BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

*Biochemistry* **2010,** *49,* 3305–3316 3305
DOI: 10.1021/bi100009p

# Stimulation of Lignocellulosic Biomass Hydrolysis by Proteins of Glycoside Hydrolase Family 61: Structure and Function of a Large, Enigmatic Family[†]

Paul V. Harris,*[,‡,∥] Ditte Welner,[§,∥] K. C. McFarland,[‡] Edward Re,[‡] Jens-Christian Navarro Poulsen,[§] Kimberly Brown,[‡] Rune Salbo,[§] Hanshu Ding,[‡] Elena Vlasenko,[‡] Sandy Merino,[‡] Feng Xu,[‡] Joel Cherry,[‡] Sine Larsen,[§] and Leila Lo Leggio[§]

[‡]*Novozymes Inc., 1445 Drew Avenue, Davis, California 95618, and* [§]*Biophysical Chemistry Group, Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark* [∥]*These authors contributed equally to this work*

*Received January 5, 2010; Revised Manuscript Received March 12, 2010*

ABSTRACT: Currently, the relatively high cost of enzymes such as glycoside hydrolases that catalyze cellulose hydrolysis represents a barrier to commercialization of a biorefinery capable of producing renewable transportable fuels such as ethanol from abundant lignocellulosic biomass. Among the many families of glycoside hydrolases that catalyze cellulose and hemicellulose hydrolysis, few are more enigmatic than family 61 (GH61), originally classified based on measurement of very weak endo-1,4-$\beta$-D-glucanase activity in one family member. Here we show that certain GH61 proteins lack measurable hydrolytic activity by themselves but in the presence of various divalent metal ions can significantly reduce the total protein loading required to hydrolyze lignocellulosic biomass. We also solved the structure of one highly active GH61 protein and find that it is devoid of conserved, closely juxtaposed acidic side chains that could serve as general proton donor and nucleophile/base in a canonical hydrolytic reaction, and we conclude that the GH61 proteins are unlikely to be glycoside hydrolases. Structure-based mutagenesis shows the importance of several conserved residues for GH61 function. By incorporating the gene for one GH61 protein into a commercial *Trichoderma reesei* strain producing high levels of cellulolytic enzymes, we are able to reduce by 2-fold the total protein loading (and hence the cost) required to hydrolyze lignocellulosic biomass.

Developing a biomass-based, sustainable energy industry is of strategic importance (*1*−*3*). Although fuel ethanol production from sugar- and starch-rich biomaterials is an economically viable industry, lignocellulosic biomass (including agricultural byproducts, forestry residues, and woody/herbaceous crops) has a much larger potential as a renewable energy source in the future (*4*, *5*). The structural rigidity and stability of lignocellulose account for its evolution as the primary constituent of plant cell walls. These same properties make lignocellulose a difficult substrate to degrade into soluble small sugars that are a substrate for the envisioned "biorefinery" producing fuel ethanol and higher value coproducts (*6*, *7*). Currently, thermomechanical and/or chemical pretreatment followed by enzymatic hydrolysis is considered to have the greatest potential as an economical process for producing soluble sugars from lignocellulosic biomass (reviewed in ref *8*). However, technoeconomic analysis suggests that the cost of the enzymes must be further reduced in order for lignocellulose-derived fuel ethanol to be competitive with either gasoline or starch-derived fuel ethanol (*9*, *10*).

Degradation of plant cell walls in nature typically occurs through the action of a complex community of microbes secreting an array of enzymes, each enzyme acting on particular constituents of the cell wall (*11*). One of the best known and most commercially utilized cellulolytic systems is that of *Trichoderma reesei* (teleomorph *Hypocrea jecorina*) (*12*, *13*), which secretes two exocellobiohydrolases (Cel7A and Cel6A) that degrade cellulose chains processively from the ends, releasing cellobiose, several endoglucanases that cleave cellulose chains internally, and one or more $\beta$-glucosidases that hydrolyze cellobiose to glucose. Also secreted are multiple hemicellulases and other minor proteins such as carbohydrate esterases (*14*−*17*). These enzymes and their homologues from other organisms have been the object of considerable research, and substantial progress has been made in improving their characteristics with respect to reactivity, stability, and synergy (*11*, *18*−*23*). In contrast, only a limited effort has been made in studying accessory proteins that are coregulated and coexpressed by microbes during growth on cellulosic substrates. Recently, we carried out a multidisciplinary study aimed at significantly reducing the enzyme dosage and cost for biomass-based ethanol (*22*, *24*). By comparing and combining the extracellular cellulolytic systems of *T. reesei* and a cellulose-degrading thermophilic ascomycete *Thielavia terrestris*, we discovered three *T. terrestris* proteins that by themselves were not active in hydrolyzing either cellulosic biomass or typical glycosidase substrates, yet were capable of significantly enhancing the activity of cellulases on pretreated corn stover (PCS).[1] Subsequent

