PROTEINS: Structure, Function, and Genetics 41:257–269 (2000)

Comparative Modeling of the Three-Dimensional Structures of Family 3 Glycoside Hydrolases

Andrew J. Harvey,[1] Maria Hrmova,[1] Ross De Gori,[2] Joseph N. Varghese,[2] and Geoffrey B. Fincher[1]*
[1]*Department of Plant Science, University of Adelaide, Waite Campus, Glen Osmond, South Australia, Australia*
[2]*Biomolecular Research Institute, Parkville, Victoria, Australia*

ABSTRACT **There are approximately 100 known members of the family 3 group of glycoside hydrolases, most of which are classified as β-glucosidases and originate from microorganisms. The only family 3 glycoside hydrolase for which a three-dimensional structure is available is a β-glucan exohydrolase from barley. The structural coordinates of the barley enzyme is used here to model representatives from distinct phylogenetic clusters within the family. The majority of family 3 hydrolases have an NH_2-terminal $(\alpha/\beta)_8$ barrel connected by a short linker to a second domain, which adopts an $(\alpha/\beta)_6$ sandwich fold. In two bacterial β-glucosidases, the order of the domains is reversed. The catalytic nucleophile, equivalent to D285 of the barley β-glucan exohydrolase, is absolutely conserved across the family. It is located on domain 1, in a shallow site pocket near the interface of the domains. The likely catalytic acid in the barley enzyme, E491, is on domain 2. Although similarly positioned acidic residues are present in closely related members of the family, the equivalent amino acid in more distantly related members is either too far from the active site or absent. In the latter cases, the role of catalytic acid is probably assumed by other acidic amino acids from domain 1. Proteins 2000;41:257–269.** © 2000 Wiley-Liss, Inc.

Key words: catalytic amino acids, domains, phylogenetic tree, protein folds

INTRODUCTION

Glycoside hydrolases are widely distributed in nature, where they catalyze the hydrolysis of glycosidic linkages in glycosides, oligosaccharides, and polysaccharides. The enzymes have been classified traditionally on the basis of their substrate specificity, according to the EC numbering system of the International Union of Biochemistry and Molecular Biology.[1] However, many glycoside hydrolases are able to hydrolyze several substrates and some exhibit a very broad specificity. For example, the barley β-glucan exohydrolase described here is able to hydrolyze synthetic aryl β-glycosides such as 4-nitrophenyl β-D-glucoside and might, therefore, be classified as a β-glucosidase in the EC 3.2.1.21 group.[2] Nevertheless, the preferred substrates for the barley β-glucan exohydrolase are (1→3)-β-glucans, from which the enzyme releases nonreducing terminal β-glucosyl[3]; this specificity would place the enzyme in the EC 3.2.1.58 group.

Although substrate specificity will remain a cornerstone of enzyme classification, additional criteria are clearly needed to identify evolutionary and functional relationships between individual glycoside hydrolases. Structural information is increasingly used to discriminate between enzymes or, conversely, to classify them into related groups. Thus, glycoside hydrolase classifications based on linear, one-dimensional (1D) amino acid sequences were initially proposed to take into account primary protein structures[4] ,and these were subsequently supplemented with hydrophobic cluster analysis (HCA).[5] The HCA analyses provide 2D structural information and are particularly useful where amino acid sequence identities between proteins are low.[6] The 1D and 2D structural information has been applied to the glycoside hydrolases, which fall into around 80 different families under this classification scheme[7,8] (http://afmb.cnrs-mrs.fr). Crystallographic data are available for representatives from almost half of the families, and it has become evident that members of each family share a common protein fold.[7] It should be emphasized that classification into a single structural family does not suggest a common substrate specificity or function. For example, barley (1→3)-β-glucan endohydrolases (EC 3.2.1.39) and (1→3,1→4)-β-glucan endohydrolases (EC 3.2.1.73) share highly similar 3D structures and are both classified with the family 17 group of glycoside hydrolases.[9] However, a small number of amino acid substitutions along the substrate-binding cleft have resulted in quite distinct substrate specificities and in dramatic differences in the biologic functions of the two enzymes.[10]

Related to the application of structural criteria in the classification of glycoside hydrolases is the identification and classification of these structures solely from amino acid sequence information. Genomics programs in organisms as diverse as yeast, rice, and humans are generating nucleotide sequence data at an unprecedented rate, and it is likely that genes encoding previously unknown glycoside hydrolases will soon be represented in these genome sequencing programs. The current challenge is to recognize individual genes and to determine the functions of

Grant sponsor: Australian Research Council; Grant sponsor: Grains Research and Development Corporation.

Drs. Harvey and Hrmova have contributed equally to this work.

*Correspondence to: Geoffrey B. Fincher, Department of Plant Science, University of Adelaide, Waite Campus, Glen Osmond, South Australia 5064, Australia. E-mail: gfincher@waite.adelaide.edu.au

Received 4 January 2000; Accepted 15 June 2000

their expressed products. Again, alignment of primary structures of unknown enzymes with DNA and protein databases will allow the identification of many genes, and the use of 2D HCA procedures will enhance this process, especially where sequence similarities are low. Furthermore, proteins and enzymes can now be identified rapidly in high throughput genomics programs by automated, comparative 3D structural modeling[11] (http://guitar.rockefeller.edu/modbase).

The identification of glycoside hydrolases through 3D structural models and their classification into family groups rely on the availability of at least one experimentally determined 3D structure per family. Until recently, no 3D structures were available for family 3 glycoside hydrolases, which currently include around 100 known enzymes that are variously designated as β-glucosidases, *N*-acetyl β-glucosaminidases and β-xylosidases (http://afmb.cnrs.mrs.fr). However, the structure of a family 3 β-glucan exohydrolase from barley has now been solved to 2.2 Å resolution by X-ray crystallography.[12,13] In the present work, we use comparative protein modeling of 14 family 3 enzymes to examine domain folds within this group of glycoside hydrolases, to define the disposition of catalytic amino acids and structures of active sites, and to determine local variations in structures across the family.

EXPERIMENTAL PROCEDURES

Evolutionary Analysis

A total of 99 members of the family 3 group of glycoside hydrolases, as classified on the Carbohydrate-Active enZYmes (CAZY) server[14] (http://afmb.cnrs-mrs.fr/~pedro; as at June 1, 2000), were aligned by using the PileUp program. An unrooted, radial phylogenetic tree was constructed by using the EprotPars program of the University of Wisconsin Genetics Computers Group (GCG) package.[15]

The accession numbers at the GenBank/EMBL and SWISS-PROT databases of the 99 family 3 glycoside hydrolases are as follows: *Acetobacter xylinus* AB010645 (a), *Acetobacter xylinus* AB015802 (b), *Agrobacterium tumefaciens* M59852, *Ajellomyces capsulatus* U20346, *Alteromonas sp* D17399, *Arabidopsis thaliana* AL353995 (a), AL133292 [ORF F13I12.100] (b), AL133292 [ORF F13I12.50] (c), AL133292 [ORF F13I12.60] (d), AL133292 [ORF F13I12.90] (e), AL353994 (f), AL162651 (g), AC009243 (h), *Aspergillus aculeatus* D64088, *Aspergillus kawachi* AB003470, *Aspergillus nidulans (Emericella nidulans)* Y13568, *Aspergillus niger* Z84377, *Aspergillus oryzae* AB009972, *Aspergillus wentii* P29090, *Azospirillum irakense* AF090429 β-glucosidase (SalA) (a), *Azospirillum irakense* AF090429 β-glucosidase (SalB) (b), *Bacillus sp*. AB009411, *Bacillus subtilis* L19954, *Bacteroides fragilis* AF006658, *Borrelia burgdorferi* AE001115 (a), *Borrelia burgdorferi* AE001163 (b), *Botryotinia fuckeliana* AJ130890, *Butyrivibrio fibrisolvens* M31120, *Cellulomonas biazotea* AF005277, *Cellvibrio gilvus* D14068, *Chryseobacterium meningosepticum* AF015915, *Clostridium stercorarium* Z94045, *Clostridium thermocellum* X15644, *Coccidioides immitis* U87805 (a), *Coccidioides immitis* AF022893 (b), *Cochliobolus heterostrophus* AF027687, *Debaryomyces hansenii* AJ223815, *Deinococcus radiodurans* AE001979, *Dictyostelium discoideum* L21014, *Erwinia chrysanthemi* U08606, *Escherichia coli* U15049 (a), *Escherichia coli* AE000211 (b), *Gaeumannomyces graminis* U35463 (a), *Gaeumannomyces graminis* U17568 (b), *Glycine max* AF000378, Haemophilus *influenza* U32777, *Hordeum vulgare* β-glucan exohydrolase isoenzyme Exo1 AF102868, *Hordeum vulgare* β-glucan exohydrolase isoenzyme Exo2 U46003, *Kluyveromyces fragilis* X05918, *Listeria monocytogenes* U78883, *Microbispora bispora* L06134, *Mycobacterium tuberculosis* AL021929 (a), Z97050 (b), *Neisseria meningitidis* AE002408 (a), *Neisseria meningitidis* AL162754 (b), *Neurospora crassa* AL355929, *Nicotiana tabacum* AB017502, *Phanerochaete* avenaria AJ276675, Phanerochaete *chrysosporium* AF036872, *Pichia anomala (Candida pelliculosa)* X02903, *Prevotella ruminicola* U35425, *Pseudomonas aeruginosa* U56077, *Pseudomonas fluorescens* X65527, *Ruminococcus albus* U92808 (a), *Ruminococcus albus* X15415 (b), *Ruminococcus flavefaciens* AJ132472, *Saccharomycopsis fibuligera* M22475 (a), M22476 (b), *Saccharopolyspora erythraea* Y14327, *Salmonella typhimurium* D86507, *Schizophyllum commune* M27313, *Schizosaccharomyces pombe* AL355920, *Septoria lycopersici* U24701, *Streptomyces antibioticus* AF055579, *Streptomyces coelicolor* AL121596 (a), AL136519 (b), AL023702 (c), AL352956 (d), AL117385 (e), AL355913 (f), AL031013 (g), *Streptomyces lividans* AF043654, *Streptomyces thermoviolaceus* AB008771, *Streptomyces venezuelae* AF079762, *Synechocystis sp.* D90914, *Thermoanaerobacter brockii* Z56279, *Thermoanaerobacter ethanolicus* AF135015, *Thermotoga maritima* AE001690 (a), AE001694 (b), AE001748 (c), *Thermotoga neapolitana* Z77856 (a), Y17983 (b), U58632 (c), *Trichoderma reesei (Hypocrea jecorina)* U09580 (a), Z69257 (b), *Tropaeolum majus* AJ006501, unidentified bacterium U12011, *Vibrio furnissii* U52818, and *Zea mays* AF064707. Lower case letters after accession numbers indicate individual members when multiple entries of a species are listed.

