Structure-Function Analysis of a Broad Specificity *Populus trichocarpa* Endo-β-glucanase Reveals an Evolutionary Link between Bacterial Licheninases and Plant XTH Gene Products

Received for publication, February 20, 2013, and in revised form, March 28, 2013. Published, JBC Papers in Press, April 9, 2013, DOI 10.1074/jbc.M113.462887

Jens M. Eklöf,* Shaheen Shojania, Mark Okon, Lawrence P. McIntosh, and Harry Brumer

From the ‡ Division of Glycoscience, School of Biotechnology, Royal Institute of Technology (KTH), SE-106 91 Stockholm, Sweden, the § Michael Smith Laboratories, University of British Columbia, Vancouver V6T 1Z4, British Columbia, Canada, the ¶ Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada, and the ¶¶ Department of Chemistry, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada

**Background:** The evolution of the plant xyloglucan endotransglycosylase/hydrolase (XTH) genes in glycoside hydrolase family 16 (GH16) is enigmatic.

**Results:** A unique, mixed function endo( xylo)glucanase from black cottonwood has been biochemically and structurally characterized.

**Conclusion:** This enzyme is an important link between extant bacterial endoglucanases and plant XTH gene products.

**Significance:** New insights into the molecular evolution of XTH gene products and further unification of GH16 enzymes have been gained.

The large xyloglucan endotransglycosylase/hydrolase (XTH) gene family continues to be the focus of much attention in studies of plant cell wall morphogenesis due to the unique catalytic functions of the enzymes it encodes. The XTH gene products compose a subfamily of glycoside hydrolase family 16 (GH16), which also comprises a broad range of microbial endoglucanases and endogalactanases, as well as yeast cell wall chitin/β-glucan transglycosylases. Previous whole-family phylogenetic analyses have suggested that the closest relatives to the XTH gene products are the bacterial licheninases (EC 3.2.1.73), which specifically hydrolyze linear mixed linkage β(1→3)/β(1→4)-glucans. In addition to their specificity for the highly branched xyloglucan polysaccharide, XTH gene products are distinguished from the licheninases and other GH16 enzyme subfamilies by significant active site loop alterations and a large C-terminal extension. Given these differences, the molecular evolution of the XTH gene products in GH16 has remained enigmatic. Here, we present the biochemical and structural analysis of a unique, mixed function endoglucanase from black cottonwood (*Populus trichocarpa*), which reveals a small, newly recognized subfamily of GH16 members intermediate between the bacterial licheninases and plant XTH gene products. We postulate that this clade comprises an important link in the evolution of the large plant XTH gene families from a putative microbial ancestor. As such, this analysis provides new insights into the diversification of GH16 and further unites the apparently disparate members of this important family of proteins.

Plant cell walls are complex biocomposites, comprising semicrystalline and amorphous polysaccharides, polyphenolics, structural proteins, and inorganics, which are of biological, ecological, and industrial importance (1–3). In addition to forming the basis of the agricultural and forest products industries, plant cell walls represent a vast sink in the global carbon cycle. Thus, understanding molecular mechanisms of plant cell wall biosynthesis and morphogenesis is fundamentally important (4).

Depending on the cell type and developmental stage, carbohydrates typically compose at least three-fourths of the plant cell wall dry weight (1, 5). Semicrystalline cellulose is the main load-bearing polymer embedded in a hydrated network of matrix glycans. A diversity of primary plant cell wall polysaccharide compositions can be observed (6). For example, the walls of dicots and non-commeneloid monocots contain xyloglucans as the primary matrix polysaccharide, whereas the grasses utilize mixed linkage β(1→3)/β(1→4)-glucans in this role (1, 5). Due to this predominance of structural carbohydrates, the biosynthesis and enzyme-catalyzed rearrangement of polysaccharides underpin contemporary cell wall models (4). Indeed, plant genomes encode large repertoires of glycosyltransferases, glycoside hydrolases, and transglycosidases, often in large, multigene families (7–10).

In the context of postbiosynthetic plant cell wall remodeling, there is a sustained interest in the xyloglucan endotransglyco-
sylase/hydrolase (XTH) gene family. This is due to the remarkable ability of some gene products to catalyze matrix polysaccharide rearrangement with essentially no chain hydrolysis (XET activity, EC 2.4.1.207). On the other hand, a limited number of XTH genes encode predominant hydrolases (XEH, EC 3.2.1.151) that operate in germinating seeds, ripening fruit, or expanding tissues (11, 12). Notably, contemporary enzyme structure-function analyses have led to a refined phylogenetic delineation of the disparate transglycosylating and hydrolytic activities individual XTH gene products, which number in the range of ~20–60 in diverse plants (11, 13).

However, the molecular evolution of the XTH gene products in the larger context of glycoside hydrolase family 16 (GH16) (14) remains enigmatic. In addition to the plant XTHs, GH16 counts a broad range of microbial endogalactanases and endogalactanases among its functionally characterized members, spanning those active on terrestrial and marine polysaccharides to yeast cell wall chitin/β-glucan transglycosylases (13). Despite significant sequence divergence, all GH16 enzymes are predicted to share both a common overall β-jellyroll protein fold and the canonical retaining GH catalytic mechanism (15–17).

Previous whole-family phylogenetic analyses have suggested that the closest relatives to the XTH gene products are the bacterial licheninases (EC 3.2.1.73), which specifically hydrolyze β(1→4) linkages in mixed linkage β(1→3)/β(1→4)-glucans (15). XTH gene products are, however, distinguished from the licheninases and all other GH16 enzyme classes by significant differences in their common β-jellyroll protein structure. These differences include major loop alterations and a unique C-terminal extension, in addition to a singular specificity for a highly branched polysaccharide substrate (13). Exactly how the XTH gene products may have arisen from licheninases has heretofore remained unclear.

Here, we present the biochemical and structural analysis of an unusual, mixed function endoglucanase from black cottonwood (Populus trichocarpa). This research reveals a small clade of GH16 members intermediate between the bacterial licheninases and plant XTH gene products. We postulate that this clade comprises an important link in the evolution of the large plant XTH gene families from a putative microbial ancestor. As such, this analysis provides new insights into the diversification of GH16 and further unites the apparently disparate members of this important family of proteins.

**EXPERIMENTAL PROCEDURES**

Substrates—Xyloglucan, barley mixed linkage glucan (medium viscosity), Icelandic moss lichenan, laminarin, lupin β-galactan, wheat arabinoxylan, carboxymethyl cellulose (CMC), celluloligosaccharides, and mixed linkage glucooligosaccharides were from Megazyme. Hydroxethyl cellulose was from Fluka. Xyloglucooligosaccharides were prepared as described previously (18).

