Oryza sativa L.): substrate-binding subsites demonstrated by kinetic and molecular modelling studies. *Plant Mol. Biol. 52*, 43–52.
14. Fukamizo, T., Minematsu, Y., Yanase, Y., Hayashi, K., and Goto, S. (1986) Substrate size dependence of lysozyme-catalysed reaction. *Arch. Biochem. Biophys. 250*, 312–321.
15. Hollis, T., Honda, Y., Fukamizo, T., Marcotte, E., Day, P. J., and Robertus, J. D. (1997) Kinetic analysis of barley chitinase. *Arch. Biochem. Biophys. 344*, 335–342.
16. Honda, Y., and Fukamizo, T. (1998) Substrates binding sites of chitinase from barley seeds and lysozyme from goose egg white. *Biochem. Biophys. Acta 1388*, 53–65.
17. Huet, J., Wyckmans, J., Wintjens, R., Boussard, P., Raussens, V., Vandenbusshe, G., Ruysschaert, J.-M., Azarkan, M., and Looze, Y. (2006) Structural characterisation of two papaya chitinases, a family GH 19 of glycosyl hydrolases. *Cell. Mol. Life Sci. 63*, 3042–3054.
18. Huet, J., Azarkan, M., Looze, Y., Villeret, V., and Wintjens, R. (2008) Crystallization and X-ray analysis of a family 19 glycosyl hydrolase from Carica papaya latex, *Acta Crystallogr. F64*, 371–374.
19. Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr. 26*, 795–800.
20. Matthews, B. W. (1968) Solvent content of protein crystals. *J. Mol. Biol. 33*, 491–497.
21. The, C. C. P. 4. (1994) CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D50*, 760–763.
22. Šali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol. 234*, 779–815.
23. Vagin, A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr. 30*, 1022–1025.
24. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126–2132.
25. Murshudov, G. N., Vagin, A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D53*, 240–255.
26. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr. 26*, 283–291.
27. Kabsch, W., and Sander, C. (1983) Dictionnary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*, 2577–2637.
28. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) LIGPLOT: a program to generate diagrams of protein-ligand interactions. *Protein Eng. 8*, 127–134.
29. McDonald, I. K., and Thornton, J. M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol. 238*, 777–793.
30. Boutonnet, N. S., Rooman, M. J., Ochagavia, M. E., Richelle, J., and Wodak, S. J. (1995) Optimal protein structure alignments by multiple linkage clustering: application to distantly related proteins. *Protein Eng. 8*, 647–662.
31. Kezuka, Y., Kitazaki, K., Itoh, Y., Watanabe, J., Takaha, O., Watanabe, T., Nishizawa, Y., and Nonaka, T. (2004) Crystallization and preliminary X-ray analysis of plant class I chitinase from rice. *Protein Pept. Lett. 11*, 401–405.
32. Hahn, M., Hennig, M., Schlesier, B., and Höhne, W. (2000) Structure of jack bean chitinase. *Acta Crystollagr. D56*, 1096–1099.
33. Ubhayasekera, W., Tang, C. M., Ho, S. W. T., Berglund, G., Bergfors, T., Chye, M.-L., and Mowbray, S. L. (2007) Crystal structures of a family 19 chitinase from *Brassica juncea* show flexibility of binding cleft loops. *FEBS J. 274*, 3695–3703.
34. Hoell, I. A., Dalhus, B., Heggset, E. B., Aspmo, S. I., and Eijsink, V. G. H. (2006) Crystal structure and enzymatic properties of a bacterial family 19 chitinase reveal differences from plant enzymes. *FEBS J. 273*, 4889–4900.
35. Kezuka, Y., Ohishi, M., Itoh, Y., Watanabe, J., Mitsutomi, M., Watanabe, T., and Nonaka, T. (2006) Structural studies of a two-domain chitinase from *Streptomyces griseus* HUT6037. *J. Mol. Biol. 358*, 472–484.
36. Tompson, J. D., Gilson, T. J., Plewniak, F., Janmougin, F., and Higgins, D. G. (1997) The clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res. 24*, 4876–4882.
37. Bokma, E., Rozeboom, H. J., Sibbald, M., Dijkstra, B. W., and Beintema, J. J. (2002) Expression and characterization of active site mutants of hevamine, a chitinase from the rubber tree *Hevea brasiliensis*. *Eur. J. Biochem. 269*, 893–901.
38. Blake, C. C. F., Johnson, L. N., Mair, G. A., North, A. C. T., Phillips, D. C., and Sarma, V. R. (1967) Crystallographic studies of the activity hen egg-white lysozyme. *Proc. R. Soc. London, Ser. B 167*, 378–388.
39. Brünger, A. T. (1998) *Crystallographic refinement by simulation annealing in crystallographic computing 4: Techniques and new technologies* (Isaac, N. W., and Taylor, R. R., Eds.) pp 126−140, Clarendron Press, Oxford.
40. Davies, G. D., Wilson, K. S., and Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J. 321*, 557–559.
41. Chipman, D. M., Grisaro, V., and Sharon, N. (1967) The binding of oligosaccharides containing N-acetylglucosamine and N-acetylmuramic acid to lysozyme. *J. Biol. Chem. 242*, 4388–4394.
42. Andersen, M. D., Jensen, A., Robertus, J. D., Leah, R., and Skriver, K. (1997) Heterologous expression and characterization of wild-type and mutant forms of a 26 Kda endochitinase from barley (*Hordeum vulgare* L.). *Biochem. J. 322*, 815–822.
43. Verburg, J. G., Rangwala, S. H., Samac, D. A., Luckow, V. A., and Huynh, Q. K. (1993) Examination of the role of tyrosine-174 in the catalytic mechanism of the *Arabidopsis thaliana* chitinase: comparaison of variant chitinases generated by site-directed mutagenesis and expressed in insect cells using baculovirus vectors. *Arch. Biochem. Biophys. 300*, 223–230.
44. Jörnvall, H., and Persson, B. (1983) Amino acid sequence restriction in relation to proteolysis. *Biosci. Rep. 3*, 225–232.
45. Vanhoof, G., Goossens, F., Demeester, I., Hendriks, D., and Scharpé, S. (1995) Proline motifs in peptides and their biological processing. *FASEB J. 9*, 736–744.
46. Markert, Y., Köditz, J., Ulbrich-Hofmann, R., and Arnold, U. (2003) Proline versus charge concept for protein stabilization against proteolytic attack. *Protein Eng. 16*, 1041–1046.

BI800655U