Analysis of Domain Arrangements

Analyses of all the family 3 glycoside hydrolases were effected through the ProDom 99.1 server[16] (http://www.toulouse.inra.fr/multalin.html), by using the sequence of barley β-glucan exohydrolase isoenzyme ExoII. The automated program compares the nonfragmentary sequences with the PSI-BLAST tool,[17] taken from the SWISS-PROT database, and the resultant families are further processed with the MKDOM program.[18] The original output is filtered to yield nonredundant similarities.

Protein Structure Modeling

Comparative (homology) protein modeling was performed by using the Modeler version 4.0 program,[19] which uses satisfaction of spatial restraints and statistical analysis of the relationships between pairs of homologous structures to produce the protein fold. The CHARM 22 energy function was implemented during energy minimization to produce an objective function. The resultant model was refined by using a conjugate gradient and molecular

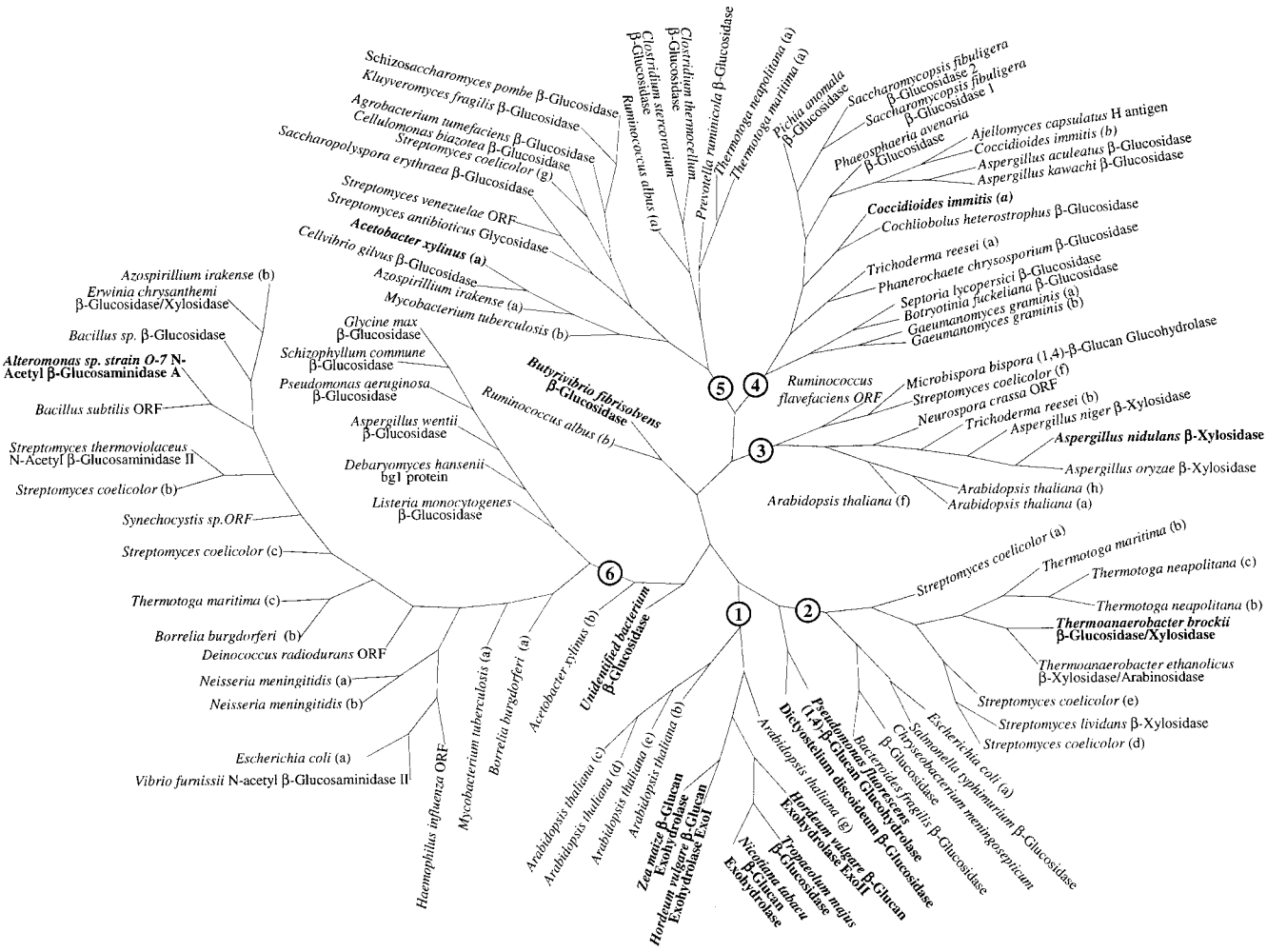

Fig. 1. Unrooted radial phylogenetic tree of family 3 glycoside hydrolases. The sequences were aligned by using the PileUp program, and the tree was constructed with the EprotPars program of the University of Wisconsin GCG package.[15] The 99 sequences (for accession numbers see Materials and Methods section) are subdivided into six major branches, or clusters. The representatives that have been modeled are shown in bold typeface. To obtain good alignments for the bacterial β-glucosidases from *Butyrivibrio fibrisolvens*[30] and *Ruminococcus albus* (K. Ohmiya, unpublished results, GenBank/EMBL accession number X15415), it was necessary to change the order of the NH_2-terminal and COOH-terminal domains. Lower case letters after accession numbers indicate individual members when multiple entries of a species are listed. All family 3 glycoside hydrolases listed in the databases[14] on June 1, 2000 are included in this Figure, which is available on the Web at http://planta.waite.adelaide.edu.au/labs/gbf/modeling.htm.

dynamics with simulated annealing.[19] The first step of the experimental approach requires the identification of a known 3D structure (template) related to the target sequences of family 3 glycoside hydrolases and uses the coordinates of the template protein as a basis for further modeling. The template in this case was the barley β-glucan exohydrolase isoenzyme ExoI.[13] The second step in the process of modeling is the alignment of the template with the target sequence. This was done with the Bestfit program, by using the implemented gap penalty function and the local homology algorithm of Smith and Waterman[20] in the University of Wisconsin Genetics Computer Group (GCG) package,[15] with all other parameters unchanged from the default settings. During the sequence alignments, no correlations for loops connecting secondary structural elements were introduced. In some instances the Swiss-Pdb Viewer's alignment tool[21] was used and the resultant alignments were found to be similar. In the case of the *Aspergillus nidulans* β-xylosidase,[22] the alignment with the template sequence was also effected with the ClustalW algorithm[23] (http://www2.ebi.ac.uk/clustalw), by using the Blosum matrix. In the final step, 14 protein models, based on the template-target alignments, were constructed on a Silicon Graphics model O2 computer, running IRIX 6.4.