Bioinformatic and Phylogenetic Analyses—The EG16-like enzymes were found using PEG16 as the query with BLASTp at the Phytozone Web site (October 2012). The Charophyta sequence CHARA2 was manually transcribed from Ref. 19, and an expressed sequence tag from Coleochaete nitellum was obtained from GenBankTM (accession number HO204633). Possible localization and post-translational modifications were analyzed by SignalP (20), ChloroP (21), LipoP (22), and GPP (23). These sequences were also included in the phylogeny: XTH gene products from rice, *Oryza sativa* (OsXTHs), and *Arabidopsis* (AtXTHs); Group III-A sequences from Ref. 11; and selected Group IIIA sequences from TIGR (24). Two *Bacillus* licheninases (PDB codes 1bgg and 1u0a) were included to root the tree. Sequences were aligned using MUSCLE (25), and the alignment was edited manually in Bioedit (26), guided by the available three-dimensional structures. Maximum likelihood and Bayesian phylogenies were built using PhyML 3.0 (27) and MrBayes 3.1.2 (28), respectively. In PhyML, the reliability of nodes was tested by 100 resamplings. In MrBayes, 2 × 10⁶ generations were run with a sample frequency of 100. Blosum62 was used as the amino acid substitution matrix in both PhyML and MrBayes. Trees were drawn with MEGA5 (29).

**Protein Expression and Purification—**PEG16 cDNA (originally codon-optimized for *Pichia pastoris* expression; GENERAT AG) was cloned into Expresso N-His vector with an N-terminal His₆ tag (Lucigen). PEG16 was produced in H1 control *Escherichia coli* BL21 (DE3) cells (Lucigen) grown in Terrific Broth at 37 °C at 200 rpm to an A₅₆₀ of 0.8 and then induced by the addition of isopropyl β-D-galactopyranoside to a final concentration of 0.5 mM. During the induction phase, the culture was kept at 25 °C overnight. Cells were then collected by centrifugation at 4800 × g for 10 min at 4 °C. The cells were resuspended in buffer A (25 mM sodium phosphate, 0.5 mM NaCl, and 25 mM imidazole, pH 7.5) and ultrasonicated to liberate cytoplasmic proteins. The supernatant was collected by centrifugation at 24,700 × g for 15 min at 4 °C and passed through a 5-ml HiTrap FF Crude column (GE Healthcare). The protein was eluted using a 10-column volume-long linear gradient with buffer B (25 mM sodium phosphate, 0.5 mM NaCl, and 0.5 mM imidazole, pH 7.5). Fractions containing PEG16 were pooled, and DTT was added to a final concentration of 0.5 mM. After incubation at room temperature for 2 h, the sample was purified further using an XK 16/100 Superdex 75 column (GE Healthcare), pre-equilibrated with 20 mM MOPS, 1 mM tris(2-carboxyethyl)phosphine, pH 7.5. Monomeric PEG16 fractions were pooled, concentrated using Vivaspin20 5 kDa (PES) centrifugal concentrators, and stored at 4 °C.

PEG16 enriched in ¹⁵N for NMR experiments was produced using *E. coli* cells grown at 37 °C at 200 rpm to an A₅₆₀ of 1 in LB, centrifuged at 3000 × g for 10 min, and then gently resuspended in M9 medium containing 1 g/liter ¹⁵NH₄Cl and 0.5 mM isopropyl β-D-galactopyranoside. The induced culture was kept at 16 °C for 30 h before the cells were collected by centrifugation at 4800 × g for 10 min at 4 °C. The ¹⁵N-labeled PEG16 was treated and purified as described above for the non-labeled enzyme.
Triple-labeled PtEG16-1 (2D, 13C, and 15N) was produced by first inoculating 600 ml of M9 medium with 3 ml of overnight LB culture (H2O) and kept at 37 °C at 200 rpm until an A600 of 0.5. The cells were gently pelleted by centrifugation at 1000 × g for 8 min and resuspended in 600 ml of M9 medium containing 1 g/liter 15NH4Cl, 2.5 g/liter 2H2/13C6-glucose in D2O. At A600 0.65, the culture was moved to 25 °C and induced after 20 min with isopropyl β-D-galactopyranoside to a final concentration of 1 mM. After overnight induction, the cells were harvested at 4800 × g for 15 min at 4 °C. Cells were sonicated and initially purified on a 5-ml HisTrap FF crude column as for the unlabeled protein. After initial purification, the buffer was exchanged to 25 mM MOPS, 2 mM tris(2-carboxyethyl)phosphine, pH 7.5. To reintroduce 1H on the nitrogens, the triple-labeled PtEG16 was denatured by 1:20 dilution in 8 mM urea, 25 mM Tris, pH 7.5, with 2 mM DTT prepared in H2O. On-column refolding was performed essentially as described (30), with the exception that 2 mM DTT was added to all buffers. The refolded triple-labeled PtEG16 was buffer-exchanged to 20 mM MOPS, 2 mM tris(2-carboxyethyl)phosphine, pH 7.5, before NMR experiments.

Activity on Polysaccharides Detected by Gel Permeation Chromatography—Assays in a total volume of 150 μl, containing 2.5 g/liter polysaccharide, 30 mM NH4OAc, 0.25 mM DTT, and 24 μg/ml PtEG16, were incubated for 0, 6, 20, 60, 180, and 480 min and overnight at 22 °C. The reactions were stopped by heating to 95 °C for 10 min. The samples were then lyophilized and dissolved in 200 μl of DMSO before analysis by gel permeation chromatography, as described previously (18). Control assays without enzyme were also monitored to detect background hydrolysis. Carboxymethyl cellulose and Konjac glucosaminan (Megazyme) were insoluble in DMSO and were therefore analyzed by high performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD).

HPAEC-PAD—Samples were analyzed on a Dionex ICS-5000 HPAEC-PAD system using a Dionex CarboPac PA-200 column. Four different programs (supplemental Table S1) were used, depending on the sample. Program A was used to identify the limit digestion products of lichenan and mixed linkage glucans versus standard samples of cello-oligosaccharides (degree of polymerization 1–6) and three mixed linkage glucan tetraoses: MLGA, G3G4G4G; MLGB, G4G4G3G; and MLGC, G4G3G4G. Program B (18) was used to analyze the limit digestion products of all other polysaccharides. Potential transglycosylation products at maximal substrate concentrations (10-min reactions) were analyzed by Program C (for cello-oligosaccharides and mixed linkage glucan oligosaccharides) and Program D (for XXXGXXXG).

Quantitative Kinetic Analysis of PtEG16 Activity on Oligosaccharides—Kinetic quantitation under initial rate conditions was determined on cello-oligosaccharides (degree of polymerization 2–6) and the xyloglucan-oligosaccharide XXXGXXXG using HPAEC-PAD. The pH optimum of the enzyme was determined by incubation of 30 nM PtEG16 with 400 μM cellohexaose and 50 mM buffer (sodium citrate for pH 3–6.5 and MOPS for pH 6.5–8.1) in a final volume of 40 μl. All subsequent assays with oligosaccharide substrates were performed in 40-μl reactions buffered with 25 mM sodium citrate, pH 5.25. In all cases, assays were maintained at 22 °C and terminated by the addition of 10 μl of 1 N NaOH prior to analysis using Program A (supplemental Table S1). Peak areas were quantified by integration and converted to molar amounts based on standard curves in the range 1–50 μM.