[†]Supported in part by Subcontract No. ZCO-1-30017-02 with The National Renewable Energy Laboratory under Prime Contract No. DE-AC36-99GO10337 with the U.S. Department of Energy and by The Danish Natural Science Research Council. The Dansync and ARI programs provided travel support to the MAX-lab synchrotron.

*To whom correspondence should be addressed. Phone: (530) 757-4703. Fax: (530) 758-0317. E-mail: paha@novozymes.com.

[1]Abbreviations: GH61, glycoside hydrolase family 61; PCS, pretreated corn stover; NREL, National Renewable Energy Laboratory; TrGH61B, *Trichoderma reesei* GH61B; PASC, phosphoric acid-swollen cellulose; SEM, standard error of the mean; MIR, multiple isomorphous replacement.


Published on Web 03/15/2010

Table 1: Data Collection (100 K) and Other Statistics for the GH61E Structures[a]

| | data set | | | |
|---|---|---|---|---|
| | native | $K_2Au(CN)_2$ | PHMBS | $ZnSO_4$ |
| data collection | | | | |
| space group | *F*23 | *F*23 | *F*23 | *F*23 |
| cell dimensions | | | | |
| *a*, *b*, *c* (Å) | 221.461 | 221.044 | 221.200 | 220.918 |
| α, *β*, *γ* (deg) | 90.0 | 90.0 | 90.0 | 90.0 |
| resolution (Å) | 15−1.9 (1.97−1.90) | 15−2.3 (2.38−2.30) | 15−2.3 (2.38−2.30) | 20−2.25 (2.33−2.25) |
| $R_{merge}$ | 0.094 (0.483) | 0.099 (0.318) | 0.170 (0.564) | 0.109 (0.333) |
| $I/\sigma I$ | 31.0 (7.3) | 26.6 (8.6) | 22.3 (6.2) | 9.9 (2.1) |
| completeness (%) | 99.6 (100.0) | 99.9 (99.2) | 98.0 (100.0) | 100 (99.9) |
| redundancy | 15.9 | 12.7 | 17.5 | 6.2 |
| refinement | | | | |
| resolution (Å) | 15−1.9 (1.97−1.90) | | | 20−2.25 (2.33−2.25) |
| no. of unique reflections | 69834 | | | 42338 |
| $R_{work}/R_{free}$ | 18.4/20.4 | | | 21.7/25.1 |
| no. of atoms | | | | |
| protein | 6474 | | | 6393 |
| divalent metal ion | 4 | | | 6 |
| water | 851 | | | 696 |
| NAG/GOL/$SO_4$ | 123 | | | 123 |
| *B*-factors | | | | |
| protein | 15.5 | | | 18.8 |
| divalent metal ion | 14.4 | | | 34.8 |
| water | 24.3 | | | 28.3 |
| NAG/GOL/$SO_4$ | 43.3 | | | 53.6 |
| rms deviations | | | | |
| bond lengths (Å) | 0.009 | | | 0.010 |
| bond angles (deg) | 1.1 | | | 1.5 |

[a]Values in parentheses are for the highest resolution shell.

investigations demonstrated that several but not all members of this protein family, currently classified as glycoside hydrolase family 61 (GH61), have this capacity. The biochemical function of this family has never been satisfactorily elucidated, and in this study we demonstrate for the first time that GH61 proteins can dramatically influence the activity of cellulases on lignocellulose, are unlikely to be canonical glycoside hydrolases, and require divalent metal ions for activity.