Refinement of Selected Models

The energy configurations outputs from the Modeler program were used as starting structures for optimizations minimized with the X-PLOR program.[24] Because the energy minimizer in the Modeler program[19] distorted the aromatic rings in some of the phenylalanine, tryptophan, and tyrosine residues, it was necessary to replace the distorted rings in each of the starting structures with

TABLE I. Protein Sequence Identities and Similarities, Root Mean Square Deviation Values, Overall G-Factors and Ramachandran Plot Statistics of the Template Barley β-Glucan Exohydrolase Isoenzyme ExoI and Target Sequences Selected From Family 3 Glycoside Hydrolases

Entry[a]	Identity (%) Bestfit[b]	Similarity (%) Bestfit[b]	RMSD[c] Å (aa)	G-factor PROCHECK[d]	Ramachandran values[e] % Allowed	Disallowed
Barley Exo1	100	100	0	−0.28	99.8	0.2
Barley Exo2	72	82	0.50 (596)	−0.15	99.4	0.6
Maize	83	91	0.25 (598)	−0.04	99.6	0.4
Nasturtium	70	82	0.47 (599)	−0.18	99.4	0.6
Tobacco	71	84	0.53 (601)	−0.08	99.4	0.6
Acetobacter	26	49	1.10 (495)	−0.64	95.8	4.2
Alteromonas	22	48	0.96 (502)	−0.46	96.2	3.8
Aspergillus	29	51	1.02 (527)	−0.56	95.3	4.7
Butyrivibrio $(\alpha/\beta)_8$	25	49	0.20 (278)	−0.52	97.1	2.9
Butyrivibrio $(\alpha/\beta)_6$	22	50	1.39 (166)	−0.74	94.6	5.4
Coccidioides	28	51	1.42 (495)	−0.75	96.4	3.6
Dictyostelium	33	55	0.91 (577)	−0.51	97.6	2.4
Pseudomonas	45	64	0.64 (579)	−0.27	98.6	1.4
Thermoanaerobacter	33	56	0.76 (510)	−0.42	96.1	3.9
Unidentified bacterium	36	60	0.96 (320)	−0.25	98.3	1.7
Barley β-glucosidase (family 1)	19	44	1.83 (317)	−1.05	94.6	5.4

[a]Representatives were chosen from different branches on the unrooted phylogenetic radial tree.
[b]Using Bestfit program of the University of Wisconsin GCG package.[15]
[c]Root mean square deviations (RMSD) of the C^{α} backbone were determined for the Modeler pdb files using program O.[25] Numbers in brackets (aa) indicate the number of amino acid residues aligned to generate the RMSD.
[d]Overall G-factors of models[29] calculated by PROCHECK.[28]
[e]Ramachandran plot statistics excluding Gly and Pro residues.

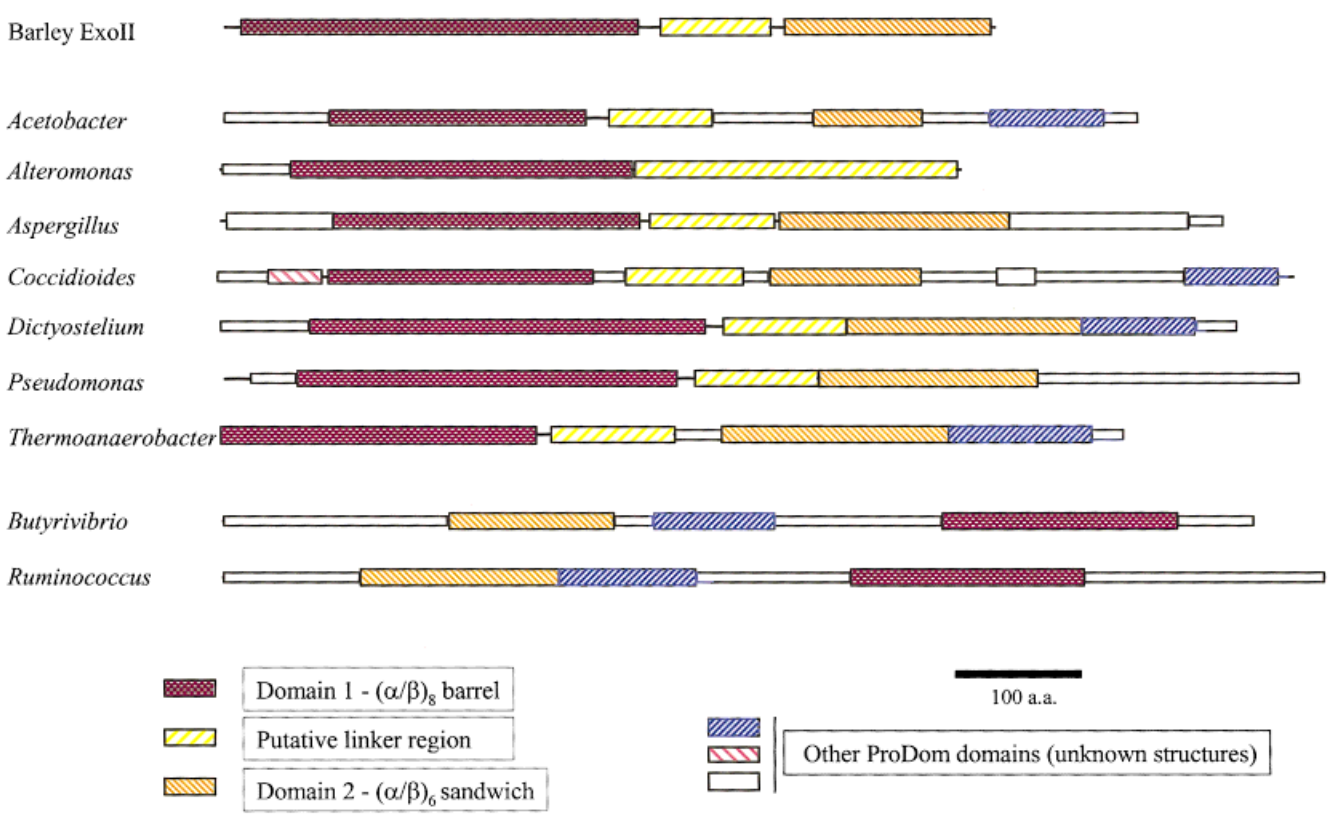

Fig. 2. Schematic display of domain arrangements of selected members of family 3 glycoside hydrolases. Entries available through the SWISS-PROT database were analyzed with the automated ProDom server by using the BLAST, MKDOM, and MultAlin programs.[16] Note that 2 members of the group (*Butyrivibrio* and *Ruminococcus*) show reverse orientation of domains 1 and 2, in comparison with the barley β-glucan exohydrolase isoenzyme ExoII. The amino acid sequence of barley isoenzyme ExoII has 72% identity with that of barley β-glucan exohydrolase isoenzyme ExoI.

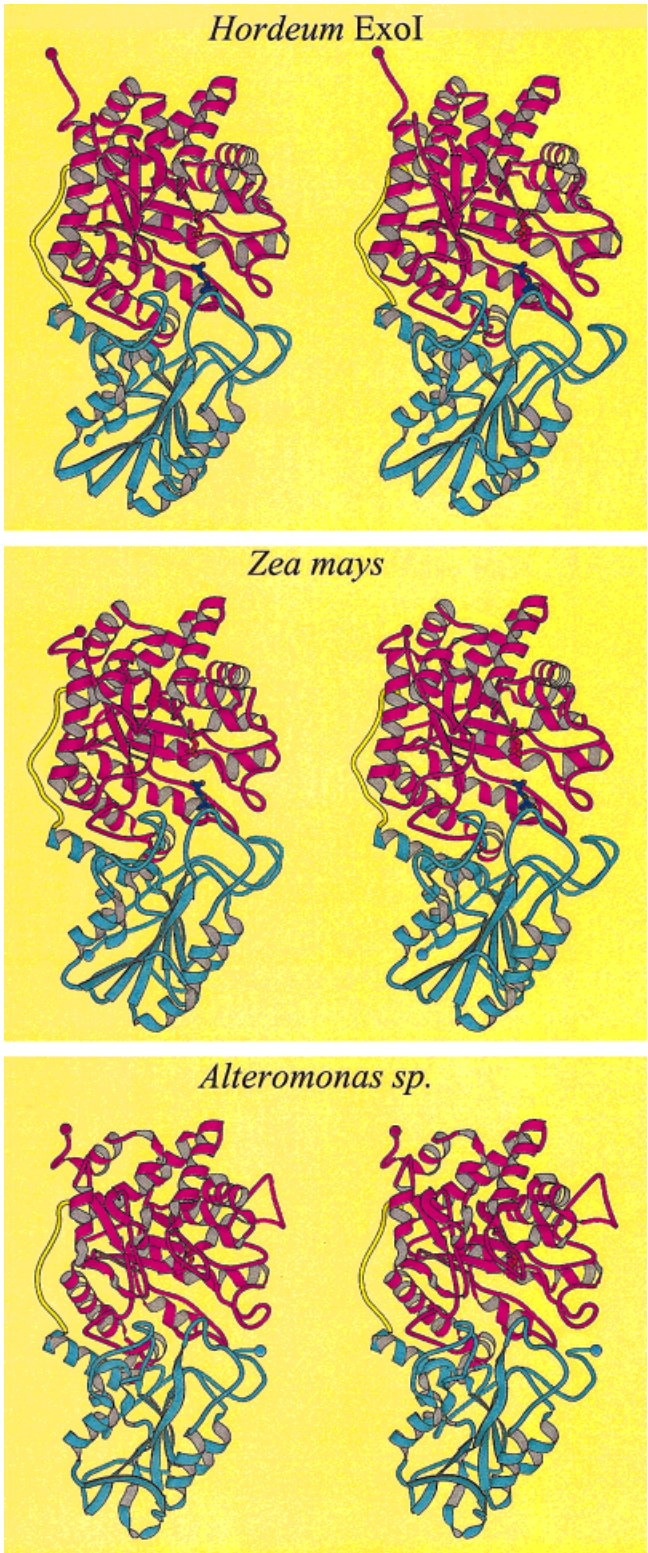

Fig. 3. Stereoview ribbon representation of barley β-glucan exohydrolase isoenzyme ExoI and modeled structures of *Zea mays* β-glucan exohydrolase and *Alteromonas sp. N*-acetyl β-D-glucosaminidase A. Domain 1, domain 2, and the linker of each enzyme are coloured in magenta, cyan, and yellow, respectively. The catalytic nucleophiles and catalytic acids are shown in red and blue, respectively. No other catalytic acid is required in the *Alteromonas* enzyme. The figure was generated with MOLSCRIPT.[54]