Matrix-assisted Laser Desorption Ionization Time-of-flight (MALDI-TOF) Analysis—Oligosaccharide products were analyzed by MALDI-TOF mass spectrometry in positive ion mode on an LT3 Plus mass spectrometer (SAI Ltd.) operated by the MALDI Mainframe 2, MALDI Control software (version 1.03.51, SAI Ltd.). Samples were supplemented with NaCl to a final concentration of 20 mM, and 2,5-dihydroxybenzoic acid (10 g/liter in water) was used as the matrix.

NMR Spectroscopy—Resonance assignments for the main chain 1H, 15N, and 13C nuclei in the 220-residue carbon-per-deuterated 13C/15N-labeled His6-tagged PtEG16 were obtained using standard pulse sequences recorded at 25 °C on a Bruker Avance III 600-MHz spectrometer equipped with a TCI Cryo-Probe. The resulting 15N TROSY HSQC (31–35), HNCA CB (36, 37), HNCOCAB (36, 37), HNCO (38), and HN(CA)CO (37) spectra were processed with NMRPipe (39) and analyzed with Sparky (40). The 1H and 15N dimensions were corrected by +46 Hz and −46 Hz, respectively, to adjust for the TROSY shift. Chemical shift-based secondary structural predictions were obtained using the programs MICS (41), TALOS+ (42), and 82D (43). Referencing and adjustment of the carbon shifts for deuterium isotope effects were done using TALOS+ (42).

RESULTS

Bioinformatics Analysis Reveals a Unique XTH-like Gene Product of the P. trichocarpa Genome—Using amino acid sequence alignments, we identified a unique protein encoded by the P. trichocarpa genome (genome version 2.2, locus P0PTR_0002s15460, also known as gene3.00021425 or PtXTH8 (8)). This protein clearly lacked the C-terminal extension diagnostic of GH16 XTH gene products (44, 45) ( Pfam XET_C (PF06955) (46) but otherwise showed significant sequence similarity to known XETs and XEHs (Fig. 1). A BLASTp search of this sequence against predicted proteomes available at Phytozome revealed that the genomes of many embryophytes encode one or more homologous proteins with Expect (E) values less than 10−80. In comparison, proteins possessing the XTH-specific C-terminal extension typically had E values above 10−35 in this analysis (data not shown), which further strengthened the conclusion that the POPTR_0002s15460 gene product (hereafter renamed as “P. trichocarpa endoglucanase 16” (PtEG16) in accordance with biochemical data; see below) and its homologs indeed composed a separate clade. A multiple-protein sequence alignment of these homologs from diverse plants is shown in supplemental Fig. S1. Notably, none has predicted signal peptides for apoplast, organelle, or membrane localization, according to SignalP (20), ChloroP (21), and LipoP (22).

We then performed independent Bayesian and maximum likelihood phylogenetic analyses with these sequences and all 33 Arabidopsis thaliana and 29 Oryzae sativa XTH gene products (9) as representative dicot and monocot sequences, respec-
tively. Crystallographically characterized XETs and XEHs (11, 45) as well as select phylogenetic Group III-A XTH gene products were also included in this analysis. The sequences of two Bacillus GH16 licheninases with known tertiary structures (47, 48) were used as more distant outliers to root the phylogenetic trees. In this analysis, only the common GH16 domain was used (i.e. the unique XTH gene product C terminus was excluded from the alignments).

Notably, the inclusion of the PtEG16-like sequences in Bayesian and maximum likelihood phylogenies did not alter the overall relationships of the true XTH gene products observed previously (9, 11, 49, 50). Rather, PtEG16 and homologs formed a major distinct clade with strong statistical support (Fig. 2A). Relative tree branch lengths further indicated that this clade is intermediate between the true XTH gene products and their presumed bacterial ancestors, the licheninases (15). Monocot and dicot sequences are clearly separated in the tree, thus indicating a potential divergent evolution of PtEG16-like proteins in these two groups of flowering plants. A similar distinction can be seen among the hydrolytic Group III-A XTH gene products, where monocot, dicot, and non-angiosperm sequences cluster separately (Fig. 2A).

A Census of PtEG16-like Proteins in Plants—After establishing that the PtEG16-like proteins formed a unique phylogenetic clade, we were then able to examine their presence or absence as a class within predicted plant proteomes available via Phytozone (51), thereby extending our previous census of Group I/II, III-A, and III-B XTH gene products (13). Although plants generally maintain large XTH gene families with tens of members, the PtEG16-like proteins occur in very limited numbers and are not found across all lineages (Fig. 2B). Notably, members of the clade containing PtEG16 are not found in the two chlorophytic algal genomes presently available; these genomes also lack XTH or licheninase-encoding genes. In the current absence of whole genome sequences, we were able to identify two partial expressed sequence tag sequences from charophycean green algae (Chara vulgaris CHARA2 (19) and C. nitellarum GenBank™ accession number HO204633 (52)) whose proteins appear to exhibit greater resemblance to PtEG16 than to true XTH gene products (supplemental Fig. S1and Fig. 2A). These observations imply that PtEG16-like proteins may have first arisen in the charophytes.

Regardless, PtEG16-like proteins are conclusively found in an extant member of one of the oldest lineages of land plants, the moss Physcomitrella patens, which has three homologs. The lycophyte Selaginella moellendorfii, an early tracheophyte, contains only one PtEG16 homolog. Among the angiosperms, all available grass in silico proteomes (order Poales) likewise harbor one PtEG16 homolog each. The grass PtEG16 homologs are, however, distinguished by longer C termini, which none-
theless do not have significant similarity to the diagnostic C-terminal extensions of true XTH gene products (supplemental Fig. S1). Interestingly, PtEG16 homologs are not consistently found in dicotyledonous genomes. For example, they are missing in *Medicago truncatula* but are found in *Glycine max*; both are species within the Fabales. PtEG16 homologs are notably absent from the predicted proteomes of two *Arabidopsis* model species as well as other important dicots (Fig. 2B).

Biochemical Analysis Defines PtEG16 as a Broad Specificity Endo(xylo)glucanase—To elucidate the function of the *P. trichocarpa* POPTR_0002s15460 gene product, the protein was produced recombinantly in *E. coli*. Purification to homogeneity was achieved by immobilized metal ion affinity chromatography and subsequent gel filtration chromatography. The protein was then assayed for activity against typical polysaccharide substrates for the most closely related GH16 enzymes (i.e. the mixed linkage $\beta(1\rightarrow3)/\beta(1\rightarrow4)$-glucans from barley and Icelandic moss and the highly branched galactoxyloglucan from tamarind seed, as well as wheat arabinoxylan, lupin $\beta$-galactan, laminarin (a mixed linkage $\beta(1\rightarrow3)/\beta(1\rightarrow6)$-glucan), hydroxethyl cellulose, and CMC).