## MATERIALS AND METHODS

*Culture and Fermentation of Fungal Species*. *T. reesei* strain SMA135 was derived from Novozymes' commercial Celluclast production strain by transformation with a *β*-glucosidase gene from *Aspergillus oryzae* (*25*). Strain SaMe MF-268 was similarly generated by transformation with a different construct of the same *β*-glucosidase gene from *A. oryzae* (*26*) along with the *gh61a* gene from *Thermoascus aurantiacus* (*27*). High-level large-scale production and formulation of cellulase product from SMA135 and SaMe MF-268 were performed by Novozymes North America.

*Cloning of GH61-Encoding Genes and Recombinant Protein Expression*. Cloning and high-level recombinant expression of *T. aurantiacus* GH61A (Genbank Accession ABW56451) and *T. terrestris* GH61B (Accession ACE10231) and GH61E (Accession ACE10234) were performed as previously described (*27*, *28*).

*Assay for Cellulose and Lignocellulose Hydrolysis*. Corn stover substrate was pretreated by the U.S. National Renewable Energy Laboratory (NREL) to generate pretreated corn stover (PCS) essentially as described by Schell et al. (*29*) using total solids of 28%, sulfuric acid concentration of 0.05 g of acid/g of dry biomass, and a residence time of 1 min at 190 °C. This material was further processed by extensive washing with reverse osmosis water until the pH was above 4.0 and soluble sugars were less than 0.6 mg/mL. Washed PCS was ground to a uniform size, sterilized by autoclaving, and stored at 4 °C. The starting material has a wide size distribution based on sieve filtration, with 28% of the particles less than 75 *μ*m in at least one dimension, 24% between 75 and 150 *μ*m, 26% between 150 and 425 *μ*m, and the remainder larger. Approximately 85% of the final substrate passes readily through a 200-mesh screen and is therefore less than about 75 *μ*m in its smallest dimension. PCS used for assaying stimulation of GH61 by metal ions was additionally incubated with 10 mM EDTA for 48 h at 50 °C, washed extensively with Milli-Q (Millipore) water until the conductivity was less than 10 *μ*S, and further washed extensively with 5 mM NaCl in Milli-Q water. According to NREL, the washed PCS solids' composition is 57.5% glucan, 7.0% xylan, 0.9% galactan, 0.7% arabinan, and 27.2% lignin. The standard hydrolysis assay was carried out in a volume of 1 mL of 50 mM sodium acetate, pH 5.0, 1 mM $MnSO_4$, and 5.0% (w/v) washed and ground PCS or Avicel PH 101 (FMC). Incubation was typically at 50 °C with one daily inversion for mixing. Samples were taken after 3−7 days, filtered (Millipore Multiscreen MAHV N45 50), and diluted 1:1 with 5 mM $H_2SO_4$. Sugars were fractionated by HPLC on an Aminex HPX 87H column eluted with 5 mM $H_2SO_4$. Detection was by refractive index and quantitation by integration of the signal compared to purified sugar standards.

*Phylogenetic Analysis of Family GH61*. GH61 proteins detected by BLAST (*30*, *31*) searching of the Uniprot (*32*) and GENESEQ (Thomson Reuters Corp.) databases were aligned

with the "einsi" algorithm of MAFFT version 6.624 or later (*33*, *34*). Obvious duplicate sequences and clearly bad or incomplete gene models were discarded, and the alignment was truncated by removing the predicted signal peptide and any C-terminal linker and cellulose-binding modules. A phylogenetic tree was constructed using the maximum likelihood program PhyML version 3.0.1 (*35*, *36*). The initial tree was constructed with the BioNJ algorithm (*37*), and the tree topology was refined using the best of nearest neighbor interchange (*36*) and subtree pruning and regrafting (*38*) under a discrete gamma model (estimated) with four rate categories and variable sites fixed. The LG model of amino acid substitution was used (*39*).