planar aromatic rings, by using the program O.[25] The distorted side chains were replaced with rotamer library side chains, by using the lego_sidechain module of program O, before any molecular dynamics simulations were performed. Optimizations were always done with a constrained C^{α} backbone unless otherwise stated. Energy minimization by using X-PLOR was done in several stages with the starting structure coordinates of one relaxed system taken from the end point of a previous protein optimization. For each dynamics run, Verlet parameters[26] were used with a time step for a finite difference integration of 10^{-3} ps for 2,000 cycles and a decrease of 100 steps at a time. The mode of initial velocity was determined by the Maxwell method with the frequency of velocity rescaled every 25 steps. Temperatures were subsequently cooled from 2500 to 300°K. Backbone constraints were removed, and the model allowed to relax further by 5,000 cycles through the same dynamics setup described above, except that a constant temperature of 300°K was used. The model was finally refined for 10,000 cycles by using the conjugate gradient method.[27] The stereochemical properties of each model were examined after each minimization step was complete, to ensure that the models did not deviate significantly from ideality.

Evaluation of Models

The stereochemical quality of models was evaluated by using PROCHECK.[28] Overall G-factors, which are measures of normality of main chain bond lengths and bond angles obtained from the small molecule library of Engh and Huber,[29] were calculated from PROCHECK. The root mean square deviation (RMSD) values in the C^{α} positions between barley β-glucan exohydrolase isoenzyme ExoI and the modeled structures of family 3 glycoside hydrolases were determined with the program O.[25]

RESULTS

Phylogenetic Tree

The amino acid sequences of 99 members of the family 3 group of glycoside hydrolases were aligned and an unrooted, radial phylogenetic tree was constructed (Fig 1; http://planta.waite.adelaide.edu.au/labs/gbf/modeling-.htm). Amino acid sequence identities between the barley β-glucan exohydrolase isoenzyme ExoI and other family 3 members vary from 19 to 83%, although it must be emphasized that not all sequences are full-length. Approximately 65% of the family members are 20–30% identical with the barley enzyme and approximately 21% fall in the 30–40% identity range.

For subsequent analysis and modeling, the phylogenetic tree (Fig. 1) was divided into six major branches of sequences and 14 members were modeled (Table I). The majority of higher plant enzymes are closely related and are located in a cluster on one branch of the tree. Members of this group exhibited amino acid sequence identities of between 57% and 83%, compared with the sequence of barley β-glucan exohydrolase isoenzyme ExoI. Most of the other major branches of the tree included representatives from bacteria, yeast and fungi (Fig. 1). To obtain good

alignments for the bacterial β-glucosidases from *Butyrivibrio fibrisolvens*[30] and *Ruminococcus albus* (K. Ohmiya, unpublished results, GenBank/EMBL accession number X15415), it was necessary to change the order of the NH_2-terminal and COOH-terminal domains, because in these two enzymes there is a reversal of the two domains compared with their order in the barley β-glucan exohydrolase isoenzyme ExoI (see section on *Domain Reversal in Two Bacterial Enzymes* below). However, another β-glucosidase from *Ruminococcus albus* has its two domains arranged in the same order as the barley enzyme (Fig. 1; *Ruminococcus albus* [a]).

Domain Structures

The arrangements of protein domains in selected enzymes from the family 3 glycoside hydrolases were further investigated by using the ProDom server.[16] The enzymes were selected so that each of the major branches on the phylogenetic tree was represented and include β-glucosidases from fungal and bacterial sources, β-glucan exohydrolases from bacteria and higher plants, a bacterial *N*-acetyl β-glucosaminidase, a β-glucosidase from *Dictyostelium discoideum*, and a β-glucosidase from a thermophilic bacterium. Some of the selected enzymes are shown in bold type in Figure 1, and possible domain arrangements assigned by the ProDom program are represented diagrammatically in Figure 2.

The barley β-glucan exohydrolases are known to have an NH_2-terminal domain (domain 1) that adopts an $(\alpha/\beta)_8$ TIM barrel fold and is connected by means of a 16-amino acid linker to a second domain, which forms an $(\alpha/\beta)_6$ sandwich (domain 2); a 42 amino acid antiparallel loop is located at the COOH-terminus.[13] The ProDom program detected the two major domains in all the sequences analyzed and indicated that in each enzyme their lengths were similar (Fig. 2). The $(\alpha/\beta)_8$ barrel domain is usually separated from the $(\alpha/\beta)_6$ sandwich domain by a 'linker' region of variable length and other possible domains of unknown structure are also predicted by the program. As mentioned above, in the bacterial β-glucosidases from *Butyrivibrio* and *Ruminococcus* domains 1 and 2 are reversed in order, compared with their arrangements in the other enzymes (Fig. 2). Because the ProDom program is based on sequence alignments that indicate likely domain boundaries, rather than on 3D structural information, the data presented in Figure 2 must be interpreted with some caution. Nevertheless, the program does predict that almost all the family 3 glycoside hydrolases examined have the two major protein domains, that both of these are similar in length to the equivalent domains in the barley enzyme, and that in two bacterial enzymes domain 2 precedes domain 1 (Fig. 2). In some instances, such as the *Alteromonas N*-acetyl β-glucosaminidase, the program could not detect the presence of domain 2 in the sequence. This could be related to the absence of a clear signature that is required by the program to delineate the domain boundary, rather than the absence of blocks of amino acid residues equivalent to those in domain 2 of the barley β-glucan exohydrolase (Fig. 4).

Protein Modeling

A total of 14 models were built from the amino acid sequences of family 3 glycoside hydrolases, by using the Modeler program.[19] With the exception of some *Arabidopsis* members, the known plant enzymes were modeled, together with the fungal and bacterial representatives indicated in Figure 1.

The amino acid sequence similarities of the selected enzymes, the RMSD values for modeled C^α chains, the PROCHECK program G-factors, which measure how reliable the models are, and the percentage of "allowed" amino acid residues from Ramachandran plots[28] are compared in Table I. Amino acid sequence identities of the modeled enzymes, compared with barley β-glucan exohydrolase isoenzyme ExoI, range from approximately 22% to more than 80% (Table I). The plant enzymes share the highest degree of amino acid sequence identity, at 70–83% identity, but the amino acid sequences of the other selected enzymes are only 22% to 45% identical when aligned with the barley enzyme (Table I).

The RMSDs in C^α positions between the modeled enzymes and the barley β-glucan exohydrolase template range from 0.25 Å to more than 1.4 Å, over 500–600 amino acid residues (Table I). These values may be compared with an RMSD of 0.65 Å in C^α positions for two very closely related family 17 β-glucan endohydrolases from barley.[9] The predicted 3D structure of the maize β-glucan exohydrolase is more similar to barley β-glucan exohydrolase isoenzyme ExoI than is the other barley enzyme, isoenzyme ExoII (Table I). A model of the negative control enzyme, the family 1 β-glucosidase from barley,[31] shows that the sequence of this β-glucosidase can be "forced" onto the β-glucan exohydrolase structure, but that the RMSD is more than 1.8 Å (Table I).

The values for RMSDs in C^α positions between the models and the template β-glucan exohydrolase isoenzyme ExoI are also reflected in the overall G-factors and Ramachandran values (Table I). The G-factors are a measure of model reliability, in which more positive values generally indicate better reliability; values of less than −1.0 indicate an unreliable model.[25,29] Similarly, Ramachandran plot statistics which place 95–97% or more of the amino acid residues in "allowed" positions are considered reliable in modeling experiments, and this indicates how well the structures fit the expected main chain length and torsion angle distribution.[28,32–34] Higher numbers of residues in the disallowed region reflect a distorted geometry or steric clashes in the models, because there are higher proportions of residues falling outside the limits of main chain bond length and torsion angles of the small molecule library.[29]

Table I shows that correlations between sequence similarity, RMSD of C^α chains, G-factors or Ramachandran values could be detected, and some general trends were evident (Table I). The higher plant enzyme models were the most reliable, and it can be concluded that all of these

modeled enzymes are likely to adopt very similar protein folds (Table I). This is consistent with their position on a single branch of the phylogenetic tree (Fig. 1). However, some of the *Arabidopsis* entries, specifically those designated "β-xylosidase-like," are positioned on other branches of the phylogenetic tree (Fig. 1). Enzymes that are relatively close to the higher plant enzymes on the phylogenetic tree, in particular the β-glucosidases from *Dictyostelium*,[35] *Thermoanaerobacter*,[36] and an unidentified bacterium,[37] and the (1→4)-β-glucan glucohydrolase from *Pseudomonas*,[38] also appear to be most similar in 3D structure with the barley β-glucan exohydrolase (Table I; Fig. 1). Conversely, the *Alteromonas N*-acetyl β-glucosaminidase[39] and the *Coccidioides* β-glucosidase (JJ Yu, PW Thomas, K Seshan, GT Cole; unpublished results in GenBank/EMBL entry U87805) are located relatively further from the barley enzyme on the phylogenetic tree (Fig. 1), and their 3D structures are less similar to the barley enzyme, as indicated by the lower overall G-factors and Ramachandran plot values (Table I). Despite these differences, it remains likely that the family 3 glycoside hydrolases consist of an $(\alpha/\beta)_8$ TIM barrel domain and an $(\alpha/\beta)_6$ sandwich domain.

