Molecular modeling of family GH16 glycoside hydrolases: Potential roles for xyloglucan transglucosylases/hydrolases in cell wall modification in the poaceae

MARCO STROHMEIER,¹ MARIA HRMOVA,² MARKUS FISCHER,¹
ANDREW J. HARVEY,² GEOFFREY B. FINCHER,² AND JÜRGEN PLEISS¹

¹Institute of Technical Biochemistry at the University of Stuttgart, D-70569 Stuttgart, Germany

²Faculty of Sciences, School of Agriculture and Wine, and the Australian Centre for Plant Functional Genomics, The University of Adelaide, Glen Osmond, South Australia 5064, Australia

(RECEIVED April 21, 2004; FINAL REVISION August 30, 2004; ACCEPTED August 30, 2004)

Abstract

Family GH16 glycoside hydrolases can be assigned to five subgroups according to their substrate specificities, including xyloglucan transglucosylases/hydrolases (XTHs), (1,3)- β -galactanases, (1,4)- β -galactanases/ κ -carrageenases, “nonspecific” (1,3/1,3;1,4)- β -D-glucan endohydrolases, and (1,3;1,4)- β -D-glucan endohydrolases. A structured family GH16 glycoside hydrolase database has been constructed (<http://www.ghdb.uni-stuttgart.de>) and provides multiple sequence alignments with functionally annotated amino acid residues and phylogenetic trees. The database has been used for homology modeling of seven glycoside hydrolases from the GH16 family with various substrate specificities, based on structural coordinates for (1,3;1,4)- β -D-glucan endohydrolases and a κ -carrageenase. In combination with multiple sequence alignments, the models predict the three-dimensional (3D) dispositions of amino acid residues in the substrate-binding and catalytic sites of XTHs and (1,3/1,3;1,4)- β -D-glucan endohydrolases; there is no structural information available in the databases for the latter group of enzymes. Models of the XTHs, compared with the recently determined structure of a *Populus tremulos* $\times$ *tremuloides* XTH, reveal similarities with the active sites of family GH11 (1,4)- β -D-xylan endohydrolases. From a biological viewpoint, the classification, molecular modeling and a new 3D structure of the *P. tremulos* $\times$ *tremuloides* XTH establish structural and evolutionary connections between XTHs, (1,3;1,4)- β -D-glucan endohydrolases and xylan endohydrolases. These findings raise the possibility that XTHs from higher plants could be active not only on cell wall xyloglucans, but also on (1,3;1,4)- β -D-glucans and arabinoxylans, which are major components of walls in grasses. A role for XTHs in (1,3;1,4)- β -D-glucan and arabinoxylan modification would be consistent with the apparent overrepresentation of XTH sequences in cereal expressed sequence tags databases.

Keywords: database; β -galactanases, (1,3/1,3;1,4)- β -D-glucan endohydrolases; (1,3;1,4)- β -D-glucan endohydrolases; homology modeling; xyloglucan transglucosylases

Progressive changes in the composition of cell walls and in the fine structure of constituent polysaccharides are of central importance in plant development. Different cell types

within a plant can be distinguished from each other by the chemistry and organization of their walls, and walls differ in a way that is related to developmental stage or to their exposure to different environmental conditions (Fincher 1993; Pennel 1998; Cosgrove 1999; Carpita et al. 2001). Wall composition and fine structure can also change in response to microbial attack, when rapid deposition of the polysaccharide callose and layers of lignin, together with the formation of a cross-linked protein network, create a

Reprint requests to: Jürgen Pleiss, Institute of Technical Biochemistry at the University of Stuttgart, Allmandring 31, D-70569 Stuttgart, Germany; e-mail: Juergen.Pleiss@itb.uni-stuttgart.de; fax: +49-711-685-3195.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.04828404>.

physical barrier that is believed to impede the progress of invading fungal and bacterial pathogens (Mengden et al. 1996; Knogge 2002). Thus, the wall is a dynamic structure in which newly deposited polysaccharides can be modified or restructured during normal growth and development, or in response to abiotic and biotic stresses.

Primary walls of dicotyledonous plants consist of a cellulosic network embedded in a matrix of complex polysaccharides of which xyloglucans, glucuronoarabinoxylans, and pectic polysaccharides are most abundant (Carpita and Gibeaut 1993). Extensive intermolecular hydrogen bonding between matrix-phase polysaccharides and cellulose are thought to be important determinants of overall wall integrity (Fry 1989; Hayashi 1989; Carpita et al. 2001). Walls of the monocotyledons are constructed in essentially the same way, although glucuronoarabinoxylans and (1,3;1,4)- β -D-glucans predominate in the matrix phase, particularly in the grasses; levels of pectic polysaccharides and xyloglucans are relatively low in these species (Carpita and Gibeaut 1993). For example, xyloglucan levels in walls from barley (*Hordeum vulgare*) aleurone, starchy endosperm, and young leaves are extremely low or absent (Sakurai and Masuda 1978; Fincher 1993). The highest xyloglucan content so far detected in barley cell walls is 8% to 10% in coleoptiles (D.M. Gibeaut, M. Pauly, and G.B. Fincher, unpubl.). Xyloglucan levels in suspension-cultured maize cells are ~4% (Rose et al. 2002), and low levels are found in walls from the starchy endosperm of rice (Shibuya and Misaki 1978; Shibuya et al. 1983).

Xyloglucans are subject to modification following their initial deposition into the wall. In particular, their molecular mass distribution can be altered to simultaneously generate larger and smaller polysaccharide chains (Smith and Fry 1991; Nishitani 1997). This "disproportionation" is catalyzed by xyloglucan transglycosylases (EC 2.4.1.207) that are abundant in the apoplastic space during wall development. It is now known that xyloglucan-modifying enzymes include both xyloglucan endotransglycosylases (XETs) and xyloglucan endohydrolases (XEHs), which are collectively referred to as xyloglucan transglycosylases/hydrolases (XTHs) (Farkas et al. 1992; Fanutti et al. 1993; Rose et al. 2002).

Our interest in barley XTHs was originally stimulated by observations that XTH sequences were surprisingly abundant in barley expressed sequence tag (EST) databases. These observations were consistent with earlier reports that XET activity extractable from vegetative tissues of grasses was higher than that extractable from dicotyledons (Fry et al. 1992; Wu et al. 1994). Given the low levels of xyloglucans in walls of most barley tissues, it seemed possible that the enzymes might be active on the more abundant matrix-phase polysaccharides, namely, the arabinoxylans and/or the (1,3;1,4)- β -D-glucans. In support of this possibility, it was recently reported that the molecular mass of xylans in

walls of suspension-cultured maize cells increased dramatically in the first few hours after deposition in the walls (Kerr and Fry 2003); this was suggestive of (1,4)- β -D-xylan transglycosylation activity.

Here, homology modeling has been used to develop potential three-dimensional (3D) structures for XTHs. The models have subsequently been compared with known 3D structures of enzymes that hydrolyze arabinoxylans and (1,3;1,4)- β -D-glucans, to assess the possibility that XTHs from grasses could bind to and disproportionate these abundant wall constituents. The plant XTHs are all classified in the family GH16 group of glycoside hydrolases, although a small number of microbial XTHs are classified within family GH12 (Coutinho and Henrissat 1999) (<http://afmb.cnrs-mrs.fr/CAZY/>). Initial analyses of the family GH16 enzymes have revealed an evolutionary and structural link between higher plant XTHs and microbial (1,3;1,4)- β -D-glucan endohydrolases, while more detailed comparisons of active site regions have uncovered similarities between the XTHs and family GH11 (1,4)- β -D-xylan endohydrolases. Our results are supported by the recently published 3D structure of the *Populus tremula* $\times$ *tremuloides* XTH (Protein Data Bank [PDB] entries 1UN1 and 1UMZ; Johansson et al. 2004).