*Deglycosylation and Purification of GH61E for Structure Determination*. Fermentation broth of *A. oryzae* expressing *T. terrestris* GH61E was concentrated and desalted into sodium citrate buffer, pH 5.5. For deglycosylation, EndoH (New England Bioloabs) was added to ~1500 units/mg, and the mixture was sterile filtered and incubated for 48 h at 37 °C. Deglycosylated protein was desalted into 20 mM Tris-HCl, pH 8.91, applied to a MonoQ column (GE Healthcare), and eluted with a gradient of the same buffer from 0 to 1 M NaCl. The peak fractions were buffer-exchanged to 2.0 mM each imidazole, diethanolamine, Tris, piperazine, and *N*-methylpiperazine plus 10% (v/v) glycerol, pH 9.5, and applied to a MonoP column (GE Healthcare) equilibrated with the same buffer. Protein was eluted with a gradient of the same buffer from pH 9.5 to pH 4.5. The protein eluted as a broad peak from pH 7.6 to pH 5.0.

*Crystallization and Data Collection*. Crystals were obtained at room temperature within 5−7 days using the vapor diffusion method in Linbro plates. The 1 mL reservoir contained 1.6 M $MgSO_4$ and 0.1 M MES, pH 6.5, while the hanging drop consisted of 2 μL of the same buffer mixed with 2 μL of GH61E protein (3.1 mg/mL). Two heavy atom derivatives were obtained by soaking crystals in 5 mM $K_2Au(CN)_2$ and 10 mM PHMBS for 73 and 85 days, respectively. All three data sets were collected at 100 K on a rotating Cu-anode diffractometer. A derivative with zinc ion was obtained by soaking a crystal in 1.8 M $ZnSO_4$ and 0.1 M sodium cacodylate, pH 6.5, for a few minutes. Data were then collected at beamline 911-3 at MAX-lab, Lund, Sweden, with a data collection temperature of 100 K at a wavelength of 1.1 Å to ensure significant anomalous signal for Zn. All data were processed with Denzo/Scalepack from the HKL 2000 software package (*40*) or a previous version of the program (1.97.9). Data collection statistics for the collected X-ray data sets are given in Table 1.

*Structure Determination and Refinement*. Phases were determined by the MIR method. Eight heavy atom sites were located in each derivative using SOLVE (*41*) and experimental phases calculated to 2.3 Å (FOM = 0.59). Density modification and automatic tracing of ~64% of the amino acids were performed with RESOLVE (FOM = 0.80). The remaining fragments were manually built in O (*42*) or COOT (*43*), and the model for the native structure was finally refined to 1.9 Å in CNS (*44*) and REFMAC (*45*) using NCS restraints. There are four molecules in the asymmetric unit. The largest coordinate difference between them is between molecules A and B with a root-mean-square deviation of 0.486 Å over 1568 atoms. An *N*-acetyl-D-glucosamine linked to Asn-51 is present in each molecule in the asymmetric unit as a residual from the glycosidase treatment. The $ZnSO_4$ structure was refined in CNS to 2.25 Å resolution using the native structure as a starting model. Refinement statistics for the final deposited structure models are given in Table 1. The percentage of residues in the Ramachandran most favored regions was 90.7% for the structure with bound $Mg^{2+}$ and 88.1% for the structure with bound $Zn^{2+}$. Ramachandran plots were calculated with PROCHECK (*46*). The coordinates of the structure models and the structure factors have been deposited at the PDB (www.pdb.org) and have been given accession codes 3EJA and 3EII, respectively, for the $Mg^{2+}$- and $Zn^{2+}$-bound structures.

*Site-Directed Mutagenesis of GH61E*. Several sites in the gene coding for GH61E were mutagenized using a QuikChange site-directed mutagenesis kit (Stratagene) according to the manufacturer's instructions. The specific amino acid changes targeted were H1N, H68A, E137A, H142A, Q151L, Q151N, Q151E, Y153F, and Y192A.