Initial activity screening of purified recombinant PtEG16 against these polysaccharides was unsuccessful. Closer inspection of the amino acid sequence and a tertiary structure homology model (see below) revealed that the 25,000Da protein had 11 cysteine residues, 5–7 of which were likely to be surface-exposed. Subsequently, we assayed the protein in the presence of DTT and analyzed polysaccharide depolymerization by gel permeation chromatography; the presence of the reducing agent effectively precluded the use of carbohydrate reducing end measurement chromatography; the presence of the reducing agent DTT and analyzed polysaccharide depolymerization by gel permeation chromatography. The protein was then assayed for activity against typical polysaccharide substrates for the most closely related GH16 enzymes (i.e. the mixed linkage $\beta(1\rightarrow3)/\beta(1\rightarrow4)$-glucans from barley and Icelandic moss and the highly branched galactoxyloglucan from tamarind seed, as well as wheat arabinoxylan, lupin $\beta$-galactan, laminarin (a mixed linkage $\beta(1\rightarrow3)/\beta(1\rightarrow6)$-glucan), hydroxethyl cellulose, and CMC).

Initial activity screening of purified recombinant PtEG16 against these polysaccharides was unsuccessful. Closer inspection of the amino acid sequence and a tertiary structure homology model (see below) revealed that the 25,000Da protein had 11 cysteine residues, 5–7 of which were likely to be surface-exposed. Subsequently, we assayed the protein in the presence of DTT and analyzed polysaccharide depolymerization by gel permeation chromatography; the presence of the reducing agent effectively precluded the use of carbohydrate reducing end assays, such as the BCA assay (53) or the Nelson-Somogyi assay (54). In this way, the ability of PtEG16 to efficiently cleave barley $\beta$-glucan, Icelandic moss lichenan, and tamarind seed xyloglucan was clearly revealed (Fig. 3). In contrast, no change in polysaccharide molecular mass was observed under identical conditions for wheat arabinoxylan, lupin galactan, or laminarin, even after extensive incubation (24 h with 1 $\mu$m PtEG16; data not shown).

PtEG16 was also active on the artificial, soluble cellulose derivatives hydroxethyl cellulose and CMC. The activity of PtEG16 on hydroxethyl cellulose was complex; an initial rapid molecular mass shift was observed, followed by a much slower phase of limited degradation, which may indicate that only a few sites on this modified cellulose were susceptible to PtEG16-1 (supplemental Fig. S2). Analysis of CMC degradation was complicated by the observation that this polymer, as obtained from the supplier, was insoluble in the gel permeation chromatography eluent (100% DMSO). An overnight digest sample of CMC was, however, soluble in this solvent, which was indicative of reduction of the polymer molecular mass. HPAEC-PAD (data not shown) and MALDI-TOF MS (supplemental Fig. S3) analysis of this reaction subsequently revealed the presence of short cello-oligosaccharides and carboxymethylated derivatives.

Limit Digestion Products from Mixed Linkage Glucans—Further analysis of the products from overnight digestion of mixed linkage $\beta(1\rightarrow3)/\beta(1\rightarrow4)$-glucans and (galacto)xyloglucan using HPAEC-PAD and MALDI-TOF MS confirmed that PtEG16 is indeed a broad specificity endo-$\beta(1\rightarrow4)$-glucanase. Upon extended incubation with the enzyme, barley $\beta$-glucan and laminarin were reduced to glucose, disaccharides (cellobiose/laminaribiose), and tetrasaccharides as predominant products (supplemental Fig. S4). Closer analysis of the tetrasaccharide peak resulting from either substrate (supplemental Fig. S4) indicated that this represented a single isomer with a retention time distinct from the available G4G4G3G, G3G4G4G, and G4G3G4G tetrasaccharide standards (where “G” represents Glc and the Arabic numeral represents a corresponding $\beta(1\rightarrow3)$ or $\beta(1\rightarrow4)$ glycosidic linkage). Considering the native polysaccharide structure (55), the other possible tetrasaccharide products from the mixed linkage glucans are cellotetraose (G4G4G4G) and G3G4G3G. Because the retention time of cellotetraose is shorter than the available mixed linkage tetrasaccharide standards under these HPAEC conditions (data not shown), we tentatively assign the observed product peak as G3G4G3G.

When explicitly tested for activity on the three commercially available mixed linkage tetraoses, PtEG16 only hydrolyzes the central $\beta(1\rightarrow4)$ linkage of G4G4G3G, thereby yielding G4G and G3G. This suggests that the possible modes for hydrolysis of the $\beta(1\rightarrow4)$ glucosidic bonds in G3G4 ↓ G ↓ G and G4 ↓ G3G4 ↓ G (potential cleavage sites indicated by arrows) are catalytically insignificant for these short substrates due to geometrical requirements (i.e. rejection of Glc$\beta(1\rightarrow3)$ in sub-sites $−1$, $−2$, and $−3$) and/or a lack of a sufficient number of glucosyl units bound in the positive or negative enzyme subsites (see Ref. 56 for GH subsume nomenclature). Nonetheless, the production of G3G4G3G from mixed linkage glucans, as suggested above, would require acceptance of Glc$\beta(1\rightarrow3)$ units in subsite $−2$. This implies that extended binding of polysaccharides in the active-site cleft may overcome such limitations. Regardless, the mode of action of PtEG16 on mixed linkage glucans appears distinct from that of archetypical GH16 licheninases, which explicitly require a $\beta(1\rightarrow3)$ linkage to span the $−2$ and $−1$ subsites to yield limit digestion products of the series G3G, G4G3G, G4G4G3G, and G4G4G4G3G (16).

Limit Digestion Products from Galactoxyloglucan—Despite a long incubation time, the hydrolysis of tamarind seed galactoxyloglucan by PtEG16 did not go to the same level of completion (supplemental Fig. S5) previously observed for plant and microbial XEHs (12, 18). In addition to the mixture of oligosaccharides expected from cleavage at the unbranched $\beta(1\rightarrow4)$-
FIGURE 3. Time-dependent endohydrolytic cleavage of polysaccharides by PtEG16. The action of PtEG16 on barley β-glucan (A), Icelandic moss lichenin (B), and tamarind seed galactoxyloglucan (C) was analyzed using high performance size exclusion chromatography with evaporative light-scattering detection (ELS). The t = 0 trace corresponds to that of the corresponding polysaccharide before the addition of the enzyme. Compositional analyses of the oligosaccharide mixtures resulting from overnight incubation are shown in supplemental Figs. S3–S5. Peaks marked with an asterisk arise from low components the enzyme preparation (e.g. buffer and salts).

linked glucosyl residues of the polysaccharide backbone (i.e. XXXG, XLXG, XXLG, and XLLG), a series consisting of longer oligosaccharide repeats was clearly observed by HPAEC-PAD (xyloglucooligosaccharide nomenclature is according to Ref. 57), where G represents GlcP or β-GlcP-(1→4), X is [α-Xylp-(1→6)]-β-GlcP-(1→4), and L is [β-Galp-(1→2)-α-Xylp-(1→6)]-β-GlcP-(1→4). Moreover, HPAEC-PAD and MALDI-TOF MS indicated that the ratio of the shortest oligosaccharide products deviated from the expected ratio of 1:4:3:5:4 XXXG/XLXG/XXLG/XLLELG (58, 59). Specifically, the amount of the bis-galactosylated XLLELG (Glc3Xyl3Gal) appeared to be reduced (supplemental Fig. S5), which suggests that extended chain branches may limit hydrolysis.