Results

EST database analysis reveals an abundance of XTH sequences

Analysis of cell wall-modifying enzymes and proteins from 345,000 barley ESTs from the National Center for Biotechnology Information (NCBI) Web site revealed an abundance of XTH sequences in the databases. Although xyloglucans account for <5% to 10% of walls in most barley tissues, the XTH ESTs represent >20% of all ESTs encoding enzymes and proteins that have been implicated in the synthesis, restructuring, and/or degradation of major wall polysaccharides in barley (Fig. 1A). Within the XTH ESTs, five genes account for most of the entries (Fig. 1B). These highly represented genes are designated *HvXTH1* to *HvXTH5* (Fig. 1B). Further analyses of the XTH sequences identified in Figure 1 showed that the barley genome carries at least 22 independent XTH genes (data not shown). The detection of additional XTH ESTs that did not overlap with regions used to define the 22 different genes would suggest that further XTH genes are likely to be present on the barley genome. Examination of the draft rice genome sequence (Goff 2002) indicated that the rice genome carries at least 29 XTH genes (data not shown). Cellulose synthases (*CesAs*) comprise ~25% of the barley ESTs, while expansins, which are involved in wall elongation (Cosgrove 1999), comprise ~15% of wall-modifying proteins in the barley EST databases (Fig. 1A).

Figure 1. Analysis of EST databases. (A) Relative abundance of ESTs encoding cell wall-modifying enzymes and proteins taken from ~345,000 barley ESTs. Numbers indicate the actual number of EST entries for each class of enzyme or protein. CesA indicates cellulose synthase, Gsl indicates glucan synthase-like proteins, Ara/Xyl indicates α -L-arabinofuranosidases and β -D-xylosidases, and AXAH indicates arabinoxylan arabinofuranohydrolases. (B) Relative abundance of ESTs encoding XTHs in the barley EST databases. The five most abundantly represented genes are designated *HvXTH1* (equivalent to *PM2*; Smith et al. 1997), *HvXTH2* (equivalent to *PM5*; Schünmann et al. 1997), *HvXTH3* (equivalent to *XEA*; Schünmann et al. 1997), *HvXTH4* (equivalent to *XEB*; Schünmann et al. 1997), and *HvXTH5* (equivalent to *EXT*; Smith et al. 1997). The unnamed relatively low abundance ESTs are likely to arise from nine separate genes.

The family GH16 glycosyl hydrolases can be divided into five subgroups

The Lipase Engineering Database (LED) system (Fischer and Pleiss 2003) was used to collect ~260 proteins from the family GH16 group of glycoside hydrolases, as in the CAZy Database (Coutinho and Henrissat 1999) (server at http://afmb.cnrs_mrs.fr/~cazy/CAZY/index.html). Based on multiple sequence alignment, phylogenetic analysis, and substrate specificity, the proteins were assigned to five subgroups, each of which has distinct substrate specificity. The subgroups included XTHs (110 proteins), (1,3;1,4)- β -D-glu-

canases (49 proteins), (1,3)- β -galactanases (10 proteins), (1,4)- β -galactanases/ κ -carrageenases (5 proteins), and non-specific (1,3/1,3;1,4)- β -D-glucanases (74 proteins). The sequences of two SKN1 proteins, two KRE6 proteins, and four putative β -D-glucan synthesis-associated proteins were assigned to the nonspecific (1,3/1,3;1,4)- β -D-glucanase subgroup, as proposed previously (Barbeyron et al. 1998). Several proteins classified in the CAZy Database as family 16 glycoside hydrolases (Coutinho and Henrissat 1999), including four (1,3)- β -D-glucan recognition proteins, four Gram-negative bacteria binding proteins, a p50 protein, an allergen, an external agarase, and eight (1,3)- β -D-glucan-binding proteins, could not be assigned to any of the subgroups.

An unrooted phylogenetic tree showing a random selection of 57 enzymes of the 260 proteins from the five subgroups is shown in Figure 2. Inclusion of all 260 proteins resulted in a cluttered tree in which individual proteins were difficult to discern (Harvey et al. 2000). Enzymes for which 3D structures are available are indicated, together with enzymes used for subsequent modeling experiments (Fig. 2). It appears that the large nonspecific (1,3/1,3;1,4)- β -D-glucanase subgroup is centrally placed in the tree, and that the two smaller subgroups of β -galactanases and the two subgroups of (1,3;1,4)- β -D-glucanases and XTHs lie at the edges of the tree (Fig. 2). Whether this means that the non-specific (1,3/1,3;1,4)- β -D-glucanase subgroup has acted as a common ancestor for the other four subgroups in family GH16 remains to be demonstrated.

Multiple sequence alignments further revealed that the catalytic nucleophile and the catalytic acid/base residues of the β -galactanases and nonspecific (1,3/1,3;1,4)- β -D-glucanase subgroups were separated by four amino acid residues (Fig. 3A), while those in the XTH and (1,3;1,4)- β -D-glucanase subgroups were separated by three amino acid residues (Fig. 3B). In the expectation that the different spacing of catalytic amino acid residues was likely to have an effect on substrate specificity and probably reflected an important evolutionary change in active site geometry (Michel et al. 2001), we have subdivided the family GH16 enzymes into subfamily GH16a, comprising the β -galactanase and nonspecific (1,3/1,3;1,4)- β -D-glucanase subgroups, where the catalytic residues are separated by four amino acid residues, and into subfamily GH16b, comprising the XTHs and the (1,3;1,4)- β -D-glucanases, where the catalytic residues are separated by three amino acid residues (Fig. 3B).

The family GH16 database

The family GH16 glycoside hydrolase database (GHDB) is available at <http://www.ghdb.uni-stuttgart.de>. In addition to the alphabetic listing of family GH16 glycoside hydrolases, the database provides experimentally determined 3D protein structures, phylogenetic trees, and multiple

Zobellia galacinovorans (Barbeyron et al. 1998). The sequence identity between the target and template sequences was 39%, and the models suggested that secondary structural elements of the β -jelly roll fold of the enzymes were conserved, while differences were detected in loop regions (data not shown).

There are no crystal structures available for the nonspecific (1,3/1,3;1,4)- β -D-glucanase group of family GH16 enzymes, but the solved structure of the *P. carragenovora* κ -carrageenase (Michel et al. 2001) was used successfully as a template for molecular modeling of the nonspecific (1,3/1,3;1,4)- β -D-glucanases from *Bacillus circulans* (Ya-

Figure 3. (Continued on next page)

hata et al. 1990) and *Strongylocentrotus purpuratus* (Bachman and McClay 1996). The sequences of the two nonspecific (1,3/1,3;1,4)- β -D-glucanases contained 682 and 499 amino acid residues, respectively, and were about twice as long as the sequence of the available κ -carrageenase template structure. Differences in sequence lengths and domain compositions also made it difficult to align the nonspecific (1,3/1,3;1,4)- β -D-glucanases with each other and with the κ -carrageenase template. However, the PRODOM program (Corpet et al. 1998) used in conjunction with multiple sequence alignment eventually allowed 268 out of 682 amino acid residues of the *B. circulans* nonspecific (1,3/1,3;1,4)- β -D-glucanase and 276 out of 499 amino acid residues of the *S. purpuratus* nonspecific (1,3/1,3;1,4)- β -D-glucanase to be aligned with the κ -carrageenase template sequence for homology modeling of the target protein structures (Fig. 4A,B). Both models again showed the typical family GH16 β -jelly roll fold, and the 3D dispositions of the catalytically active region were conserved in both molecular models.

The substrate-binding clefts of both enzymes stretch across the surface of the protein, and in the case of the *S. purpuratus* nonspecific (1,3/1,3;1,4)- β -D-glucanase model, the cleft appears to be closed at the top (Fig. 4, cf. B and A). Nevertheless, the dispositions of the catalytic amino acid residues and of the aromatic residues that are believed to be involved in substrate binding are similar. In addition, the active site aromatic residue that distinguishes the various subgroups (Trp118 in the κ -carrageenase from *P. carragenovora*, a Phe residue in the (1,3;1,4)- β -D-glucanases, and a Tyr residue in the XTHs) (Fig. 3A) is located in a similar position in all enzymes (Fig. 4A,B). The conservation of these residues in Cartesian space around the catalytic site (Fig. 4) is consistent with their conservation in the amino acid sequence alignments of the enzymes (Fig. 3A).

The protein sequence identities, root mean square (RMS) deviations, overall G-factors, and Ramachandran plot statistics for template and target sequences of selected members from family GH16a are shown in Table 1. Despite the relatively low sequence identity values, the figures indicate

that the models constructed for the nonspecific (1,3/1,3;1,4)- β -D-glucanase from the κ -carrageenase template sequence are acceptable.