*Purification of GH61 Proteins for Measurement of Carbohydrase Activity*. Fermentation broth containing recombinantly expressed (*A. oryzae*) *T. terrestris* GH61B was desalted and concentrated in 20 mM Tris-HCl, pH 8.0. It was applied to a Q-Sepharose Big Beads column (GE Healthcare) in 20 mM Tris-HCl, pH 8.0, and eluted with a 0 to 1 M linear NaCl gradient. Fractions were pooled based on SDS−PAGE (major band at ~40 kDa) and then desalted into 20 mM Tris-HCl, pH 8.0. The desalted pool was then applied to a MonoQ column (GE Healthcare) in 20 mM Tris-HCl, pH 8.0, and eluted using a 0 to 100 mM NaCl gradient. Fractions were pooled based on SDS−PAGE and desalted and concentrated into 20 mM Tris-HCl, pH 8.0, with a final volume of 100 mL. Five milliliters was slowly mixed with 5 mL of 20 mM Tris-HCl, pH 7.5, and 3.4 M $(NH_4)_2SO_4$. This was loaded on a HiTrap Phenyl FF column (GE Healthcare) with 20 mM Tris-HCl, pH 7.5, and 1.7 M $(NH_4)_2SO_4$ and eluted using a linear gradient from 1.7 to 0 M $(NH_4)_2SO_4$ over 20 column volumes. Fractions eluting from 0.64 to 0.38 M were pooled, buffer-exchanged into 20 mM Tris-HCl, pH 7.5, and stored at −20 °C. The protein retained full activity in a cellulase enhancement assay.

Fermentation broth containing recombinantly expressed (*A. oryzae*) *T. terrestris* GH61E was concentrated and desalted by tangential flow filtration into 20 mM Tris-HCl, pH 8.5, and applied to a Q-Sepharose high-performance QHP-75 column (GE Healthcare) equilibrated with the same buffer. Protein was eluted with a 0 to 1.0 M NaCl linear gradient over 10 column volumes. GH61E protein eluted in two peaks at approximately 50 and 120 mM NaCl. The peak eluting at 50 mM was adjusted to 1.5 M $(NH_4)_2SO_4$ and applied to a phenyl-Superose HR16/10 column (GE Healthcare) equilibrated with 20 mM Tris-HCl, pH 8.0, and 1.5 M $(NH_4)_2SO_4$ and eluted with a 1.5 to 0 M linear $(NH_4)_2SO_4$ gradient over 20 column volumes at 1 mL/min. GH61E eluted in two peaks at approximately 0.6 and 0.4 M $(NH_4)_2SO_4$. Fractions eluting around 0.6 M $(NH_4)_2SO_4$ were pooled, desalted, and stored at −20 °C. The protein retained full activity in a cellulase enhancement assay.

*Assay for Carbohydrase and Transglycosidase Activity*. The activity of two partially purified GH61 proteins was tested on a variety of carbohydrate substrates in 0.2 mL of 50 mM sodium acetate, pH 5.0, 0.01% (v/v) Tween-20, 1 mM $MnSO_4$, 1.0 μg GH61 protein, and substrates at 4 mg/mL (except for PASC and xylan at 1.6 mg/mL). Incubation was for 72 h at 50 °C. The reactions were stopped by adding 50 μL of 0.1 M NaOH, and reducing sugars produced were analyzed by the PHBAH method (*47*), which was modified and adapted to a 96-well microplate format. In some cases reaction products of the hydrolysis were analyzed and quantified by HPLC analysis using a Dionex

ICS3000 CarboPac PA1 column. The column was equilibrated with 1.5% solution B (1.00 M NaOH), and reaction products were first eluted isocratically for 20 min at a flow rate of 1.0 mL/min followed by a 1 min gradient from 1.5% to 11.0% solution B and 0% to 5.0% solution C (0.50 M sodium acetate, 0.10 M NaOH). This was followed by a 29 min gradient up to 14.0% solution B and 10.0% solution C and a final 1 min gradient to 25.0% solution B and 10.0% solution C. Substrates tested were NREL-washed PCS, Avicel, CMC, sugar beet arabinan, barley $\beta$-glucan, amyloid xyloglucan, wheat arabinoxylan, mannan, carob galactomannan, potato galactan, potato amylose, potato amylopectin, PASC, pectic galactan, apple pectin, and crab shell chitin. All substrates were obtained from Megazyme except for PCS (NREL), Avicel (FMC), CMC (Hercules), PASC, amylose, amylopectin, pectin, and chitin (Sigma). PASC was prepared by suspending moistened Avicel (about 5 g) in ice-cold 85% phosphoric acid at 0.033 g/L, stirring on ice for 1 h, diluting with cold acetone to 0.0077 g/L, and vacuum-filtering through a glass fiber filter. The solids were rinsed three times with 100 mL of ice-cold acetone and twice with 500 mL of water. The final vacuum-filtered material was suspended in water at 0.01 g/mL and blended to homogeneity using an Ultra Turrax homogenizer and stored at 4 °C.