To illustrate the range of molecular conformations in the family 3 group, the barley β-glucan exohydrolase isoenzyme ExoI structure is compared with models of the most closely-related β-glucan exohydrolase, which is from maize, and the most distantly related enzyme, which is an *N*-acetyl β-glucosaminidase from the marine bacterium *Alteromonas*[39] (Fig. 3). For comparative purposes, the amino acid sequence alignments of the same three enzymes are shown in Figure 4, together with the secondary structural elements of the barley β-glucan exohydrolase isoenzyme ExoI.

Inserted and Deleted Blocks of Amino Acids

The most noteworthy differences between the plant and the bacterial enzymes are the insertions and deletions of blocks of up to 13 amino acid residues, and the observation that approximately half of the COOH-terminal loop is missing in the bacterial enzyme (Fig. 4). The inserted or deleted groups of amino acid residues are generally located in surface loop regions of the protein fold or near the ends of α-helices (Fig. 4). Furthermore, there seem to be more surface insertions and deletions in the first 300–350 amino acid residues, which constitute domain 1 of the enzymes (Figs. 3 and 4). Similar effects are observed when the model of the (1→4)-β-glucan glucohydrolase from *Pseudomonas*[38] is compared with the 3D structure of the barley β-glucan exohydrolase (Fig. 5), where insertions are shown in red and the positions of deletions are indicated in blue. In all cases, the insertions and deletions in the models are treated as loops.

Domain Reversal in Two Bacterial Enzymes

As noted in Figure 2, the two domains in the β-glucosidases from *Butyrivibrio* and *Ruminococcus* are arranged in the reverse sequence, compared with all the other members of the family. Thus, the $(\alpha/\beta)_6$ sandwich domain precedes the $(\alpha/\beta)_8$ barrel in these two enzymes. Although the reliability of the $(\alpha/\beta)_6$ sandwich domain of the *Butyrivibrio* enzyme model is relatively poor, the other domain models well (Table I). It was possible to generate a composite model, in which partial models of the two domains could be arranged so that the overall shape of the enzyme was similar to that of the barley β-glucan exohydrolase (Fig. 6). It is not clear exactly how the two domains are connected, but there are 96 amino acid residues between the COOH-terminal end of the $(\alpha/\beta)_6$ sandwich domain and the NH_2-terminus of the $(\alpha/\beta)_8$ barrel domain; this would be easily long enough to connect the two domains or even to insert a new subdomain into the vacant gap (Fig. 6). However, the absence of an appropriate template that corresponded to this 96-residue sequence precluded further modeling of regions between the two domains.

Catalytic Amino Acids

Careful examination of the 3D structure,[13] together with studies of the effects of active site-specific inhibitors (M. Hrmova, J.N. Varghese, H. Driguez, and G.B. Fincher, unpublished), indicate that the catalytic nucleophile and the catalytic acid of barley β-glucan exohydrolase isoenzyme ExoI are D285 and E491, respectively. Sequence alignments indicated that the position of the catalytic nucleophile (D285 or its equivalent) is absolutely conserved in all family 3 glycoside hydrolases. In the plant enzymes the catalytic nucleophile is located in a highly conserved **GFVISDW** sequence, and is surrounded by a landscape of amino acid residues with similar characteristics in other family 3 enzymes (Table II). Molecular modeling and HCA procedures indicate that the catalytic nucleophile is always found near the COOH-terminus of β-strand "g" in domain 1 (Fig. 4), and that this β-strand is

Fig. 4. (Overleaf.) Amino acid sequence alignments of family 3 glycoside hydrolases. The amino acid sequence of barley β-glucan exohydrolase isoenzyme ExoI (AJ Harvey, M Hrmova, and GB Fincher, unpublished data, GenBank/EMBL accession number AF102868) is aligned with related enzymes from maize (YY Zhao and L Bogorad 1999, unpublished results, GenBank accession number AF064707) and *Alteromonas sp.*[39] The sequences were aligned by using the PileUp program. Green areas highlight identical residues, and yellow areas represent regions of high conservation. Open arrowheads point to the catalytic residues, and black arrowheads point to the *N*-glycosylation sites. Numbers below the sequences indicate amino acid residues of the mature barley β-glucan exohydrolase isoenzyme ExoI. Secondary structural elements of domain 1 of β-glucan exohydrolase isoenzyme ExoI are shown as purple cylinders (α-helices) and purple horizontal arrows (β-strands), connected by purple lines (coils). Cyan cylinders (α-helices) and cyan horizontal arrows (β-strands), connected by cyan lines (coils), indicate secondary structural elements of the domain 2 $(\alpha/\beta)_6$ sandwich. Yellow and red lines correspond to the linker and the COOH-terminal antiparallel loop, respectively. The figure was prepared by using ALSCRIPT.[55]

Fig. 5. (Overleaf.) Stereoview ribbon representation of modeled structure of *Pseudomonas fluorescens* (1→4)-β-glucan glucohydrolase, showing positions of amino acid deletions and insertions. The *Pseudomonas* enzyme structure[38] was modeled by using the barley β-glucan exohydrolase isoenzyme ExoI as a template. Domains 1 and 2 are coloured in white and grey, respectively. Residues that are inserted in the *Pseudomonas* enzyme but are absent from the barley enzyme are coloured red, and amino acid residues flanking deletions are in blue. Deletions and insertions are generally located on surface loops. The figure was generated with MOLSCRIPT.[54]

Figure 4. (Legend on previous page.)

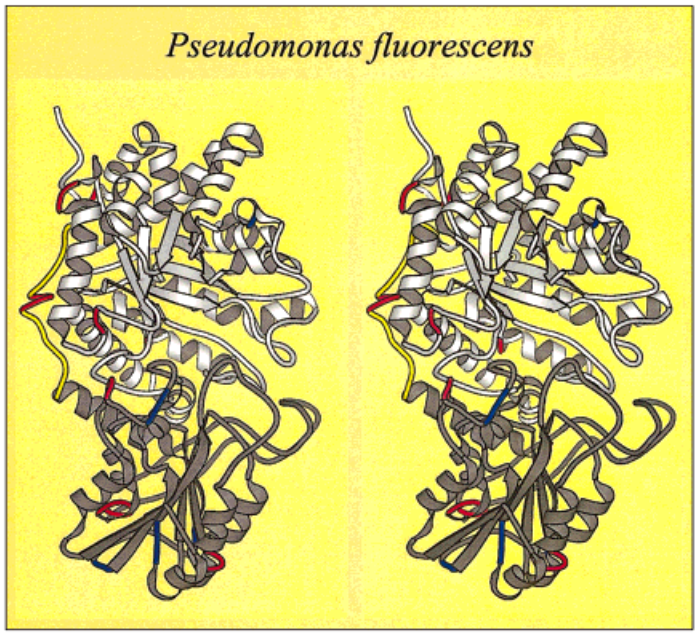

Figure 5. (Legend on previous page.)

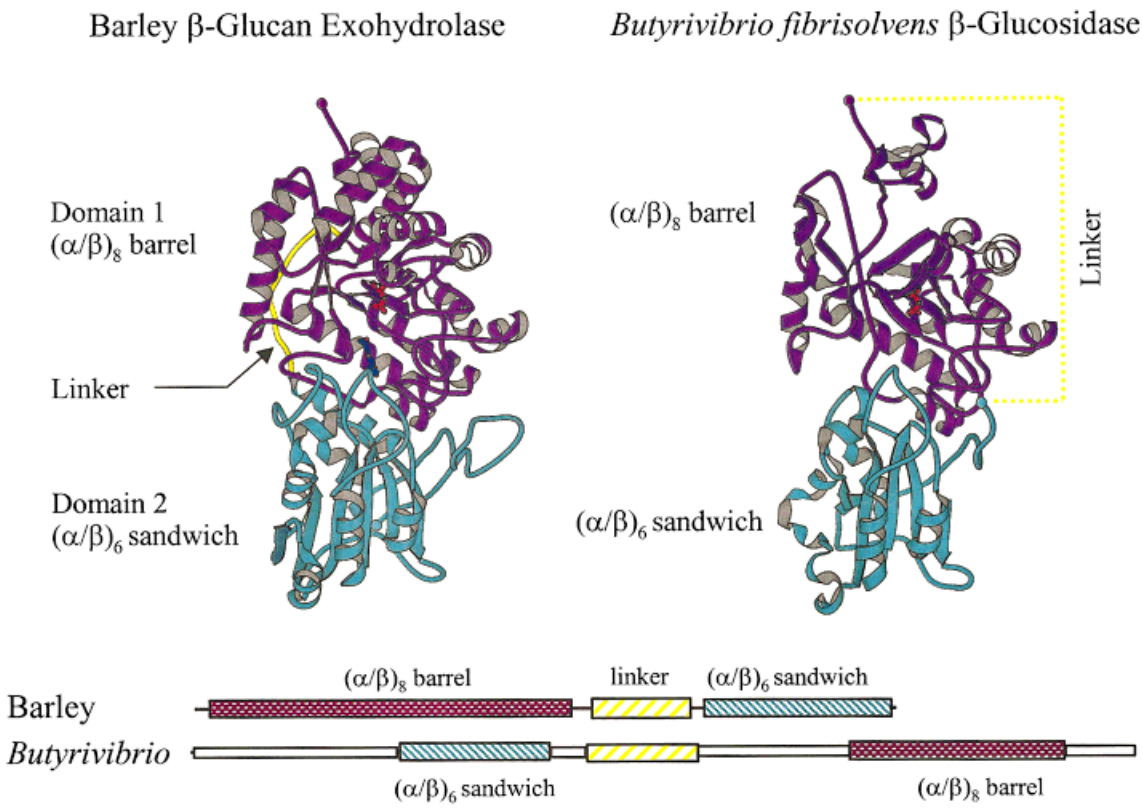

Figure 6. (Legend on following page.)