Homology modeling of 3D structures from family GH16b

In subfamily GH16b, the 3D structures of two *Bacillus* (1,3;1,4)- β -D-glucanases have been solved by X-ray crystallography (PDB entries 1GBG [Hahn et al. 1995] and 1BYH [Keitel et al. 1993]). Based on sequence identities of ~60% between template and target sequences, it was possible to use the solved 3D structures to model the structure of a (1,3;1,4)- β -D-glucanase from *Bacillus brevis* (Louw et al. 1993) and 219 of 330 amino acid residues of a (1,3;1,4)- β -D-glucanase from *Clostridium thermocellum* (Schimming et al. 1992). The only differences between the two solved structures and the two models were in loop regions of the enzymes; all secondary structural elements, residues required for substrate binding, and the 3D dispositions of catalytic residues were conserved (data not shown).

Amino acid sequence identities of ~30% allowed the 3D structures of the two *Bacillus* (1,3;1,4)- β -D-glucanases to be used as templates to build reliable models of two XTHs, namely, the XTHs from *Vigna angularis* (175 of 292 amino acid residues) (Okazawa et al. 1993), and *Vitis labrusca* (197 of 291 amino acid residues) (Ishimaru and Kobayashi 2002). Because the template sequences of the *Bacillus* (1,3;1,4)- β -D-glucanases lacked ~100 amino acid residues found at the C termini of the XTH target sequences, these regions of the XTH proteins could not be modeled. However, the catalytic amino acid residues of the two modeled XTHs were conserved and similar in spatial disposition to catalytic residues of the (1,3;1,4)- β -D-glucanases, as were substrate-binding aromatic residues and the distinguishing aromatic residues Phe92 from the (1,3;1,4)- β -D-glucanases and Tyr75 from the XTHs (Fig. 4C–E). Again, the conservation of these residues in Cartesian space around the catalytic site is consistent with their conservation in the amino

Figure 3. Multiple sequence alignments of the modeled (target) sequences and the 3D structures used as templates. The enzymes of family GH16 are divided into the two subfamilies GH16a (A) and 16b (B). The numbers above the sequences in A and B indicate amino acid residues of 3D structures of the (1,4)- β -galactanase/ κ -carrageenase from *P. carragenovorans* (Michel et al. 2001) and of the (1,3;1,4)- β -D-glucan endohydrolase from *B. macerans* (Hahn et al. 1995), respectively. Multiple sequence alignments of amino acid residues in the indicated regions show the difference between subfamily GH16a enzymes, comprising the (1,4)- β -galactanases/ κ -carrageenases and the “nonspecific” (1,3/1,3;1,4)- β -D-glucanase subgroups. Subfamily GH16b contains the XTHs and the (1,3;1,4)- β -D-glucanases. In subfamily GH16a four amino acid residues separate the catalytic Glu residues (colored in red), while in subfamily GH16b three amino acid residues separate the catalytic Glu residues. Conserved hydrophobic residues (yellow), residues potentially involved in catalytic amino acid function (blue), and the conserved aromatic platform amino acid residues (green) are also indicated. The positions of these amino acid residues in the 3D structures and molecular models are shown in Figures 4 and 5. Putative Ca^{2+} -binding amino acid residues (aqua blue) and nonaromatic substrate binding residues (orange) are also shown. The PYX motif that is inserted in the XTHs is shaded in gray/yellow in B. The multiple sequence alignments were calculated by using ClustalW (Thompson et al. 1994) with the following sequences: *P. carragenovorans* (PDB entry 1DYP), *Z. galactanivorans* (Swis-Prot: AAF21820), *S. purpuratus* (Swis-Prot: Q26660), *B. circulans* (Swis-Prot: P23903), *B. macerans* (PDB entry 1BYH), *B. lichiformis* (PDB entry 1GBG), *B. brevis* (Swis-Prot: P37073), *C. thermocellum* (P29716), *V. labrusca* (NCBI: BAB78506), and *V. angularis* (Swis-Prot: A49539).

Figure 4. Molecular surface representations of family GH16 and GH11 enzymes. Family GH16a members include the 3D structure of the *P. carragenovora* (1,4)- β -galactanase/ κ -carrageenase (PDB entry 1DYP) (A), and the model of the *S. purpuratus* nonspecific (1,3/1,3;1,4)- β -D-glucanase (Swis-Prot: Q26660) (B). The family GH16b members include the 3D structure of the *B. macerans* (1,3;1,4)- β -D-glucanase (PDB entry 1BYH) (C) and the models of the *V. angularis* XTH (Swis-Prot: A49539) (D) and *V. labrusca* XTH (NCBI: BAB78506) (E). Family GH11 is represented by the *A. niger* (1,4)- β -D-xylan endohydrolase (PDB entry 1UKR) (F). The catalytic amino acid residues are colored in red, amino acid residues affecting the function of catalytic amino acid residues are blue, aromatic amino acid residues are in yellow, and the aromatic platform residues (Tyr, Trp, or Phe) are highlighted in green. The figure was generated by using the PyMol program (<http://www.pymol.org>).

acid sequences of the enzymes (Fig. 3B), and reliability checks showed acceptable values for the models (Table 1). In contrast, models of the XTHs calculated by using the structure of the *Pseudomonas carragenovora* κ -carrageenase as a template showed distorted active site structures (data not shown), which showed that this template was not suitable for modeling XTHs.

The substrate-binding cleft in the *V. angularis* and *V. labrusca* XTH model appears to be somewhat wider than that in the (1,3;1,4)- β -D-glucanase templates (Fig. 4C–F). Multiple sequence alignments of template and target

sequences, together with alignments of all sequences belonging to the XTH subgroup, showed a set of conserved amino acid residues that are located in the substrate binding cleft of the XTHs. The frequencies of these residues are shown in Table 2. Amino acid residues corresponding to Tyr75, His83, Phe106, Tyr170, and Trp174 in the XTH of *V. angularis* are conserved in the substrate-binding cleft of the two modeled XTH structures (Fig. 4D,E; Table 2). The highly conserved His83 and Asp87 residues of the XTHs correspond to Trp103 and Asp107 of the *Bacillus macerans* (1,3;1,4)- β -D-glucanase, respectively. These residues are likely to stabilize the catalytic nucleophile through hydrogen bonds (Keitel et al. 1993) and Asp87 may take part in proton trafficking (Michel et al. 2001). Similarly, Tyr75 of the *V. angularis* XTH corresponds to Phe92 of the *B. macerans* (1,3;1,4)- β -D-glucanase (Table 2), and is 5 to 7 Å from the O ϵ 1 and O ϵ 2 of the catalytic amino acid residues Glu85 and Glu89, while Tyr170 of the *V. angularis* XTH and corresponding amino acid residues in other XTHs are in the same position as Met180 of the *B. macerans* (1,3–1,4)- β -D-glucanase (Table 2). Thus, these Met and Tyr residues are conserved in both the (1,3;1,4)- β -D-glucanase and XTH subgroups of family GH16b enzymes (Fig. 3; Table 2). Amino acid residues Asn121 and Glu131 of the *B. macerans* (1,3–1,4)- β -D-glucanase also involved in substrate binding (Hahn et al. 1995) and are conserved in the XTHs; they correspond to Asn104 and Glu114 of the *V. angularis* XTH. The two aromatic amino acid residues in positions corresponding to Tyr24 and Tyr94 of the *B. macerans* (1,3–1,4)- β -D-glucanase could not be detected in the sequences and the modeled 3D structure of the XTHs (Figs. 3, 4C–E). Because of a low sequence identity between the template and target sequences of *Bacillus licheniformis* and *V. angularis*, respectively (Table 1), the C-terminal portion of 90 to 100 amino acid residues of the *V. angularis* XTH could not be modeled, including the conserved Trp179 (Fig. 3; Table 2). In contrast, the modeled structure of the *V. labrusca* XTH reached further toward the C terminus, including Trp191 (Figs. 3B, 4D,E).