Initial Rate Kinetics on Defined Xyloglucooligosaccharides and Celloooligosaccharides—To further analyze the effect of polysaccharide branching on catalysis, we quantified the hydrolysis of well defined xylogluco- and cello-ooligosaccharides by PtEG16 under initial rate kinetic conditions. The tetradecasaccharide XXXG ↓ XXXG was exclusively hydrolyzed at the internal, unbranched β(1→4)-linked glucosyl residue (cleavage site indicated by an arrow), with a $k_{cat}$ value of 62 ± 6 min$^{-1}$ and a $K_m$ value of 295 ± 64 μM (Fig. 4A). Under these conditions of low substrate conversion, transglycosylation to form (XXXG)$_3$ by disproportionation of the substrate was kinetically insignificant. However, after extended incubation of 1 mM XXXGXXXG with PtEG16, minor amounts of the transglycosylation products (XXXG)$_{3-5}$ could be observed by HPAEC-PAD (supplemental Fig. S6A). In an independent assay, the heptasaccharide XXXG (Glc$_3$Xyl$_3$) was not hydrolyzed by PtEG16-1, thus providing further support that the enzyme can only cleave the glycosidic bond of non-xylolyslated glucosyl units.

In contrast to the singular mode of cleavage of XXXGXXXG by the enzyme, unbranched cello-ooligosaccharides gave rise to multiple products, which complicated initial rate kinetic analysis. The longest water-soluble cello-ooligosaccharide, cellohexaose, produced Glc$_3$ as well as pairwise equimolar amounts of Glc and Glc$_3$, and Glc$_2$ and Glc$_4$ (Fig. 4B). The Glc$_3$/Glc$_4$ hydrolysis mode dominated, with a rate approximately equal to that of XXXGXXXG hydrolysis at saturation. Summation of all cellohexaose hydrolysis modes yielded a rate approximately twice that of XXXGXXXG at saturation (Fig. 4B). As with XXXGXXXG, extended incubation of cellopentaose with PtEG16-1 yielded apparent transglycosylation products (up to Glc$_{11}$ in this case), as suggested by HPAEC-PAD (supplemental Fig. S6B).

Whereas the initial rate kinetics of cellohexaose were apparently uncomplicated by transglycosylation modes, this was not the case for cellopentaose and cellooctaose. For cellopentaose, a clear discrepancy in the rates of formation of Glc$_2$ and Glc$_3$ was observed across a range of initial substrate concentrations (supplemental Fig. S7A). Here, simple hydrolysis of either of the two indicated linkages, G4G4 ↓ G4 ↓ G4G4, would be expected to yield equimolar amounts of the products. The observation of a HPAEC peak with a retention time longer than cellohexaose (tentatively assigned as Glc$_{11}$, in the absence of a standard sample) was further evidence for transglycosylation.

The initial rate kinetics observed for cellooctaose were likewise complicated, including the observation of increasing production of cellohexaose with increasing substrate concentration (supplemental Fig. S7B). Formation of this product is
consistent with initial binding of cellotetraose in subsites −2 to +2, followed by cleavage to form a covalent, α-linked cellobiosyl enzyme intermediate, which is subsequently turned over by glycosyl transfer to a second molecule of cellotetraose. Of the shorter congeners, cellotriose was a very poor substrate, and cellobiose was not a substrate for PtEG16 (data not shown).

Analysis of Heterotransglycosylation Potential—In light of recent interest in the capacity of certain GH16 members and crude plant enzyme extracts to catalyze xylolucan/β-glucan heterotransglycosylation (60, 61), we examined the reaction products from two representative glycosyl donor/acceptor substrate pairs by HPAEC-PAD. For the XXXGXGG/cellobiose pair, the product distribution was identical to that of XXXGXGXXG alone (i.e. only longer (XXXG), products were observed (data not shown)). In the case of cellotetraose/XXXG, repeated analysis indicated the presence of very low abundance peaks, possibly corresponding to GXXG and GGXXXG, in addition to dominating hydrolysis and homotransglycosylation products (data not shown). However, the low amounts of these alternative products precluded isolation and definitive assignment, and we conclude that heterotransglycosylation is not kinetically significant for PtEG16.

Tertiary Structure of PtEG16 and Comparison with GH16 XEHs, XETs, and Licheninases—PtEG16 has thus far resisted crystallization.3 We therefore performed in silico structure homology modeling, supported by experimental protein NMR spectroscopic analyses, to obtain a three-dimensional representation of the enzyme. The M4T Server version 3.0 (62) selected a bacterial GH16 licheninase (PDB code 2ayh (63)) and a plant GH16 XEH (PDB code 2uwa (11)) as templates for modeling full-length PtEG16. For comparison, we also subjected PtEG16 to structural modeling using Protein Homology/Analysis Recognition Engine version 2.0 (Phyre2 (64)). Both approaches produced tertiary structures with a β-jellyroll fold typical of GH16 enzymes. Moreover, superimposition of the Ca traces indicated that these predicted structures were nearly identical (Fig. 5A; root mean square deviation values are given in supplemental Table S3).

To provide experimental validation for the models, we investigated the recombinant PtEG16 using NMR spectroscopy. Under reducing conditions, the protein yielded an excellent quality, well dispersed 15N TROSY-HSQC spectrum with peak line widths indicative of a stably folded, monomeric protein (Fig. 5B). In total, of the 213 residues (not including the His₆ tag), main chain 1H, 15N, 13C, 13C, and 13C nuclei from 166 were unambiguously assigned via a manual analysis of standard triple resonance correlation experiments. An additional 14 spin systems were assigned using the PINE assignment server (65). Chemical shift assignments for only the 13C nuclei were obtained for a further 29 residues (including 8 of the 11 prolines). Four residues were left without any assigned resonances (Pro-22, Pro-163, Gly-183, and Ser-213). The non-proline residues that are without amide proton and nitrogen assignments are located throughout the protein (supplemental Fig. S8). The lack of assignments may reflect spectral overlap, rapid amide hydrogen exchange, or conformational exchange broadening.

The NMR chemical shifts of main chain nuclei are sensitive indicators of protein secondary structure (66, 67). We therefore used the SPP (68), MICS (41), and TALOS+ (42) algorithms to determine the secondary structural elements of PtEG16 from the assigned chemical shifts. All three algorithms yielded similar results, identifying numerous β-strands with a limited number of short helical regions. Importantly, these secondary structural elements, derived from experimental data, agreed well with those present in the M4T structural model, as defined by VADAR (69) (supplemental Fig. S8).

Further experimental validation of the model was obtained by calculating the overall PtEG16 fold from backbone chemical shift data using the CS23D Web server (70). The resulting
FIGURE 5. Modeling and experimental validation of the PtEG16 tertiary structure. A, superimposition of the M4T (marine) and Phyre2 (magenta) homology models and the chemical shift-derived CS23D (orange) model of PtEG16 in a wall-eyed stereo representation. Models are oriented with the positive enzyme subsites on the right, relative to the conserved active-site residues Glu-88, Asp-90, and Glu-92 (bottom to top) of PtEG16 (cf. Fig. 1). B, assigned $^{15}$N TROSY-HSQC spectrum of carbon-perdeuterated and $^{13}$C,$^{15}$N-labeled PtEG16 in H$_2$O buffer at pH 7.5 and 25 °C.
model superimposes very well the homology models from M4T and Phyre2 (Fig. 5A and supplemental Table S3). Moreover, superimposition with the M4T templates (licheninase (PDB code 2ayh) and XEH (PDB code 2uwa)), indicate that the PtEG16 models possess the canonical GH16 fold with a broader active-site cleft, which is typical of plant XTH gene products (supplemental Fig. S9 and Table S3; cf. Fig. 6).