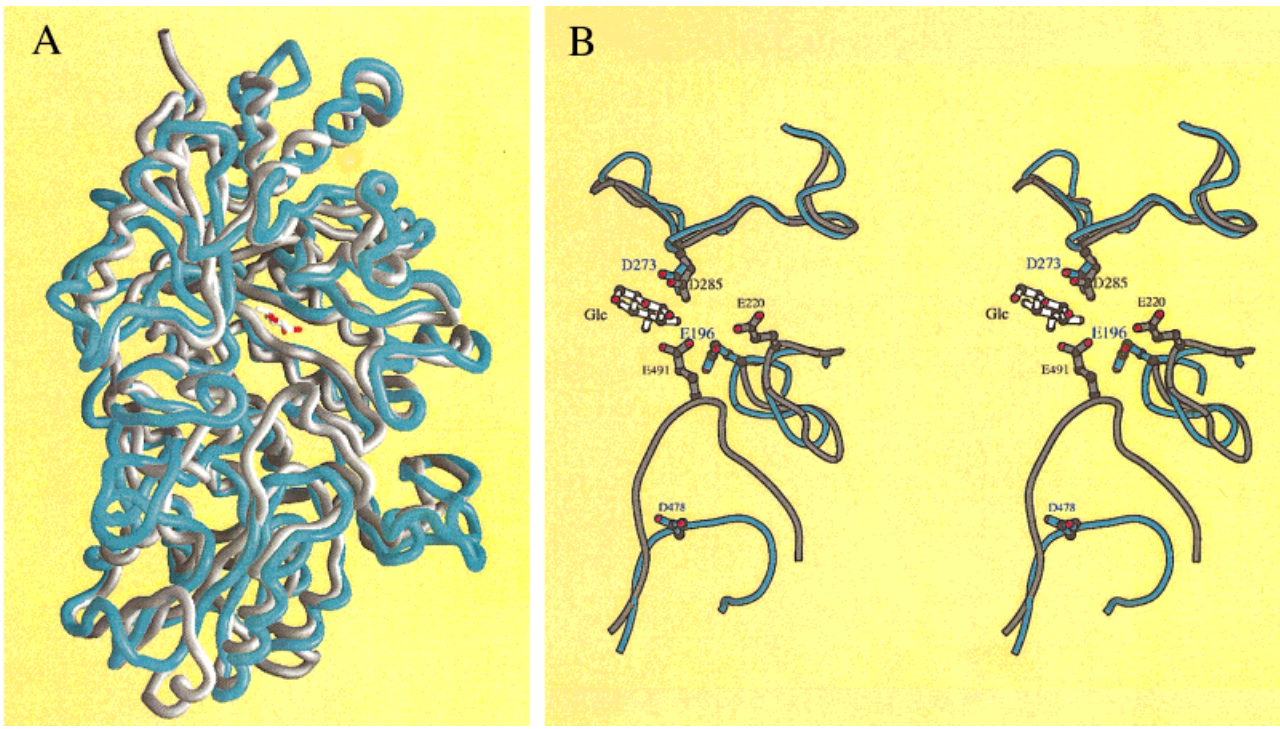

Figure 7. (Legend on following page.)

always positioned in the shallow active site pocket that is located at the interface of the two domains.[13]

The equivalent amino acid residue to the putative catalytic acid of the barley enzyme, E491, is highly conserved in the plant enzymes but is sometimes difficult to detect in more distantly related members of the family (Table II). The catalytic acid in barley is positioned on the $(\alpha/\beta)_6$ sandwich domain in a loop region that varies across the family (Figs. 3 and 6; Table II). A potentially equivalent acidic amino acid residue is often detected on that loop of domain 2, but models suggest that it is located further from the catalytic nucleophile as the particular enzyme diverges from the barley enzyme. For example, in the cell-wall associated β-xylosidase from *Aspergillus nidulans*,[22] for which the C^α chain is superimposed on the C^α chain of the barley enzyme in Figure 7A, the equivalent amino acid, E475, is positioned on the loop that joins β-strand "l" and α-helix "L," but is predicted to be 30 Å from the catalytic nucleophile, D273. There are two other acidic residues on this loop of domain 2 in the *Aspergillus* β-xylosidase (D478, E480). The closest to the catalytic nucleophile is D478, but this is still 15 Å away from the nucleophile (Fig. 7B). A more likely catalytic acid in the *Aspergillus* enzyme is E196, which is located on domain 1 approximately 6.5 Å from the catalytic nucleophile D273. The E196 residue is similar in both sequence and 3D position to E220 of the barley enzyme; this residue is highly conserved in the plant enzymes (Figs. 4 and 7). As shown in Figure 7B, the distance between D285 and E220 in the barley enzyme is 6.4 Å.

Fig. 6. (Overleaf.) Ribbon representation of barley β-glucan exohydrolase isoenzyme ExoI (left), and domain 1 and domain 2 of the *Butyrivibrio fibrisolvens* β-glucosidase (right). Domain 1, domain 2, and the linker region of each enzyme are coloured in magenta, cyan, and yellow, respectively, and the domain arrangement of the two enzymes are compared below the structures, by using the same colour code. The $(\alpha/\beta)_8$ barrel and $(\alpha/\beta)_6$ sandwich of the *Butyrivibrio* enzyme represent partial folds. The catalytic nucleophiles (D285 in the barley enzyme and D769 in the *Butyrivibrio* enzyme)[13,30] and catalytic acid (E491 in the barley enzyme) are shown in red and blue, respectively. The dotted yellow line represents the linker between the NH_2-terminus of the $(\alpha/\beta)_8$ barrel and COOH-terminus of the $(\alpha/\beta)_6$ sandwich of the composite *Butyrivibrio* model. The figure was generated with MOLSCRIPT.[54]

Fig. 7. (Overleaf.) Comparisons of the three-dimensional (3D) structures and the catalytic regions of barley β-glucan exohydrolase isoenzyme ExoI and the *Aspergillus nidulans* β-xylosidase. **A:** Superposition of the C^α backbones of the barley enzyme (grey) and the *Aspergillus* model (cyan). The root mean square deviation between the two enzymes is 1.02 Å over 527 amino acid residues. Glucose trapped in the active site of the barley enzyme is shown in CPK colours. The figure was generated with GRASP.[56] **B:** Overall, the coils and labels showing the particular amino acid residues are shown in cyan for the *Aspergillus* enzyme[22] and grey for the barley enzyme. The side chains of catalytic nucleophiles (D285 in barley and D273 in *Aspergillus*) are on domain 1; they are located in similar positions in the 3D structures of the two enzymes. The side chain of the putative catalytic acid of the barley enzyme, E491, is on domain 2. The side chain of the putative catalytic acid of the barley enzyme, E491, is 6.4 Å from the D285 nucleophile. In the *Aspergillus* enzyme, the amino acid residue that corresponds to the catalytic acid on domain 2 of the barley enzyme is D478. This residue is approximately 15 Å from the catalytic nucleophile D273; this distance is larger than usual for retaining glycoside hydrolases, and as a result, D478 might not be the actual catalytic acid. A more likely catalytic acid residue in the *Aspergillus* enzyme is E196 from domain 1. The corresponding amino acid residue on domain 1 of the barley enzyme is E220. The distance between D273 and E196 is around 6.5 Å. Glucose (Glc) trapped in the active site of the barley enzyme is shown in CPK colours. The figure was generated with MOLSCRIPT.[54]

DISCUSSION

The 3D structure of a broad specificity β-glucan exohydrolase from barley[13] has been used to examine the structures of family 3 glycoside hydrolases, which include enzymes from higher plants, fungi, and bacteria (Fig. 1). Although amino acid sequence similarities between family members are often less than 30% (Table I), it is possible through molecular modeling to show that most of the 99 family members consist of an NH_2-terminal domain that adopts an $(\alpha/\beta)_8$ barrel fold (Figs. 3, 5, and 6). The NH_2-terminal domain is attached to the second domain by means of a linker region which is 14-41 amino acids in length. Domain 2 adopts an $(\alpha/\beta)_6$ sandwich conformation (Figs. 3, 5, and 6), and is followed by a long antiparallel loop at the COOH-terminus of the enzymes, which is also conserved across most of the family (Table II). In two bacterial enzymes, from *Butyrivibrio fibrisolvens* and *Ruminococcus albus*, in which the order of the two domains is reversed, it is still possible to build models that are similar in fold to the barley β-glucan exohydrolase (Fig. 6). Heterologous expression of the *Butyrivibrio* gene (*bglA*) enzyme in *E. coli* confirmed that the domains were indeed reversed and that the expressed enzyme was catalytically active.[30] The reversal of domain sequence suggests that recombination or other shuffling machinery may have caused "domain" rearrangements in the genes that encode the two enzymes.[40] It is noteworthy that β-glucosidases from *Ruminococcus albus* can have either the "normal" or "reversed" arrangement of domains 1 and 2 (Fig. 1).