Another difference between the XTHs and the (1,3;1,4)- β -D-glucanases of the family GH16b subgroup is the insertion in the XTHs of three amino acid residues Pro-TyrXxx, where Xxx is a nonconserved amino acid residue, at a position corresponding to Pro₉₈Tyr₉₉Ile₁₀₀ in the *V. angularis* XTH structure (Table 2; Fig. 3). The insertion is located in the modeled structure of *V. angularis* at the beginning of the central β -sheet, close to the Glu89 at substrate-binding subsite 2 or –2, so that the Tyr residue points toward the catalytic amino acid residues and the bound substrate (Fig. 4D). In the molecular model of *V. labrusca* XTH the corresponding Tyr residue is not exposed at the surface but covered by the C-terminal region (Fig. 4D); therefore, it would not be expected to contrib-

Table 1. Evaluation of molecular models generated for selected enzymes from family GH16 glycosyl hydrolases

Entry ^a	Activity ^b	Identity ^c %	RMS ^d Å (aa)	G-factor ^e	Ramachandran values ^f (%)	
					Allowed	Disallowed
Templates ^g						
1. <i>Bacillus macerans</i>	(1,3;1,4)	100	— (214)	−0.75	100	0
2. <i>Bacillus licheniformis</i>	(1,3;1,4)	100	— (214)	0.16	99.5	0.5
3. <i>Pseudoalteromonas carragenovora</i>	(κ-carrag)	100	— (271)	0.26	100	0
Targets ^g						
1. <i>Clostridium thermocellum</i>	(1,3;1,4)	59	0.4 (211)	−0.16	99.4	0.6
1. <i>Bacillus brevis</i>	(1,3;1,4)	58	0.4 (211)	−0.18	100	0
1. <i>Vitis labrusca</i>	(XTH)	28	1.1 (171)	−0.46	97.6	2.4
2. <i>Vigna angularis</i>	(XTH)	30	1.3 (163)	−0.50	98	2
3. <i>Zobellia galactanivorans</i>	(κ-carrag)	40	0.4 (244)	−0.29	98.7	1.3
3. <i>Bacillus circulans</i>	(1,3/1,3;1,4)	32	0.7 (202)	−0.44	98.7	0.3
3. <i>Strongylocentrotus purpuratus</i>	(1,3/1,3;1,4)	22	1.1 (256)	−0.50	99.1	0.9

^a Representatives were chosen from different branches on the unrooted phylogenetic radial tree.

^b Indicates enzymatic activity of the entry.

^c Using Bestfit program of the University of Wisconsin GCG package (Devereux et al. 1984).

^d RMS deviations of the C^α backbone were determined by using the program O (Jones et al. 1991). Numbers in brackets indicate the number of amino acid residues aligned to calculate the RMS deviations.

^e Overall G-factors of models were calculated by PROCHECK (Laskowski et al. 1993).

^f Ramachandran plot statistics excluding Gly and Pro residues.

^g The templates for *Clostridium*, *B. brevis*, and *Vitis* was *B. macerans* enzyme (1BYH) (Keitel et al. 1993); for *Vigna* was *B. licheniformis* enzyme (1GBG) (Hahn et al. 1995); for *Zobellia*, *B. circulans* and *Strongylocentrotus* was *Pseudoalteromonas* enzyme (1DYP) (Michel et al. 2001).

ute to substrate binding. Although the two homologous XTHs are expected to have similar 3D structures, the orientation of the highly conserved Tyr residue of the Pro-TyrXxx motif and the highly conserved Trp residue corresponding to Trp179 of the *V. angularis* model cannot be predicted.

Comparison of XTH models with the experimentally determined 3D structure from *P. tremula* × *tremuloides* XTH

Following submission of this manuscript an experimentally determined 3D structure of *P. tremula* × *tremuloides* XTH

was released (PDB entries 1UN1 and 1UMZ; Johansson et al. 2004). This allowed a comprehensive evaluation of our XTH models. The amino acid sequence identities between the *P. tremula* × *tremuloides* XTH and the *V. angularis* and *V. labrusca* XTHs are 83% and 38%, respectively. The calculated RMS deviation values between C_α of the 3D structure of *P. tremula* × *tremuloides* XTH and the two XTH molecular models from *V. angularis* and *V. labrusca* are 5.8 Å (170 of 175 C_α atoms) and 7.4 Å (177 of 200 C_α atoms), respectively. The major differences resulting in the high RMS deviation values resulted in variations between the outer loop regions and in the N- and C-terminal regions, while the overall β-sandwich fold was very well

Table 2. Frequencies of conservation of selected amino acid residues in family GH16 XTHs and (1,3;1,4)-β-D-glucanases

XTHs							
Residue ^a	Tyr75 ^d	His83	Tyr99 ^d	Phe106	Tyr170 ^d	Trp174 ^d	Trp179
Frequency ^b	96	75 (90)	77 (92)	49 (100)	76 (84)	96	90
(1,3;1,4)-β-D-glucanases							
Residue ^c	Phe92	Trp103	—	Tyr123	Met180	Trp184	Trp192
Frequency ^b	96	75 (90)		49 (100)	79	96	80 (81)

Amino acid residues having the same position are aligned with each other in the table.

^a Amino acid residues in XTHs corresponding to residues of the *V. angularis* and *P. tremula* × *tremuloides* enzymes.

^b Frequencies are expressed in percentages. The numbers in parentheses are the frequencies of all aromatic amino acid residues observed in that position.

^c Amino acid residues in (1,3;1,4)-β-D-glucanases corresponding to residues of the *B. macerans* enzyme.

^d These residues have the same 3D position in the XTHs and in the β-D-xylan endohydrolases, and are designated the “aromatic triad.”

conserved. It is worth noting that the central β -sheet of *P. tremula* $\times$ *tremuloides* XTH deviated from the two models by only 1.4 Å (39 of 175 C $_{\alpha}$ atoms) and 2.5 Å (39 of 200 C $_{\alpha}$ atoms). In addition, most of the side chains forming the substrate binding site of the models have a similar orientation (data not shown). The experimental structure thus allows deciding between the two suggested orientations of the highly conserved ProTyrXxx motif in the two XTH models. According to the structure of the *P. tremula* $\times$ *tremuloides* XTH, the Tyr residue is embedded between the two major β -sheets and points toward the *N*-glycosylated Asn93 residue. Therefore, an interaction between this Tyr residue and the substrate is unlikely, but it might be essential for the conformational stability of this *N*-glycosylation site. In contrast, the highly conserved (Trp179 in *P. tremula* $\times$ *tremuloides* XTH) has a similar structural position as Trp191 in the model of the *V. labrusca* XTH, and this residue forms a part of the substrate binding site.

XTH structures share similarities with (1,4)- β -D-xylan endohydrolase 3D structures

When the two models of the family GH16 *V. angularis* and the *V. labrusca* XTH were compared with the 3D structure of the family GH11 *Aspergillus niger* (1,4)- β -D-xylan endohydrolase (Krengel and Dijkstra 1996), significant differences in the shapes of the substrate-binding clefts were evident (Fig. 4D,F). The XTH cleft seems to be relatively broad, which might be related to the highly substituted

(1,4)- β -D-glucan backbone of the XTH substrate compared with the presence of less bulky substituents on the (1,4)- β -D-xylan substrate of the (1,4)- β -D-xylanases; the top of the (1,4)- β -D-xylanase cleft appears to be closed. As expected for glycosyl endohydrolases, hydrophobic residues are located along the substrate-binding clefts of both types of enzyme (Fig. 4D,F). Nevertheless, both families GH11 and GH16 exhibit β -jelly roll folds (Coutinho and Henrissat 1999). A comparison of the catalytic and the substrate binding residues between the family GH16 *B. macerans* (1,3;1,4)- β -D-glucanase, the family GH16 *P. tremula* $\times$ *tremuloides* XTH, and the family GH11 *A. niger* β -D-xylan endohydrolase demonstrates a close structural similarity, with the catalytic acid/base and catalytic nucleophile residues having the similar 3D orientation (Fig. 5). The β -strands in this region carry aromatic residues, which have similar 3D dispositions in relation to the catalytic residues. The Asp87 residue (or a corresponding Asp107), which is conserved in family GH16 enzymes is located between the two catalytic residues in the (1,3;1,4)- β -D-glucanases and XTHs (Fig. 5A,B), is replaced by Tyr81 in the β -D-xylan endohydrolase (Fig. 5C).