Inspection of these superimposed models indicates that this broadening of the active-site cleft has resulted from the deletion of a 12-amino acid sequence found in the bacterial licheninases (Fig. 1, residues 23–34). This loop extends into the concave face of the β-jellyroll to form a part of the negatively numbered enzyme subsites (supplemental Fig. S9A), thereby narrowing the active-site cleft (Fig. 6) and contributing many structural features necessary for specific recognition of kinked mixed linkage glucan chains (16, 48, 71). The removal of this loop in plant PtEG16 and XTH gene products is a clear prerequisite for the binding of highly branched xyloglucans (72–74).

In the positively numbered subsites, PtEG16 resembles a licheninase, primarily because it lacks the C-terminal extension characteristic of true XTH gene products (supplemental Fig. S9B). In XETs and XEHs, this motif elongates the substrate binding cleft by providing one β-sheet, narrows the positive subsites (Fig. 6), and supplies two-thirds of a Xyl +2’ xyloglucan specificity pocket that affects specificity \( k_{ca}/K_m \) 500-fold (72, 75). Like the licheninases, PtEG16 also lacks the small loop insertion immediately following the catalytic motif (NRT in PttXET16-34; Fig. 1), which further affects the structure of the positively numbered subsites in XTH gene products and also contains the conserved N-glycosylation site important for the stability of XTH Group I/II members (76, 77).

DISCUSSION

Barbeyron and colleagues (15, 78) were among the first to speculate upon the evolutionary basis of the diverse catalytic functionality observed among GH16 members. Although enzyme structure-function analyses continue to shed light on protein loop differences that fine-tune specificity for particular linear galactans and glucans among bacterial members (15, 79, 80), the evolutionary steps leading to the dramatic structural features of plant XTH gene products have remained essentially unknown.

Taken together, our present data suggest that a unique protein in black cottonwood, PtEG16, is a broad specificity glycoside hydrolase with nearly equal capacity to cleave linear glucans (mixed linkage β-glucans and cello-oligosaccharides) and highly branched xyloglucan oligo- and polysaccharides (e.g. Figs. 3 and 4). Moreover, this broad specificity appears to arise from a protein scaffold that is intermediate between extant bacterial licheninases and plant XTH gene products (Figs. 2 and 5).

As such, these observations allow us to posit that PtEG16-like sequences represent an ancestral link in the evolution of extant plant XETs and XEHs.

Fig. 6 highlights a view of this potential evolution from the perspective of protein tertiary structure, supported by sequence-based phylogeny (Fig. 2A). In this scheme, an early
GH16 ancestor with a structure similar to extant licheninases may have first given rise to a PtEG16-like protein via deletion of the licheninase-specific negative subsite loop. Although we cannot rigorously exclude the possibility that a PtEG16-like protein was the ancestor and that licheninases arose by loop addition, the apparent lack of direct PtEG16 homologs in the “lower” organisms, bacteria in particular, suggests otherwise.

The lack of this loop opened the active-site cleft and expanded the potential substrate range of PtEG16 homologs (Fig. 6). The observation of PtEG16 activity on both mixed linkage β-glucans and xyloglucan is particularly noteworthy, because these are the dominant matrix polysaccharides in grasy monocots and dicots, respectively. PtEG16 homologs are present in early plant lineages (Fig. 2B), and this dual activity would have thus poised them for subsequent functional adaptation in emerging species (6, 81) that favored either type of polysaccharide as a wall component. Single PtEG16 homologs are indeed found in many monocots and dicots (Fig. 2B). Interestingly, PtEG16-like genes, in particular in early diverging plants and monocots, have simple structures that lack introns, which may be further evidence of their bacterial origins (supplemental Table S2) (82).

Subsequently in the proposed evolution (Fig. 6), an early PtEG16 homolog is suggested to have acquired the large XTH C-terminal extension (Pfam XET_C, PF06955; Fig. 1), which is unique among GH16 members. This extension is found in all true XTH gene products from all major phylogenetic clades (Group I/II, III-A, and III-B; Fig. 2B), including all biochemically characterized XETs and XEHs (13). This suggests that acquisition of the C-terminal extension predated the massive expansion of XTH genes in mosses and later-diverging plants and monocots (Fig. 2B) (e.g. see Refs. 8, 9, and 82). The genetic mechanism of this acquisition is currently unknown.

Although systematic studies are lacking for most XTH gene products (13), it appears that at least some have lost much or all of their putative ancestors’ ability to act upon mixed linkage β-glucans. Specifically, barley HvXET5 acts on barley mixed linkage β-glucan at ≤0.2% of its rate on xyloglucan, whereas nasturtium TmNXG1 (an XEH associated with seed storage polysaccharide hydrolysis) has no detectable activity on Icelandic moss lichenan or barley β-glucan (11). There is no a priori reason why XETs or XEHs should exclude unbranched β-glucans from their active-site clefts based on steric considerations. However, specific enzyme-substrate interactions have apparently been optimized to favor xyloglucan binding in these XTH gene products (11, 45, 72, 73, 75).

Finally, we have previously provided evidence that the comparatively small clade of Group III-A XTH gene products (Fig. 2A), which comprises predominant XEHs, arose from the larger body of XETs via the specific acquisition of two small loop insertions altering the active-site cleft topology (11, 13) (Fig. 6; cf. Fig. 1). Combined phylogenetic, biochemical, and tertiary structural analyses indicated that a 5-amino acid loop (YNIIIG; Fig. 1) is a primary factor contributing to the predominant hydrolytic (XEH) capacity of nasturtium TmNXG1. Deletion of this loop significantly increased the transglycosylation/hydrolysis ratio of the enzyme, making it structurally and functionally more XET-like (11). This loop insertion is likewise found in two homologous A. thaliana XEHs of Group III-A (12).

In summary, the evolutionary scheme proposed here provides a new insight into the origins of the XTH gene products and a framework for the further exploration of protein structure-function relationships in plant GH16 members. Many questions remain outstanding, especially for PtEG16 homologs. These include the following.

To What Extent Can the Endoglucanase Designation Be Reliably Extended to Other PtEG16 Homologs?—We cautiously suggest that all members of the closely related dicot and monocot EG16-like clades may bear the same general activity profile as PtEG16 (Fig. 2A). However, the lack of demonstrable XET, XEH, or licheninase activity in one of three P. patens homologs (XTH32 (82)) suggests that careful functional analysis of early-diverging plant (moss) clades may especially be warranted.