The modeling programs also allowed the relative dispositions of the catalytic nucleophiles and catalytic acids of the enzymes to be examined. The crystal structure of barley β-glucan exohydrolase isoenzyme ExoI indicates that the catalytic nucleophile, D285, and the catalytic acid, E491, are situated approximately 6.5 Å apart, and each are 3.0 to 4.5 Å from the anomeric C-1 atom of the bound glucose.[13] These distances are typical of those found in other retaining glycoside hydrolases,[41,42] and are similar in all the higher plant enzymes in family 3. Varghese and colleagues[13] noted that the catalytic nucleophile and the catalytic acid of the barley enzyme are located on different domains and suggested that the linker region between the two domains could act as a molecular hinge that would enable the second domain, which carries the catalytic acid, to move away from, or closer to, the active site on domain 1; this could regulate enzyme activity.

Although the catalytic nucleophile is absolutely conserved across the family 3 glycoside hydrolases, it has become apparent here that the catalytic acid is not highly conserved, especially in the distantly related members of the family (Table II). Nevertheless, there seem to be other acidic amino acids in the active site region, such as E196 on domain 1 of the *Aspergillus* β-xylosidase (Fig. 7B), and these might be recruited to act as catalytic acids in some members of the family. Furthermore, it is possible that the distantly related *N*-acetyl β-glucosaminidases in family 3

TABLE II. Comparison of Conserved Sequence Motifs of Family 3 Glycoside Hydrolases

Entry[a]	205–209[b]	280–286[c] ▼	429–435[d]	488–492[e] ◆	589–598[f]
Barley Exo1	AKHFV	GFVISDW	GWTIEWQ	PYTET	PLFRLGYGLT
Barley Exo2	AKHYV	GFVISDW	GWTITWQ	PYAET	PLFPFGFGLT
Maize	AKHFV	GFTISDW	GWTIEWQ	PYTET	PLFRLGYGLT
Nasturtium	AKHFV	GFVISDW	GWTIEWQ	PYAEM	PLFPFGFGLT
Tobacco	AKHFV	GFVISDW	GWTIEWQ	PYAEM	PLFPFGFGIT
Acetobacter	LKHYA	GFVMSDW			PLYPFGYGLT
Alteromonas	LKHFP	GVTVTDA			
Aspergillus	AKHYA	GYVSGDC			PVYEFGHGLF
Butyrivibrio $(\alpha/\beta)_8$	PKHFA	GFVVTDY			
Butyrivibrio $(\alpha/\beta)_6$					YPFGYGLS
Coccidioides	AKHLV	GFVMTDW			PRYHFGYGLS
Dictyostelium	AKHYF	GVAVTDW	GWSVHWQ	PEAET	PLFQFGDGLS
Pseudomonas	AKHFI	GLVVGDW	GWSVSWQ	PYAEM	PLFPYGYGLS
Thermoanaerobacter	GKHFV	GIVVSDY			PLYPFGYGLS
Unidentified bacterium	AKHYI	GFVVTDW			

[a]Representatives were chosen from different branches on the unrooted phylogenetic radial tree. Identical residues are enclosed in boxes.
[b]Conserved sequence containing a putative carbohydrate-binding site.[13]
[c]Conserved sequence containing the catalytic nucleophile[13] (▼).
[d]Conserved sequence joining helices J1 and J2 of $(\alpha/\beta)_6$ sandwich (portion of a Rossmann-like fold).[13]
[e]Conserved sequence containing the catalytic acid[13] (◆).
[f]Conserved COOH-terminal antiparallel loop sequence.[13]

require only one acidic catalytic residue in the active site. This is found in bacterial *N*-acetyl β-glucosaminidases belonging to family 20, where the *N*-acetoamido group on C-2 of the *N*-acetyl β-D-glucosaminyl residue of the substrate itself acts as the catalytic nucleophile.[43,44]

The active site of the barley β-glucan exohydrolase consists of a shallow pocket that has the appearance of a coin slot and is located near the interface of domains 1 and 2.[13] The pocket is approximately 13 Å deep and would accommodate only two glycosyl residues of bound substrate; the remainder of the substrate would project away from the surface of the enzyme. The active site geometry can be reconciled with the broad substrate specificity of the barley β-glucan exohydrolases,[3] because the shallow active site pocket would allow the enzyme to bind to the nonreducing termini of a wide range of substrates and the overall shape of the substrate would not be a key determinant for the formation of the enzyme-substrate complex. Accordingly, substrates with a range of linkage types might be accommodated in the active site and this, in turn, would explain the broad substrate specificity of the enzyme.[3]

On this basis, one might predict that many of the other family 3 glycoside hydrolases will also have broad substrate specificities. Relatively few of the enzymes have been purified and characterized in detail. Many are simply deduced from open reading frames in cloned genes or cDNAs. Those that have been partially purified have been tested against a narrow selection of potential substrates, and many are classified as β-glucosidases because they can hydrolyze synthetic aryl β-glucosides such as 4-nitrophenyl β-D-glucoside. It should be noted that β-glucosidases which hydrolyze this substrate also fall into the family 1 group of glycoside hydrolases.[7] Family 1 "β-glucosidases" do have an active site pocket on their surfaces but, in contrast to the shallow pocket found in family 3 enzymes, the family 1 enzymes have a much deeper pocket that is more like a dead-end tunnel and can accommodate approximately six glucosyl residues of the substrate.[31,45] The deep, narrow tunnel of the family 1 enzymes would undoubtedly place much greater constraints on the conformation of substrates that could fit into the active site; relatively straight (1→4)-β-linked oligoglucoside substrates are required.[31] Thus, although family 1 and family 3 glycoside hydrolases will catalyze the hydrolysis of 4-nitrophenyl β-D-glucoside, in most cases their actual substrate specificities have not been rigorously evaluated. There are likely to be fundamental differences in both the breadth of their specificities and in the length of substrates that are hydrolyzed.

These considerations of substrate specificity raise several issues regarding functions of family 3 glycoside hydrolases, which have been identified in a wide range of plants and microorganisms but not in animals. This finding suggests that they might have a fundamental function, which could be linked to plant cell wall degradation. Family 3 glycoside hydrolases have two domains; the active site is located on domain 1 but catalytic amino acids might also be located on domain 2 (Figs. 3 and 7). Could there be another function for the second domain? Hrmova and Fincher[3] presented kinetic evidence for positive cooperativity of binding for barley (1→3,1→4)-β-glucan by the barley β-glucan exohydrolases, and Varghese et al.[13] suggested that this finding might be explained by a second binding site for the (1→3,1→4)-β-glucan substrate on domain 2. A Rossmann-like fold that lies at the ends of β strands "j" and "l" and on the loop between helices J1 and J2 of the second domain is conserved in the plant enzymes

and in some bacterial enzymes (Table II). This fold is similar to conformations found in carbohydrate-binding regions of lectins and other proteins.[46]

However, (1→3,1→4)-β-glucans are not found in dicotyledonous plants, although family 3 glycoside hydrolases with two domains have been detected in the dicotyledons tobacco (N. Koizumi, unpublished results, GenBank accession number AB017502) and nasturtium.[47] Perhaps the second domains of these enzymes bind xyloglucans, which are polysaccharides analogous to the (1→3,1→4)-β-glucans in location, and possibly function, in cell walls of dicotyledons.

Observations of this kind provide circumstantial evidence for a role of domain 2 in attaching the enzyme to insoluble substrates or to cell walls, in a manner similar to that observed for the cellulose-binding domains of cellobiohydrolases,[48,49] chitinases,[50] glucoamylases,[51] and related polysaccharide hydrolases. The plant enzymes might function in wall loosening during elongation of young vegetative tissues,[52,53] whereas the microbial family 3 glycoside hydrolases could be participants in the enzymic degradation of cell walls during pathogenesis or during the normal autotrophic utilization of plant residues. In any case, detailed analyses of substrate specificities and the precise definition of kinetic properties of other family 3 glycoside hydrolases, in particular, the more distantly related enzymes from microbial sources, will be necessary before we can confidently assign functions to this widely distributed group of enzymes.

ACKNOWLEDGMENTS

We thank Professor Peter Colman for his ongoing support of the work and Dr. Richard Stewart for his assistance. G.B.F. received financial support for this work.