For a more detailed comparison, the catalytic residues and the Asp/Tyr residues that are located between these in the XTH structure and the β -D-xylan endohydrolase structure were superposed. This shows that an “aromatic triad” of three aromatic amino acid residues that are important for substrate binding in the *A. niger* β -D-xylan endohydrolase, namely, Tyr10, Tyr164, and Trp172 (Tahir et al. 2002), has the same relative position in 3D space as a triad of three

Figure 5. Comparison of the 3D dispositions of amino acid residues in the active site regions of enzymes from families GH16 and GH11. Active site amino acid residues from the 3D structure of the family GH16 (1,3;1,4)- β -D-glucanase from *B. macerans* (PDB entry 1BYH) (A) are compared with those from the 3D structure of family GH16 XTH of *P. tremula* $\times$ *tremuloides* XTH (PDB entry 1UN1) (B), and with those from the 3D structure of the family GH11 (1,4)- β -D-xylan endohydrolase from *A. niger* (PDB entry 1UKR) (C). The catalytic amino acid residues are shown in red, and the Asp residue between the two catalytic residues (Fig. 3) is shown in cyan. The distinguishing aromatic amino acid residues are colored in green, and the aromatic residues that might be involved in substrate binding are colored in yellow. In all but the two catalytic residues, the nitrogen and oxygen atoms are colored in blue and in red, respectively. Red circles show the conserved “aromatic triad.” The figure was generated by using the PyMol program (<http://www.pymol.org>).

aromatic amino acid residues, namely, Trp174, Tyr170, and Trp179, in the *P. tremula* × *tremuloides* XTH (Fig. 5, cf. B and C). Aromatic amino acid residues corresponding to Tyr6 and Tyr89 of the *A. niger* β -D-xylan endohydrolase, also identified as important for catalytic activity (Tahir et al. 2002), could not be identified unequivocally through the comparison of the 3D structures. Furthermore, the aromatic platform of Tyr75 in the *P. tremula* × *tremuloides* XTH overlaps with that of Tyr70 in the *A. niger* β -D-xylan endohydrolase (Fig. 5B,C; Table 2). All other subgroups of the family GH16 glycoside hydrolases, namely, the (1,3;1,4)- β -D-glucanases, (1,3)- β -galactanases, (1,4)- β -galactanases/ κ -carrageenases, and “nonspecific” (1,3/1,3;1,4)- β -D-glucanases, do not show this “aromatic triad” pattern in the active site (Figs. 3, 4, 5A).

Discussion

Over 20% of entries that encode cell wall-modifying enzymes or proteins in the large barley EST databases correspond to mRNAs for the XTH group of higher plant enzymes (Fig. 1). Given that barley cell walls usually contain very low amounts of the XTH substrate xyloglucan, which is of relatively low abundance in the Poaceae more generally, this high level of XTH ESTs is surprising. One of several possible explanations for the higher than expected number of XTH entries is that the substrate specificity of the enzymes is broader than originally believed and that they catalyze the hydrolysis or transglycosylation of other polysaccharides that are more abundant than are xyloglucans in barley cell walls. The most abundant noncellulosic polysaccharides of barley walls are arabinoxylans and (1,3;1,4)- β -D-glucans (Fincher 1993; Carpita et al. 2001). We therefore modeled XTH enzymes against the known 3D structures of related family GH16 (1,3;1,4)- β -D-glucan hydrolases and compared the active site regions with those of family GH11 (1,4)- β -D-xylan endohydrolases, in an attempt to identify any structural and hence evolutionary linkages between enzymes that bind xyloglucans, arabinoxylans, and (1,3;1,4)- β -D-glucans.

The higher plant XTHs have been classified by Coutinho and Henrissat (1999) into the family GH16 group of glycosyl hydrolases. This family can be divided into five subgroups on the basis of differences in their substrate specificities (Fig. 2). First, a subgroup of (1,3)- β -galactan endohydrolases (EC 3.2.1.81) can hydrolyze (1,3)- β -galactosyl linkages in complex polysaccharides that include agarose (Michel et al. 2001). The second subgroup of enzymes (EC 3.2.1.83) are specific for (1,4)- β -galactosyl linkages in other complex polysaccharides that contain sulphated and anhydro-galactosyl residues, including κ -carrageenan and keratin sulphate (Kloareg and Quatrano 1988). The third subgroup includes the “nonspecific” (1,3/1,3;1,4)- β -D-glucan endohydrolases (EC 3.2.1.6). These are nonspecific in-

sofar as they can hydrolyze both (1,3)- and (1,4)-linkages in β -D-glucans, provided there is an adjacent (1,3)- β -D-glucosyl residue on the nonreducing terminal side of the linkage hydrolyzed (Anderson and Stone 1975; Høj and Fincher 1995). These enzymes can therefore hydrolyze both (1,3)- β -D-glucans and (1,3;1,4)- β -D-glucans. The fourth subgroup comprises (1,3;1,4)- β -D-glucan endohydrolases (EC 3.2.1.73). These are absolutely specific for the hydrolysis of a (1,4)- β -D-glucosyl linkage, but only if there is an adjacent (1,3)- β -D-glucosyl residue toward the nonreducing end of the substrate. These enzymes can therefore hydrolyze only (1,3;1,4)- β -D-glucans (Parrish et al. 1960; Høj and Fincher 1995; Hrmova and Fincher 2001). The fifth subgroup contains xyloglucan-modifying enzymes that are collectively referred to as xyloglucan transglycosylases/hydrolases (XTHs; EC 2.4.1.207), but include both (XETs and XEHs (Farkas et al. 1992; Fanutti et al. 1993; Rose et al. 2002). These enzymes hydrolyse (1,4)- β -D-glucosyl linkages specifically in xyloglucans, but those with XET activity can also catalyze transglucosylation reactions, in which the nonreducing terminal product of the hydrolysis reaction can subsequently be transferred onto another xyloglucan molecule (Rose et al. 2002).

Although members of the XTH subgroup are found exclusively in higher plants while members of the other four subgroups are of microbial origin (Coutinho and Henrissat 1999), the classification of XTHs with microbial (1,3;1,4)- β -D-glucan endohydrolases in family GH16 establishes an evolutionary link between the two enzyme groups. It is noteworthy that higher plants of the Poaceae also synthesize (1,3;1,4)- β -D-glucan endohydrolases, but these have a completely different protein fold, they are classified in family GH17 and clearly converged with the microbial (1,3;1,4)- β -D-glucan endohydrolases along a distinct evolutionary route (Høj and Fincher 1995; Coutinho and Henrissat 1999). The plant XTHs and microbial (1,3;1,4)- β -D-glucan endohydrolases of subfamily GH16b show a number of features that distinguish them from the other three subgroups of the family. In particular, their catalytic amino acid residues are separated by three amino acid residues, in contrast to members of the other three subgroups that comprise subfamily GH16a, in which four amino acid residues separate the catalytic amino acid residues (Fig. 3). Moreover, the plant XTHs and microbial (1,3;1,4)- β -D-glucan endohydrolases of subfamily GH16b have a distinguishing Phe or Tyr residue in the vicinity of the catalytic amino acid residues, in a position that is occupied by a Trp in subfamily GH16a enzymes (Fig. 4; Michel et al. 2001). One key difference between the XTHs and the (1,3;1,4)- β -D-glucanases is the insertion of three amino acid residues in the XTHs, at a position corresponding to Pro₁₁₈Tyr₁₁₉Ile₁₂₀ of the *V. angularis* XTH (Fig. 3). An additional difference is the substitution of Met, at a position corresponding to Met180 of the *B. macerans* (1,3;1,4)- β -D-glucanase (Fig. 3B), with an

aromatic amino acid residue, mostly Tyr, at a position corresponding to Tyr170 of the *V. angularis* XTH (Figs. 3, 4).

Reliable models of the *V. angularis* and *V. labrusca* XTH could be generated using solved 3D structures for (1,3;1,4)- β -D-glucanases from *B. licheniformis* (Keitel et al. 1993) and *B. macerans* (Fig. 4; Table 1; Hahn et al. 1995). Similarly, reliable models of the family GH16a nonspecific (1,3/1,3;1,4)- β -D-glucanases could be generated by using the known 3D structure of the *P. carragenovora* κ -carrageenase (Michel et al. 2001) as a template (Fig. 4; Table 1), but reliable models could not be built from 3D structures across the subfamilies GH16a and GH16b (Table 1). It must be remembered that generation of the models of XTHs necessitated the removal of 90 to 100 amino acid residues from their C termini, and that this C-terminal region could play an important role in substrate specificity.