*T. aurantiacus* GH61A protein was assayed for transglycosidase activity by incubating for 3 days at 50 °C in a reaction mix containing 0.5% (w/v) glucose or cellobiose, 20 mM ammonium acetate, pH 5.0, 0.1 mM $CaCl_2$, and GH61A at 40 mg/mL. The reaction products were analyzed by mass spectrometry using a Waters QTof micro system at a flow rate of 20 $\mu$L/min in 2 mM ammonium aceate, pH 5.0. Ions were surveyed in positive mode from $m/z$ 50 to 1000, 2500 V capillary voltage, 20 V cone voltage, 300 °C desolvation temperature, 145 °C source temperature, 5 V collision energy, 0.9 s scan time, 100 L/h cone gas flow, and 500 L/h desolvation gas flow.

## RESULTS

The initial evidence for the enhancing activity of the GH61 family derived from PCS-hydrolysis experiments in which a commercial cellulase system of *T. reesei* (Novozymes Celluclast 1.5L) was mixed with broth from *T. terrestris* grown on cellulose as the sole carbon source. While the *T. terrestris* broth by itself was inferior to Celluclast in PCS hydrolysis at equal protein loading, mixing of Celluclast and *T. terrestris* fermentation broth in equal amounts allowed for a 2-fold reduction in total protein loading while still retaining the same degree of hydrolysis after 24 h of incubation (*24*). This suggested that *T. terrestris* was producing a protein or proteins that could act in synergy with the canonical *T. reesei* cellulases, either by supplementing a limiting activity and/or by providing entirely new functionality.

*Identification of GH61 Proteins As Cellulase-Enhancing Factors*. In order to identify the protein or proteins secreted by *T. terrestris* that could enhance cellulase activity, we carried out an extensive analysis of this species using protein fractionation, proteomics, random sequencing of cDNA clones, and random cloning of known glycoside hydrolases by degenerate PCR. One- and two-dimensional polyacrylamide gel fractionation of fermentation broth proteins from cultures grown on cellulose and subsequent *de novo* sequencing by tandem mass spectrometry revealed a secretome of moderate complexity consisting of the expected cellulose and hemicellulose degrading enzymes (Figure 1). Less expected was the presence of at least six

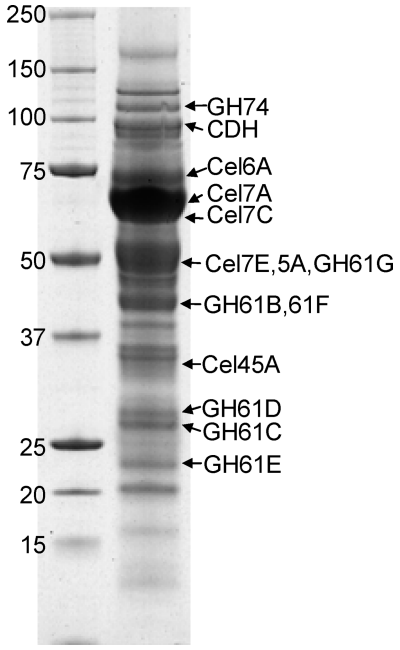


FIGURE 1: SDS−PAGE separation of the secreted proteome of *T. terrestris*. The organism was grown in shake flask culture with cellulose (Arbocel B800) as the only carbon-containing energy source. Proteins were identified by LC-MS/MS. Proteins are named according to their glycoside hydrolase family with the exception of cellobiose dehydrogenase (CDH).

representatives of the GH61 family, comprising in total about 10% of the total soluble protein under these fermentation conditions. We performed several independent chromatographic fractionations of the broth proteins and assayed individual fractions for the ability to enhance the activity of Celluclast 1.5L. Three different GH61 proteins, designated GH61B, GH61E, and GH61G, were identified as potentially stimulatory proteins. These results suggested that the GH61 family proteins might be the major factors responsible for the enhancement of Celluclast activity by crude *T. terrestris* fermentation broth.