What Is the in Vivo Role(s) of PtEG16 Homologs across Plant Lineages?—Although a comprehensive transcriptomic analysis has not been performed for any species, the gene encoding PtEG16 (previously known as XTH8) appears to be hormonally regulated (50). Likewise, a PpXTH32 promotor-glucuronidase fusion indicates tissue-specific expression in P. patens. The presence of single homologs in most species should facilitate reverse genetics analyses. Moreover, the lack of multiplication of PtEG16 homologs in plant genomes on a par with XTH genes may indicate either that they are functionally unique and under tight control or that they are simply molecular “transitional fossils,” vestigial sequences with no significant function. The observation that PtEG16 homologs are missing in many dicots might suggest the latter (Fig. 2A), although their functional importance may be inversely correlated with plant evolution. The lack of obvious signal peptides in PtEG16 homologs and a high abundance of Cys residues in some proteins (e.g. PtEG16), which may suggest intracellular localization, is particularly puzzling and relevant to this question.

What Is the Distribution and Function of PtEG16 and XTH Homologs in the Genomes of the Charophycean Green Algae?—Some members of this diverse group are the most primitive organisms known to have plantlike cell walls, including hemi-cellulosic polysaccharides like β(1→3)-glucans and xyloglucans (81). In notable contrast, the chlorophycean green algae do not possess xyloglucan (81) and, correspondingly, do not have any XTH-like genes (13) (Fig. 2B). We anticipate that forthcoming genomes of charophytes will illuminate the early origins of PtEG16 and XTH genes and their evolutionary relatedness.

In This Context, What Protein Structural Features Are Responsible for Substrate Specificity and Biochemical Function of PtEG16-like Proteins?—Our present homology models only allow reliable analysis of gross structural features on the level of overall fold and active site topology (Fig. 6). High resolution structural enzymology on par with that for other GH16 enzymes will be required to illuminate the details of substrate recognition and catalysis, which underpin biochemical and physiological function.

CONCLUSION

In harness, phylogenetic, biochemical, and protein structural analysis have revealed a unique clade of plant proteins in glyco-
side hydrolase family 16, whose members may represent a key step in the evolution of the widespread and diverse XTH gene family in plants. As such, the revelation of this clade will help refine future bioinformatics analyses of plant genomes. Moreover, further structural and functional analysis will undoubtedly bring new insight into the roles of individual clade members in the context of plant physiology.

REFERENCES

1. Carpita, N., and McCann, M. (2000) The cell wall. in *Biochemistry and Molecular Biology of Plants* (Buchanan, B., Gruissem, W., and Jones, R., eds) pp. 52–108, John Wiley & Sons, Inc., Somerset, NJ
2. Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., and Foust, T. D. (2007) Biomass recalcitrance. Plant enzymes and plants for biofuels production. *Science* **315**, 804–807
3. Ragauskas, A. J., Williams, C. K., Davison, B. H., Britvosék, G., Cairney, J., Eckert, C. A., Frederick, W. J., Jr., Hallett, J. P., Leak, D. J., Liotta, C. L., Mielzen, J. R., Murphy, R., Templer, R., and Tschapilinski, T. (2006) The path forward for biofuels and biomaterials. *Science* **311**, 484–489
4. Cosgrove, D. J. (2005) Growth of the plant cell wall. *Nat. Rev. Mol. Cell Biol.* **6**, 850–861
5. Vogel, J. (2008) Unique aspects of the grass cell wall. *Curr. Opin. Plant Biol.* **11**, 301–307
6. Fangel, J. U., Ulvskov, P., Knox, J. P., Mikkelsen, M. D., Harholt, J., Popper, Z. A., and Willats, W. G. (2012) Cell wall evolution and diversity. *Front. Plant Sci.* **3**, 152
7. Henriissat, B., Coutinho, P. M., and Davies, G. J. (2001) A census of carbohydrate-active enzymes in the genome of *Arabidopsis thaliana*. *Plant Mol. Biol.* **47**, 55–72
8. Geisler-Lee, J., Geisler, M., Coutinho, P. M., Segerman, B., Nishikubo, N., Takahashi, I., Aspeborg, H., Djeri, S., Master, E., Andersson-Gunerás, S., Sundberg, B., Karpinski, S., Teeri, T. T., Kleczkowski, L. A., Henriissat, B., and Mellerowicz, E. J. (2006) Poplar carbohydrate-active enzymes. Gene identification and expression analyses. *Plant Physiol.* **140**, 946–962
9. Yokoyama, R., Rose, J. K., and Nishitani, K. (2004) A surprising diversity and abundance of xyloglucan endotransglucosylase/hydrolases in rice. Classification and expression analysis. *Plant Physiol.* **134**, 1088–1099
10. Popper, Z. A., Michel, G., Hervé, C., Domozycz, D. S., Willats, W. G., Tuohy, M. G., Kloareg, B., and Stengel, D. B. (2011) Evolution and diversity of plant cell walls. From algae to flowering plants. *Annu. Rev. Plant Biol.* **62**, 567–590
11. Baumann, M. J., Eklöf, J. M., Michel, G., Kallas, A. M., Teeri, T. T., Cziejk, M., and Brunner, H., 3rd (2007) Structural evidence for the evolution of xyloglucanase activity from xyloglucan endotransglycosylases. Biological implications for cell wall metabolism. *Plant Cell* **19**, 1947–1963
12. Kaewthai, N., Gendre, D., Eklöf, J. M., Ibatullin, F. M., Ezcurra, I., Bhalerao, R. P., and Brunner, H. (2013) Group III-A XTH genes of *Populus trichocarpa* genome. An expert resource for glycomics. *Nucleic Acids Res.* **41**, 95–98
13. Meissner, A., Schulte-Herbruggen, T., Briand, J., and Sørensen, O. W. (1998) Sensitivity enhancement in the TROSY in triple-resonance experiments. New perspectives for sequential two-dimensional selection of multiplet components with long transverse relaxation times. *Mol. Phys.* **95**, 1137–1142
14. Weigt, J. (1998) Single scan, sensitivity- and gradient-enhanced TROSY for multidimensional NMR experiments. *J. Am. Chem. Soc.* **120**, 10778–10779
15. Rance, M., Loria, J. P., and Palmer, A. G. (1999) Sensitivity improvement of transverse relaxation-optimized spectroscopy. *J. Magn. Reson.* **136**, 92–101
16. Zhu, G., Gong, X. M., and Sze, K. H. (1999) Gradient and sensitivity enhancement of 2D TROSY with water flip-back, 3D NOESY-TROSY and TOCSY-TROSY experiments. *J. Biomol. NMR* **13**, 77–81
17. Eletsky, A., Kienöfer, A., and Persvik, K. (2001) TROSY NMR with partially deuterated proteins. *J. Biomol. NMR* **20**, 177–180
18. Salzmann, M., Wider, G., Persvik, K., Senn, H., and Wüthrich, K. (1999) TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins. *J. Am. Chem. Soc.* **121**, 844–848
19. Salzmann, M., Persvik, K., Wider, G., Senn, H., and Wüthrich, K. (1998) TROSY in triple-resonance experiments. New perspectives for sequential NMR assignment of large proteins. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13585–13590
20. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe. A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293
21. Godard, T., and Kneller, D. G. (2008) SPARKY version 3.115, University of California, San Francisco
22. Shen, Y., and Bax, A. (2012) Identification of helix capping and β-turn motifs from NMR chemical shifts. *J. Biomol. NMR* **52**, 211–232
42. Shen, Y., Delaglio, F., Cornilusc, G., and Bax, A. (2009) TALOS+. A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223