The recently published 3D structure of *P. tremula* $\times$ *tremuloides* XTH (Johansson et al. 2004) allowed a detailed evaluation of the modeled XTH enzymes from *V. labrusca* and *V. angularis*. The evaluation demonstrated the reliability of our molecular models and resolves the ambiguity about the role of the conserved Tyr residue in the Pro-TyrXxx motif, which seems to play a structure-stabilizing role rather than a role in substrate binding. However, the concept of an aromatic triad (Tyr10, Tyr164, and Trp172 in β -D-xylan endohydrolase and Trp174, Tyr170, and Trp179 in the *P. tremula* $\times$ *tremuloides* and *V. angularis* XTHs) that is essential for substrate binding has been confirmed. A highly conserved Trp179 of *P. tremula* $\times$ *tremuloides* XTH that is located at the surface of the substrate binding site has been shown to be essential for substrate binding (Johansson et al. 2004). The presence of conserved aromatic amino acids and the general structural similarity of the substrate binding sites of XTHs and (1,3;1,4)- β -D-glucanases suggests similarities in their substrate specificity. However, xyloglucans are much more bulky than are (1,3;1,4)- β -D-glucanases, which is reflected in a slightly more open structure of the XTH substrate binding sites.

In contrast to the apparent similarities between the XTHs and the (1,3;1,4)- β -D-glucanases of family GH16, structural links between the XTHs and (1,4)- β -D-xylan endohydrolases were not so obvious. There are no (1,4)- β -D-xylanases in family GH16; these enzymes are variously placed in families GH8, GH10, and GH11 (Coutinho and Henrissat 1999). The protein folds detected for the glycosyl hydrolase families GH8, GH10, and GH11 that contain (1,4)- β -D-xylan endohydrolases are $(\alpha/\alpha)_6$, $(\beta/\alpha)_8$ and β -jelly roll, respectively (Coutinho and Henrissat 1999). Because the family GH16 enzymes also adopt a β -jelly roll conformation and because the 3D structure of a family GH11 (1,4)- β -D-xylan endohydrolase from *A. niger* has been solved (Krengel and Dijkstra 1996), it was possible to compare this structure with the model of the XTH from *V. angularis* and *V. labrusca* (Fig. 4D–F). In addition, selected active site

amino acid residues from the experimentally determined structure of *P. tremula* $\times$ *tremuloides* and the models prepared for the family GH16 XTH from *V. angularis* and *V. labrusca* were shown to have a similar relative 3D disposition as those defined by the 3D structure of the family GH11 (1,4)- β -D-xylan endohydrolase from *A. niger*. Furthermore, in the *P. tremula* $\times$ *tremuloides* XTH, the structural disposition of the C terminus has been clarified. This finding also supports our predictions that a structural relationship between the XTH enzymes and the GH11 (1,4)- β -D-xylan endohydrolases exists. In the XTHs, the C-terminal region contains an exposed α -helix, which is missing in other GH16 subgroup enzymes. Interestingly GH11 (1,4)- β -D-xylan endohydrolases also have an exposed α -helix that is positioned at the same end of the β -sheet, and this α -helix faces the opposite side to the active site region. The α -helices in both groups of the enzymes carry clusters of aromatic amino acid residues and therefore could potentially be involved in binding to cell walls or other polysaccharide epitopes. These models and the recently experimentally determined 3D structure of the *P. tremula* $\times$ *tremuloides* XTH therefore established a structural and potentially functional linkage between the family GH16 XTHs and the family GH11 (1,4)- β -D-xylan endohydrolases.

In summary, structural links between family GH16 XTHs and (1,3;1,4)- β -D-glucanases, and between the XTHs and family GH11 (1,4)- β -D-xylan endohydrolases have been established. In the context of seeking a potential role for the highly abundant XTHs of the Poaceae (Fry et al. 1992; Wu et al. 1994) in the hydrolysis or modification of polysaccharides other than the assumed substrate of xyloglucan, the homology modeling experiments and the experimentally determined 3D structure of *P. tremula* $\times$ *tremuloides* XTH suggest that the substrate-binding clefts and the active sites of the XTHs are structurally similar to those of the (1,3;1,4)- β -D-glucanases and the (1,4)- β -D-xylan endohydrolases. Thus, one can postulate that the XTHs of the Poaceae might have evolved to bind these other cell wall components. Higher plants already have (1,3;1,4)- β -D-glucanases and (1,4)- β -D-xylan endohydrolases from families other than GH16 and GH11, so one might further suggest that the need for XTHs with altered specificity in the Poaceae could be related to a requirement for the transglycosylation and attendant disproportionation of cell wall arabinoxylans and (1,3;1,4)- β -D-glucans, rather than a role in hydrolysis of these wall polysaccharides. The first evidence in support of this role was recently provided by Kerr and Fry (2003), who showed an abrupt increase in the molecular mass of xylans in walls of suspension-cultured maize cells in the first few hours after deposition in the walls. This is consistent with (1,4)- β -D-xylan transglycosylation activity, which could be catalyzed by an XTH enzyme that has been modified for action on (1,4)- β -D-xylans, and is reminiscent of the well-characterized action of XTHs on xyloglucans (Smith and

Fry 1991; Nishitani 1997). Although it is likely to be difficult and time-consuming to separately purify the five XTH isoenzymes that correspond to the five most abundantly expressed XTH genes from barley (*HvXTH1* to *HvXTH5*), we have now cloned cDNAs for each of these barley genes and will attempt to express active enzymes in heterologous systems, and to investigate whether any of the individual, expressed HvXTH enzymes, isolated in highly purified form, are active on arabinoxylans or (1,3;1,4)- β -D-glucans.

Materials and methods

Database construction and evolutionary analyses

A total of ~260 amino acid sequences of family GH16 glycoside hydrolases was collected through an automated blast search from the NCBI-GenBank database (Benson et al. 1999) by using parsing tools from the LED (Fischer and Pleiss 2003) (<http://www.ghdb.uni-stuttgart.de>). Amino acid residues with potential roles in catalytic and substrate binding were annotated based on the information obtained by multiple sequence alignments with ClustalW (Thompson et al. 1997). Phylogenetic trees were constructed with TREE-PUZZLE version 5.0 using maximum-likelihood and quartet-puzzling (Schmidt et al. 2002). The parallel version of TREE-PUZZLE was run on a Linux PC-cluster of Dual Athlon MP 1800+ with 256 processors and a Myrinet interconnect system. The phylogenetic trees were edited manually and were drawn with the TreeView program (Page 1996). The accession numbers at the GenBank/EMBL, SWISS-PROT, and PDB (Berman et al. 2000) databases of the 57 selected family GH16 glycoside hydrolases are as follows:

The subgroup of “nonspecific” (1,3/1,3;1,4)- β -D-glucan endohydrolases includes the entries from *Eisenia foetida* (O77072), *B. circulans* (Q45095, P23903, and BAC06195), *Thermogota maritima* (B72428), *Thermotoga neapolitana* (Q60039), *Cochliobolus carbonum* (O14421), *Lysobacter enzymogenes* (AAN77505), *Oerskovia xanthineolytica* (O68641), *S. purpuratus* (Q26660), *Streptomyces avermitilis* MA-4680 (BAC69475), and the KRE6 protein from *Candida albicans* KRE6 (P87023).

The subgroup of (1,4)- β -galactanases/ κ -carrageenases includes the entries from *Zobellia galactanivorans* (O84907), *Pirellula* sp.1 (CAD72787.1), *Pseudoalteromonas carrageenovora* (1DYP), and *Pirellula* sp. (CAD73010.1).

The subgroup of (1,3)- β -galactanases includes the entries from uncultured bacterium AguB (AAP49346.1) and AguD (AAP49316.1), *Aeromonas* sp. (AAF03246), *Alteromonas atlantica* (Q59078), *Pseudomonas* sp. Nd137 (BAB79291), *Z. galactanivorans* (AAF21820), *Microscilla* sp. PRE1 (AAK62838), *Streptomyces coelicolor* (P07883), and *S. coelicolor* A3 (2) (NP_627674).

The subgroup of (1,3;1,4)- β -D-glucanases includes entries from *Clostridium acetobutylicum* (D97245), *C. thermocellum* (P29716), *Pseudomonas* sp. (BAC24104), *Orpinomyces* sp. PC-2 (O14412), *Streptococcus bovis* (O07856), *B. brevis* (P37073), *Brevibacillus brevis* (A48378), *Bacillus amyloliquefaciens* (P07980), *Bacillus subtilis* (P04957), *Bacillus lichiformis* (1GBG), *Hypocrea jecorina* (AAF82804), *Bacillus polymyxa* (P45797), *B. macerans* (1BYH), and *Rhizobium meliloti* (P33693).