*Cloning, Expression, and Preliminary Characterization of GH61 Proteins*. To facilitate further investigation, we cloned the genes encoding GH61B, GH61G, GH61E (GenBank accession nos. ACE10231, ACE10235, and ACE10234), and many other GH61 family members from *T. terrestris* and several other cellulose-degrading fungi. These genes were cloned into an expression vector designed for high-level expression and transformed into *A. oryzae* for production of relatively pure protein. In some cases the protein was purified further to remove low levels of contaminating activities. The proteins were then systematically tested for their ability to enhance PCS hydrolysis by *T. reesei* cellulases. Several but not all of the recombinantly produced GH61 proteins showed an enhancing activity when added in the range of 5−20% of the total protein present. Data for two of the more potent GH61 proteins are shown in Figure 2a. The fermentation broths of the *A. oryzae* strains expressing these GH61 proteins showed only very low background levels of hydrolytic activity on PCS (expected from *A. oryzae* grown on monosaccharides), indicating that the enhancement was not likely to be the result of any substantial endo- or exo-$\beta$-1, 4-glucanase activity of the GH61 proteins. The gene encoding the GH61A protein from *T. aurantiacus* (GenBank accession no. ACS05720) was additionally cloned into *T. reesei* in order to assess our ability to create a new strain of *T. reesei* capable of

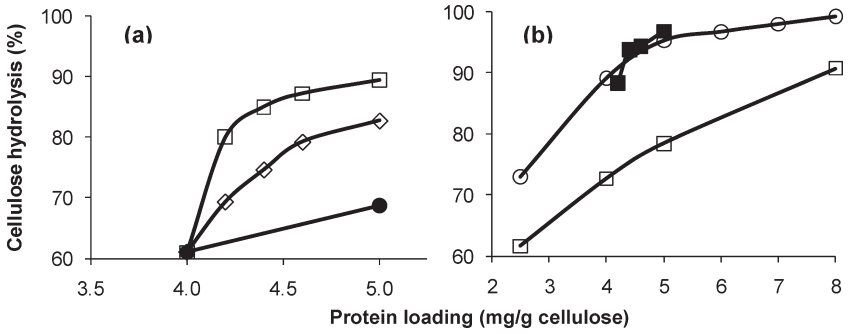


FIGURE 2: Effect of GH61 on cellulose hydrolysis of PCS by *T. reesei* cellulases. PCS was hydrolyzed for 168 h as described in Materials and Methods with $MnSO_4$ present at 1 mM. In (a), hydrolysis was performed with cellulases derived from strain SMA-135 either alone at 4 or 5 mg of protein/g of cellulose (●) or with addition to the cellulases (at 4 mg/g of cellulose) of increasing amounts of recombinant GH61E protein from *T. terrestris* (◇) or GH61A protein from *T. aurantiacus* (□). Neither GH61 protein alone produced measurable hydrolysis. In (b), cellulases were prepared from two *T. reesei* strains, one lacking heterologous expression of *T. aurantiacus* GH61A (□) and one expressing GH61A at approximately 7% of total protein (○). Also shown is the effect of adding increasing amounts of recombinantly expressed GH61A protein to the *T. reesei* cellulases (at 4 mg/g of cellulose) derived from the strain lacking heterologous expression (■).Values are the means of triplicate determinations. The SEM for each data point was smaller than the size of the symbol and is not graphed.