43. Camilloni, C., De Simone, A., Vranken, W. F., and Vendruscolo, M. (2012) Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry* **51**, 2224–2231

44. Campbell, P., and Braam, J. (1999) In vitro activities of four xyloglucan endotransglycosylases from *Arabidopsis*. *Plant J.* **18**, 371–382

45. Johansson, P., Bruner, H., 3rd, Baumann, M. J., Kallas, A. M., Henriksson, H., Denman, S. S., de Teer, T. T., and Jones, T. A. (2004) Crystal structures of a poplar xyloglucan endotransglycosylase reveal details of transglycosylation acceptor binding. *Plant Cell* **16**, 874–886

46. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C., and Eddy, S. R. (2004) The Pfam protein families database. *Nucleic Acids Res.* **32**, D138–D141

47. Hahn, M., Pons, J., Planas, A., Querol, E., and Heinemann, U. (1995) Crystallization of *Bacillus* licheniformis 1,3–1,4-β-D-glucanase at 1.8 Å resolution. *FEBS Lett.* **374**, 221–224

48. Gaiser, O. J., Piotukh, K., Ponnuswamy, M. N., Borriss, R., and Heinemann, U. (2006) Structural basis for the substrate specificity of a *Bacillus* 1,3–1,4-β-glucanase. *J. Mol. Biol.* **357**, 1211–1225

49. Del Bem, L. F., and Vincentz, M. G. (2010) Evolution of xyloglucan-related genes in green plants. *BMCE Biol.* **10**, 341

50. York, W. S., Impallomeni, G., Hisamatsu, M., Albersheim, P., and Darvill, A. G. (1995) 11 newly characterized xyloglucan oligoglycosyl alditols. The structure of plant-cell walls. *Carbohydr. Res.* **285–301

51. Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Kallas, A. M., Teeri, T. T., Brumer, H., and Agren, H. (2011) Phytozome. A comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186

52. Mark, P., Zhang, Q., Czjzek, M., Brumer, H., and Michel, G. (2010) Transfer of carbohydrate-active enzymes from marine *Bacillus* to Japanese gut microbiota. *Mol. Biol. Evol.* **27**, 2820–2830

53. Chen, J. H., Tsai, L. C., Huang, H. C., and Shyu, L. F. (2010) Structural and catalytic roles of amino acid residues located at substrate-binding pocket in *Fibrobacter succinogenes* 1,3–1,4-β-D-glucanase. *Proteins* **78**, 2802–2830

54. Mark, P., Baumann, M. J., Eklöf, J. M., Guillot, F., Michel, G., Kallas, A. M., Teeri, T. T., Brumer, H., and Czjzek, M. (2009) Analysis of natturium TmNXG1 complexes by crystallography and molecular dynamics provides detailed insight into substrate recognition by family GH16 xyloglucan endotransglycosylases and endo-hydrolases. *Proteins* **75**, 820–836

55. Mark, P., Zhang, Q., Czjzek, M., Brumer, H., and Agren, H. (2011) Molecular dynamics simulations of a branched tetradecasaccharide substrate in the active site of a *Populus* xyloglucan endo-transglycosylase with a library of xyloglucan-oligosaccharides. *J. Biol. Chem.* **286**, 2106–2111

56. Fry, S. C., York, W. S., Albersheim, P., Darvill, A., Hayashi, T., Joseleau, J. P., Kato, Y., Loncarevic, V., McNeil, M., Mort, A. J., Reid, J. S., Seitz, H. U., Selvendran, R. R., Voragen, A. G., and White, A. R. (1993) An unambiguous nomenclature for xyloglukan-derived oligosaccharides. *Physiol. Plant.* **89**, 1–3

57. Mark, P., Zhang, Q., Czjzek, M., Brumer, H., and Agren, H. (2011) Molecular dynamics simulations of a branched tetradecasaccharide substrate in the active site of a *Populus* xyloglucan endo-transglycosylase with a library of xyloglucan-oligosaccharides. *J. Biol. Chem.* **286**, 2106–2111

58. Fry, S. C., York, W. S., Albersheim, P., Darvill, A., Hayashi, T., Joseleau, J. P., Kato, Y., Loncarevic, V., McNeil, M., Mort, A. J., Reid, J. S., Seitz, H. U., Selvendran, R. R., Voragen, A. G., and White, A. R. (1993) An unambiguous nomenclature for xyloglukan-derived oligosaccharides. *Physiol. Plant.* **89**, 1–3

59. York, W. S., Harvey, L. K., Guillen, R., Albersheim, P., and Darvill, A. G. (1993) The structure of plant-cell walls. 36. Structural-analysis of tamarind seed xyloglukan oligosaccharides using β-galactosidase digestion and spectroscopic methods. *Carbohydr. Res.* **248**, 285–301

60. York, W. S., Impallomeni, G., Hisamatsu, M., Albersheim, P., and Darvill, A. G. (1995) 11 newly characterized xyloglukan oligoglycosyl alditols. The structure of plant-cell walls. *Carbohydr. Res.* **267**, 79–104

61. Hrmova, M., Farkas, V., Lahnstein, J., and Fincher, G. B. (2007) A barley xyloglucan xyloglucosyl transferase covalently links xyloglucan, cellulosic substrates, and (1,3;1,4)-β-D-glucans. *J. Biol. Chem.* **282**, 12951–12962

62. Fernandez-Fuentes, N., Madrid-Aliste, C. J., Rai, B. K., Fajardo, J. E., and Fiser, A. (2007) M4T. A comparative protein structure modeling server. *Nucleic Acids Res.* **35**, W363–W368

63. Hehemann, J. H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., and Michel, G. (2010) Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**, 908–912

64. Hehemann, J. H., Correc, G., Thomas, F., Bernard, T., Barbeyron, T., Jam, M., Helbert, W., Michel, G., and Czjzek, M. (2012) Biochemical and Struc-
Characterization of the complex agarolytic enzyme system from the marine bacterium Zobellia galactanivorans. J. Biol. Chem. 287, 30571–30584
81. Domozych, D. S., Ciancia, M., Fangel, J. U., Mikkelsen, M. D., Ulvskov, P., and Willats, W. G. (2012) The cell walls of green algae. A journey through evolution and diversity. Front. Plant Sci. 3, 82
82. Yokoyama, R., Uwagaki, Y., Sasaki, H., Harada, T., Hiwatashi, Y., Hasebe, M., and Nishitani, K. (2010) Biological implications of the occurrence of 32 members of the XTH (xyloglucan endotransglucosylase/hydrolase) family of proteins in the bryophyte Physcomitrella patens. Plant J. 64, 645–656
83. Gouet, P., Robert, X., and Courcelle, E. (2003) ESPript/ENDscript. Extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res. 31, 3320–3323