Finally, the subgroup of XTH entries includes entries from *Carica papaya* (AAK51119), *V. labrusca* (BAB78506), *Arabidopsis thaliana* (AAM91637), *Daucus carota* (AAK30204), *V. angularis* (A49539), *Actinidia deliciosa* (AAC09388), *H.*

vulgare (X91659), *Nicotiana tabacum* (BAA13163), *Gossypium hirsutum* (T09870), *Cicer arietinum* (CAA06217), *Pisum sativum* (BAA34946), *Asparagus officinalis* (AAF80591), *Beta vulgaris* (AAL04440), *Fagus sylvatica* (CAA10231), *Ananas comosus* (AAM28287.1), *Lycopersicon esculentum* (S49812), *Festuca pratensis* (CAC40809), and *P. tremula* $\times$ *tremuloides* (IUN1).

Multiple sequence alignments

Multiple sequence alignments were performed with ClustalW 1.80 (Thompson et al. 1994, 1997; Higgins et al. 1996; Jeanmougin et al. 1998) using the LED database. Individual entries were checked manually by hydrophobic cluster analysis (Lemesle-Varloot et al. 1990) to ensure that integrity of hydrophobic clusters was undisturbed and that both the distribution of secondary structure elements and topology of the active sites remained conserved. The program Bestfit from the University of Wisconsin GCG software package (Devereux et al. 1984), with the implemented gap penalty function and the Smith and Waterman local algorithm (Smith and Waterman 1981), was used to calculate sequence identities and similarities between the template and target protein sequences.

EST database analysis

Publicly available EST sequences from barley were accessed through the NCBI site (<http://www.ncbi.nlm.nih.gov>). A total of ~345,000 barley ESTs available in October 2003 was searched for sequences corresponding to enzymes and proteins involved in cell wall biosynthesis, modification, and/or degradation. These included cellulose synthases (CesA), cellulose synthase-like enzymes (Csl), glucan synthase-like proteins (Gsl), α - and β -expansins, XTHs, β -D-glucan endo- and exohydrolases, (1,4)- β -D-xylan endo- and exohydrolases, (1,4)- β -D-mannan endohydrolases, and α -L-arabinofuranosidases. Stress-related (1,3)- β -D-glucan endohydrolases were specifically excluded from the analysis. No attempts were made to discriminate between isoforms of the various proteins or between individual EST libraries, and total numbers of ESTs were simply used as a first approximation of the relative abundance of mRNA transcripts in the range of tissues from which the EST libraries had been prepared. To estimate the number of genes in the barley XTH gene family, XTH sequences were aligned for the identification of overlapping EST sequences that corresponded to fragments of the same gene, and for the identification of distinct genes. ESTs with homology to previously identified barley and rice XTH genes were aligned and clustered by using the ContigExpress program that is a part of the Vector NTi Suite 7.0 software package (Informax Inc).

Homology protein structure modeling

The 3D molecular models of protein sequences of selected representatives of family GH16 glycoside hydrolases were constructed with Modeller 6v2, which is based on homology modeling by satisfaction of spatial restraints (Sali and Blundell 1993). Default settings of the program Modeller 6v2 (Sali and Blundell 1993) and the structurally optimized aligned protein sequences were used. Homologous proteins exhibiting substrate specificities corresponding to nonspecific (1,3/1,3;1,4)- β -D-glucanases, (1,3;1,4)- β -D-glucanases, (1,3)- and (1,4)- β -galactanases, and XTHs were grouped. The N termini of the amino acid sequences were deduced from the literature data when available and/or in combination with multiple

sequence alignments (Harvey et al. 2000) and the PRODOM program that analyses domain arrangements (Gouzy et al. 1996).

Evaluation of models

The stereochemical quality (Ramachandran plots) and overall G-factors of the final models, which are measures of normality of main-chain bond lengths and bond angles (Engh and Huber 1991), were calculated with PROCHECK (Ramachandran et al. 1963; Laskowski et al. 1993). Further, the program O (Jones et al. 1991) was applied for the calculation of RMS deviations between C $_{\alpha}$ positions of the template and target protein 3D structures. Newly derived 3D models and structures were superposed with Swiss-PDBViewer V3.7 (Guex and Peitsch 1997) and viewed with PyMol (<http://www.pymol.org>).

Acknowledgments

This work was supported by grants from the Federal Ministry of Education and Research, Germany (to J.P.) and from the Australian Research Council and the Grains Research and Development Corporation (to G.B.F.).