REFERENCES

1. Webb EC. Enzyme nomenclature. San Diego: Academic Press; 1994. 862 p.
2. Hrmova M, Harvey AJ, Wang J, et al. Barley β-D-glucan exohydrolases with β-D-glucosidase activity: purification, characterization, and determination of primary structure from a cDNA clone. J Biol Chem 1996;271:5277–5286.
3. Hrmova M, Fincher GB. Barley β-D-Glucan exohydrolases: substrate specificity and kinetic properties. Carbohydr Res 1998;305: 209–221.
4. Henrissat B. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 1991;280:309–316.
5. Gaboriaud C, Bissery V, Benchetrit T, Mornon JP. Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. FEBS Lett 1987;224:149–155.
6. Callebaut I, Labesse G, Durand P, et al. Deciphering protein sequence information through hydrophobic cluster analysis (hca): current status and perspectives (review). Cell Mol Life Sci 1997;53: 621–645.
7. Henrissat B, Davies G. Structural and sequence-based classification of glycoside hydrolases. Curr Opin Struct Biol 1997;7:637–644.
8. Henrissat B. Glycosidase families. Biochem Soc Trans 1998;26:153–156.
9. Varghese JN, Garrett TPJ, Colman PM, Chen L, Hoj PB, Fincher GB. Three-dimensional structures of two plant β-glucan endohydrolases with distinct substrate specificities. Proc Natl Acad Sci USA 1994;91:2785–2789.
10. Høj PB, Fincher GB. Molecular evolution of plant β-glucan endohydrolases (review). Plant J 1995;7:367–379.
11. Sanchez R, Sali A. Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome. Proc Natl Acad Sci USA 1998; 95:13597–13602.
12. Hrmova M, Varghese JN, Høj PB, Fincher GB. Crystallization and preliminary X-ray analysis of β-glucan exohydolase isoenzyme ExoI from barley (*Hordeum vulgare*). Acta Crystallogr D 1998;54: 687–689.
13. Varghese JN, Hrmova M, Fincher GB. Three-dimensional structure of a barley β-D-glucan exohydrolase; a family 3 glycosyl hydrolase. Structure 1999;7:179–190.
14. Coutinho PM, Henrissat B. The modular structure of cellulases and other carbohydrate-active enzymes: an integrated database approach. In: Ohmiya K, Hayashi K, Sakka K, Kobayashi Y, Karita S, Kimura T, editors. Genetics, biochemistry and ecology of cellulose degradation. Tokyo: Uni Publishers Co; 1999. p 15–23.
15. Devereux J, Haeberli P, Smithies O. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 1984;12:387–395.
16. Corpet F, Gouzy J, Kahn D. The ProDom database of protein domain families. Nucleic Acids Res 1998;26:323–326.
17. Sonnhammer ELL, Kahn D. Modular arrangement of proteins as inferred from analysis of homology. Protein Sci 1994;3:482–492.
18. Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
19. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
20. Smith TF, Waterman MS. Identification of common molecular sequences. J Mol Biol 1981;147:195–197.
21. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. Electrophoresis 1997;18:2714–2723.
22. Perez-Gonzales JA, van Peij NNME, Bezoen A, MacCabe AP, Ramon D, de Graaff LH. Molecular cloning and transcriptional regulation of the *Aspergillus nidulans xlnd* gene encoding a beta-xylosidase. Appl Environ Microbiol 1998;64:1412–1419.
23. Thompson JD, Higgins DG, Gibson TJ. Clustal W. Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acid Res 1994;22:4673–4680.
24. Brünger AT. E-XPLOR Version 3.1: A system for X-ray crystallography and NMR. New Haven: Yale University Press; 1993. 382 p.
25. Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A 1991;47:110–119.
26. Verlet L. Computer experiments on classical fluids: I. Thermodynamical properties of Lennard-Jones molecules. Phys Rev 1967; 159:98–105.
27. Powell MJD. Restart procedures for the conjugate gradient method. Math Prog 1977;12:241–254.
28. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 1993;26:283–291.
29. Engh RA, Huber R. Accurate bond and angle parameters for X-ray protein structure refinement. Acta Crystallogr A 1991;47:392–400.
30. Lin LL, Rumbak E, Zappe H, Thompson JA, Woods DR. Cloning, sequencing and analysis of expression of a *Butyrivibrio fibrisolvens* gene encoding a β-glucosidase. J Gen Microbiol 1990;136: 1567–1576.
31. Hrmova M, MacGregor EA, Biely P, Stewart RJ, Fincher GB. Substrate binding and catalytic mechanism of a barley β-D-glucosidase/(1→4)-β-D-glucan exohydrolase. J Biol Chem 1998;273: 11134–11143.
32. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. J Mol Biol 1963;7: 95–99.
33. Morris AL, McArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. Proteins 1992; 12:345–364.
34. Kleywegt GJ, Jones TA. Phi/Psi-chology: Ramachandran revisited. Structure 1996;4:1395–1400.
35. Bush J, Richardson J, Cardelli JJ. Molecular cloning and characterization of the full-length cDNA encoding the developmentally regulated lysosomal enzyme beta-glucosidase in *Dictyostelium discoideum*. J Biol Chem 1994;269:1468–1476.
36. Breves R, Bronnenmeier K, Wild N, Lottspeich F, Staudenbauer

WL, Hofemeister J. Genes encoding two different beta-glucosidases of *Thermoanaerobacter brockii* are clustered in a common operon. J Appl Environ Microbiol 1997;63:3902–3910.
37. Healy FG, Ray RM, Aldrich HC, Wilkie AC, Ingram LO, Shanmugam KT. Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. Appl Microbiol Biotechnol 1995;43:667–674.
38. Rixon JE, Ferreira LMA, Durrant AJ, Laurie JI, Hazlewood GP, Gilbert HJ. Characterization of the gene *cel*D and its encoded product 1,4-β-glucan glucohydrolase D from *Pseudomonas fluorescens* subsp. *cellulosa*. Biochem J 1992;285:947–955.
39. Tsujibo H, Fujimoto K, Tanno H, et al. Gene sequence, purification and characterization of *N*-acetyl-beta-glucosaminidase from a marine bacterium, *Alteromonas* sp strain O-7. Gene 1994;146:111–115.
40. Stemmer PC. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc Natl Acad Sci USA 1994;91:10747–10751.
41. McCarter JD, Withers SG. Mechanism of enzymatic glycoside hydrolysis. Curr Opin Struct Biol 1994;4:885–892.
42. Davies G, Henrissat B. Structure and mechanism of glycoside hydrolases. Structure 1995;3:853–859.
43. Tews I, Perrakis A, Oppenheim A, Dauter Z, Wilson KS, Vorgias CE. Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease. Nature Struc Biol 1996;3:638–648.
44. Drouillard S, Armand S, Davies GJ, Vorgias CE, Henrissat B. *Serratia marcescens* chitobiase is a retaining glycosidase utilizing substrate acetamido group participation. Biochem J 1997;328:945–949.
45. Barrett T, Suresh SG, Tolley SP, Dodson EJ, Hughes MA. The crystal structure of cyanogenic, β-glucosidase from white clover, a family 1 glycosyl hydrolase. Structure 1995;3:951–960.
46. Brändén CI. Relation between structure and function of α/β proteins. Rev Biophys 1980;13:317–338.
47. Crombie HJ, Chengappa S, Hellyer A, Reid JSG. A xyloglucan oligosaccharide-active, transglycosylating β-D-glucosidase from the cotyledons of nasturtium (*Tropaeolum majus* L) seedlings - purification, properties and characterization of a cDNA clone. Plant J 1998;15:27–38.
48. Gilkes NR, Henrissat B, Kilburn DG, Miller JRC, Warren RAJ. Domains in microbial β-1,4-glycanases: sequence conservation, function, and enzyme families. Microbiol Rev 1991;55:303–315.
49. Teeri TT, Koivula A, Linder M, Wohlfahrt G, Divne C, Jones TA. *Trichoderma reesei* cellobiohydrolases: why so efficient on crystalline cellulose. Biochem Soc Trans 1998;26:173–178.
50. Watanabe T, Ito Y, Yamada T, Hashimoto M, Sekine S, Tanaka H. The roles of the C-terminal domain and type III domains of chitinase A1 from *Bacillus circulans* WL-12 in chitin degradation. J Bacteriol 1994;176:4465–4472.
51. Sigurskjøld BW, Christensen T, Payre N, Cottaz S, Driguez H, Svensson B. Thermodynamics of binding of heterobidentate ligands consisting of spacer-connected acarbose and β-cyclodextrin to the catalytic and starch-binding domains of glucoamylase from *Aspergillus niger* shows that the catalytic and starch-binding sites are in close proximity in space. Biochemistry 1998;37:10446–10452.
52. Sakurai N, Masuda Y. Auxin-induced changes in barley coleoptile cell wall composition. Plant Cell Physiol 1978;19:1217–1223.
53. Cosgrove DJ. Enzymes and other agents that enhance cell wall extensibility. Annu Rev Plant Physiol Plant Mol Biol 1999;50:391–417.
54. Kraulis P. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J Appl Cryst 1991;24:946–950.
55. Barton GJ. ALSCRIPT: a tool to format multiple sequence alignments. Protein Eng 1993;6:37–40.
56. Nicolls A, Sharp K, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 1991;4:281–296.