References

- Anderson, M.A. and Stone, B.A. 1975. A new substrate for investigating the specificity of β -D-glucan hydrolases. *FEBS Lett.* **52**: 202–207.
- Bachman, E.S. and McClay, D.R. 1996. Molecular cloning of the first metazoan β -1,3 glucanase from eggs of the sea urchin *Strongylocentrotus purpuratus*. *Proc. Natl. Acad. Sci.* **93**: 6808–6813.
- Barbeyron, T., Gerard, A., Potin, P., Henrissat, B., and Kloareg, B. 1998. The κ -carrageenase of the marine bacterium *Cytophaga drobachiensis*: Structural and phylogenetic relationships within family-16 glycoside hydrolases. *Mol. Biol. Evol.* **15**: 528–537.
- Benson, D.A., Boguski, M.S., Lipman, D.J., Ostell, J., Ouellette, B.F., Rapp, B.A., and Wheeler, D.L. 1999. GenBank. *Nucleic Acids Res.* **27**: 12–17.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res.* **28**: 235–242.
- Carpita, N.C. and Gibeault, D.M. 1993. Structural models of primary cell walls in flowering plants: Consistency of molecular structure with the physical properties of the walls during growth. *Plant J.* **3**: 1–30.
- Carpita, N.C., Defernez, M., Findlay, K., Wells, B., Shoue, D.A., Catchpole, G., Wilson, R.H., and McCann, M.C. 2001. Cell wall architecture of the elongating maize coleoptile. *Plant Physiol.* **127**: 551–565.
- Corpet, F., Gouzy, J., and Kahn, D. 1998. The ProDom database of protein domain families. *Nucleic Acids Res.* **26**: 323–326.
- Cosgrove, D.J. 1999. Enzymes and other agents that enhance cell wall extensibility. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**: 391–417.
- Coutinho, P.M. and Henrissat, B. 1999. Carbohydrate-active enzymes: An integrated database approach. In *Recent advances in carbohydrate bioengineering* (eds. H.J. Gilbert et al.), pp. 3–12. The Royal Society of Chemistry, Cambridge, UK.
- Devereux, J., Haerberli, P., and Smithies, O. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**: 387–395.
- Engh, R. and Huber, R. 1991. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. A* **47**: 392–400.
- Fanutti, C., Gidley, M.J., and Reid, J.S.G. 1993. Action of a pure xyloglucan endo-transglycosylase (formerly called xyloglucanase and transglycosylation to xyloglucan subunit oligosaccharides. *Arch. Biochem. Biophys.* **298**: 365–370.
- Farkas, V., Sulova, Z., Stratilova, E., Hanna, R., and MacLachlan, G. 1992. Cleavage of xyloglucan by nasturtium seed xyloglucanase and transglycosylation to xyloglucan subunit oligosaccharides. *Arch. Biochem. Biophys.* **298**: 365–370.
- Fincher, G.B. 1993. Cell wall metabolism in barley. In *Barley: Genetics, biochemistry, molecular biology and biotechnology* (ed. P.R. Shewry), pp. 413–437. CAB International Wallingford, Oxon, UK.
- Fischer, M. and Pleiss, J. 2003. The Lipase Engineering Database: A navigation and analysis tool for protein families. *Nucleic Acids Res.* **31**: 319–321.
- Fry, S.C. 1989. Cellulases, hemicellulases and auxin-stimulated growth: A possible relationship. *Physiol. Plant* **72**: 532–536.
- Fry, S.C., Smith, R.C., Renwick, K.F., Martin, D.J., Hodge, S.K., and Matthews, K.J. 1992. Xyloglucan endotransglycosylase, a new wall-loosening enzyme activity from plants. *Biochem. J.* **282**: 821–828.
- Goff, S.A. 2002. Collaborating on the rice genome. *Science* **296**: 45–46.
- Gouzy, J., Corpet, F., and Kahn, D. 1996. Graphical interface for ProDom domain families. *Trends Biochem. Sci.* **21**: 493.
- Guex, N. and Peitsch, M.C. 1997. SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *Electrophoresis* **18**: 2714–2723.
- Hahn, M., Keitel, T., and Heinemann, U. 1995. Crystal and molecular structure at 0.16-nm resolution of the hybrid *Bacillus* endo-1,3–1,4- β -D-glucan 4-glucanohydrolase H(A16-M). *Eur. J. Biochem.* **232**: 849–858.
- Harvey, A.J., Hrmova, M., De Gori, R., Varghese, J.N., and Fincher, G.B. 2000. Comparative modeling of the three-dimensional structures of family 3 glycoside hydrolases. *Proteins* **41**: 257–269.
- Hayashi, T. 1989. Xyloglucans in the primary cell wall. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**: 139–168.
- Higgins, D.G., Thompson, J.D., and Gibson, T.J. 1996. Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* **266**: 383–402.
- Høj, P.B. and Fincher, G.B. 1995. Molecular evolution of plant β -D-glucan endohydrolases. *Plant J.* **7**: 367–379.
- Hrmova, M. and Fincher, G.B. 2001. Structure-function relationships of β -D-glucan endo- and exohydrolases from higher plants. *Plant Mol. Biol.* **47**: 73–91.
- Ishimaru, M. and Kobayashi, S. 2002. Expression of a xyloglucan endo-transglycosylase gene is closely related to grape berry softening. *Plant Sci.* **162**: 621–628.
- Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G., and Gibson, T.J. 1998. Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* **23**: 403–405.
- Johansson, P., Brumer III, H., Baumann, M.J., Kallas, Å.M., Henriksson, H., Denman, S.E., Teeri, T.T., and Jones, A.T. 2004. Crystal structures of a poplar xyloglucan endotransglycosylase reveal details of transglycosylation acceptor binding. *Plant Cell* **16**: 874–886.
- Jones, T., Zou, J.-Y., Cowan, S., and Kjeldgaard, M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**: 100–119.
- Keitel, T., Simon, O., Borriess, R., and Heinemann, U. 1993. Molecular and active-site structure of a *Bacillus* 1,3–1,4- β -glucanase. *Proc. Natl. Acad. Sci.* **90**: 5287–5291.
- Kerr, E.M. and Fry, S.C. 2003. Pre-formed xyloglucans and xylans increase in molecular weight in three distinct compartments of a maize cell-suspension culture. *Planta* **217**: 327–339.
- Kloareg, B. and Quatrano, R.S. 1988. Structure of the cell walls of marine algae and ecophysiological functions of the matrix polysaccharides. *Oceanogr. Mar. Biol. Annu. Rev.* **26**: 259–315.
- Knogge, W. 2002. Avirulence determinants and elicitors. In *The Mycota: Agricultural applications*, Vol. XI (ed. F. Kempen), pp. 289–310. Springer Verlag, Berlin.
- Krengel, U. and Dijkstra, B.W. 1996. Three-dimensional structure of endo-1,4- β -xylanase I from *Aspergillus niger*: Molecular basis for its low pH optimum. *J. Mol. Biol.* **263**: 70–78.
- 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. Cryst.* **26**: 283–291.
- Lemesle-Varloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A., and Mornon, J.P. 1990. Hydrophobic cluster analysis: Procedures to derive structural and functional information from 2-D-representation of protein sequences. *Biochimie* **72**: 555–574.
- Louw, M.E., Reid, S.J., and Watson, T.G. 1993. Characterization, cloning and sequencing of a thermostable endo-(1,3–1,4)- β -glucanase-encoding gene from an alkalophilic *Bacillus brevis*. *Appl. Microbiol. Biotechnol.* **38**: 507–513.
- Mengden, K., Hahn, M., and Deisig, H. 1996. Morphogenesis and mechanisms of penetration by plant pathogenic fungi. *Annu. Rev. Phytopathol.* **34**: 367–386.
- Michel, G., Chantalat, L., Duee, E., Barbeyron, T., Henrissat, B., Kloareg, B., and Dideberg, O. 2001. The κ -carrageenase of *P. carrageenovora* features a tunnel-shaped active site: A novel insight in the evolution of Clan-B glycoside hydrolases. *Structure* **9**: 513–525.

- Nishitani, K. 1997. The role of endoxyloglucan transferase in the organization of plant cell wall. *Int. Rev. Cytol.* **173**: 157–206.
- Okazawa, K., Sato, Y., Nakagawa, T., Asada, K., Kato, I., Tomita, E., and Nishitani, K. 1993. Molecular cloning and cDNA sequencing of endoxyloglucan transferase, a novel class of glycosyltransferase that mediates molecular grafting between matrix polysaccharides in plant cell walls. *J. Biol. Chem.* **268**: 25364–25368.
- Page, R.D.M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Comp. Appl. Biosci.* **12**: 357–358.
- Parrish, F.W., Perlin, A.S., and Reese, E.T. 1960. Selective enzymolysis of poly- β -D-glucans, and the structure of the polymers. *Can. J. Chem.* **38**: 2094–2104.
- Pennel, R. 1998. Cell walls: Structures and signals. *Curr. Opin. Plant Biol.* **1**: 504–510.
- Ramachandran, G.N., Ramakrishnan, C., and Sasisekharan, V. 1963. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**: 95–99.
- Rose, J.K.C., Braam, J., Fry, S.C., and Nishitani, K. 2002. The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: Current perspectives and a new unifying nomenclature. *Plant Cell Physiol.* **43**: 1421–1435.
- Sali, A. and Blundell, T.L. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**: 779–815.
- Sakurai, N. and Masuda, Y. 1978. Auxin-induced changes in barley coleoptile cell wall composition. *Plant Cell Physiol.* **19**: 1217–1223.
- Schimming, S., Schwarz, W.H., and Staudenbauer, W.L. 1992. Structure of the *Clostridium thermocellum* gene licB and the encoded β -1,3–1,4-glucanase: A catalytic region homologous to *Bacillus lichenase* joined to the reiterated domain of clostridial cellulases. *Eur. J. Biochem.* **204**: 13–19.
- Schmidt, H.A., Strimmer, K., Vingron, M., and von Haeseler, A. 2002. TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**: 502–504.
- Schünmann, P.H.D., Smith, R.C., Lang, V., Matthew, R., and Chandler, P.M. 1997. Expression of XET-related genes and its relation to elongation in leaves of barley (*Hordeum vulgare* L.). *Plant Cell Env.* **20**: 1439–1450.
- Shibuya, N. and Misaki, A. 1978. Structure of hemicellulose isolated from rice endosperm cell wall: Mode of linkages and sequences in xyloglucan, β -D-glucan and arabinoxylan. *Agric. Biol. Chem.* **42**: 2267–2274.
- Shibuya, N., Misaki, A., and Iwasaki, T. 1983. The structure of arabinoxylan and arabinoglucuronoxylan isolated from rice endosperm cell walls. *Agric. Biol. Chem.* **47**: 2223–2230.
- Smith, R.C. and Fry, S.C. 1991. Endotransglycosylation of xyloglucans in plant cell suspension cultures. *Biochem. J.* **279**: 529–535.
- Smith, R.C., Matthews, P.R., Schünmann, P.H.D., and Chandler, P.M. 1997. The regulation of leaf elongation and xyloglucan endotransglycosylase by gibberellin in Himalaya barley (*Hordeum vulgare* L.). *J. Exp. Bot.* **47**: 1395–1404.
- Smith, T.F. and Waterman, M.S. 1981. Identification of common molecular subsequences. *J. Mol. Biol.* **147**: 195–197.
- Tahir, T.A., Berrin, J.G., Flatman, R., Roussel, A., Roepstorff, P., Williamson, G., and Juge, N. 2002. Specific characterization of substrate and inhibitor binding sites of a glycosyl hydrolase family 11 xylanase from *Aspergillus niger*. *J. Biol. Chem.* **277**: 44035–44043.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, 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.* **25**: 4876–4882.
- Wu, Y., Spollen, W.G., Sharp, R.E., Hetherington, P.R., and Fry, S.C. 1994. Root growth maintenance at low water potentials (increased activity of xyloglucan endotransglycosylase and its possible regulation by abscisic acid). *Plant Physiol.* **106**: 607–615.
- Yahata, N., Watanabe, T., Nakamura, Y., Yamamoto, Y., Kamimiya, S., and Tanaka, H. 1990. Structure of the gene encoding β -1,3-glucanase A1 of *Bacillus circulans* WL-12. *Gene* **86**: 113–117.