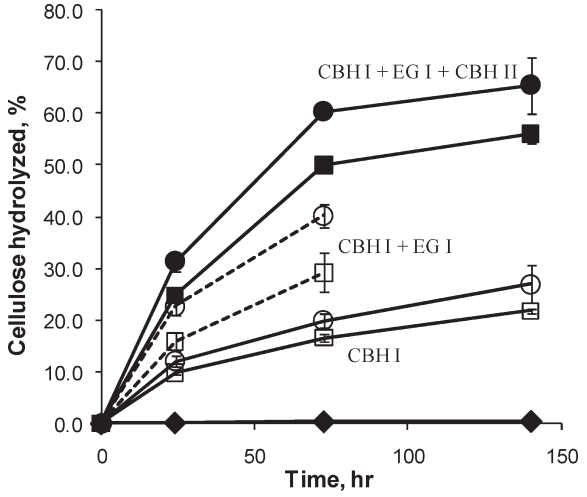


FIGURE 3: Stimulation of cellulases by *T. aurantiacus* GH61A. PCS at a loading of 4.7% (dry w/v) in 1 mL of 50 mM sodium acetate, pH 5.0, was hydrolyzed at 50 °C for the times indicated in the presence of individual *T. reesei* cellulases or cellulase mixtures, either with (circles) or without (squares) GH61A at 10% of the total protein loading of 2.5 mg/g of cellulose. The relative ratios of CBH I, CBH II, and EG I, when mixed, were 4.5:2.5:1.0. In each case purified recombinant $\beta$-glucosidase from *A. fumigatus* was also present at 10% of the total protein. Cellulose hydrolysis was measured by the PHBAH reducing sugar assay. GH61A protein alone (◆) produced no significant reducing sugar release. Values are means of triplicate determinations ± SEM.

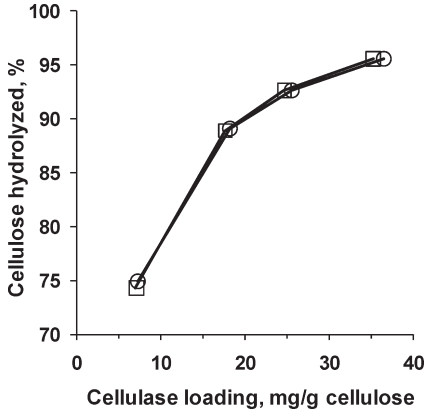


FIGURE 4: Effect of *T. aurantiacus* GH61A on hydrolysis of nearly pure cellulose by *T. reesei* cellulases. A cellulase preparation from strain SMA-135 was loaded at the indicated protein concentrations in either the absence (○) or presence (□) of GH61A protein at 7% of total protein loading. Hydrolyses were carried out in 1 mL of 50 mM sodium acetate, pH 5.0, 1 mM $MnSO_4$, and 2.3% (w/v) Avicel PH101 for 166 h at 50 °C. Conversion was measured by HPLC. Values are means of triplicate determinations. The SEM for each data point was smaller than the symbol size and is not graphed.

producing cellulases with enhanced hydrolytic activity. Heterologous GH61A expression in this strain did indeed create a cellulase mixture with markedly enhanced specific activity in the hydrolysis of PCS (Figure 2b). Compared to the strain lacking *T. aurantiacus* GH61A expression, the protein loading required to reach 91% cellulose conversion is reduced from 8.0 mg of protein/g of cellulose to 4.2 mg of protein/g of cellulose, a 1.9-fold reduction. The loading reduction is somewhat less at lower conversions but remains above 1.6-fold in the 70−80% conversion range.

*Stimulation of Individual Cellulases by GH61.* The stimulatory effect of GH61 on lignocellulose hydrolysis could result from an effect on specific enzymes in a complex cellulase mixture or might result from a more generic process that affects multiple enzymatic activities. To test this, we examined the effect of *T. aurantiacus* GH61A on the activities of the major components of the *T. reesei* cellulase mixture, CBH I (Cel7A), CBH II (Cel6A), and EG I (Cel7B). GH61A showed a small but significant stimulatory effect on the hydrolytic activity of CBH I (Figure 3). This stimulatory effect was greater with a mixture of CBH I and EG I or CBH I, CBH II, and EG I. We conclude from these data that the effect of GH61A is not predominantly stimulation of any specific enzymatic activity but rather an overall effect on the cellulase complex or substrate.

*Lack of GH61 Stimulation of Pure Cellulose Hydrolysis.* In our initial survey of GH61 proteins we used a variety of relatively pure cellulose substrates such as filter paper and phosphoric acid-swollen cellulose and failed to see any stimulatory effect of any GH61 protein on these substrates. This observed lack of effect on relatively pure cellulose was carefully examined using Avicel PH101 and one of the more active GH61 proteins known at the time, GH61A from *T. aurantiacus*. As observed in our initial survey, GH61A addition at 7% of total protein had little effect on the hydrolysis of this substrate (Figure 4), suggesting that the enhancement of cellulolytic

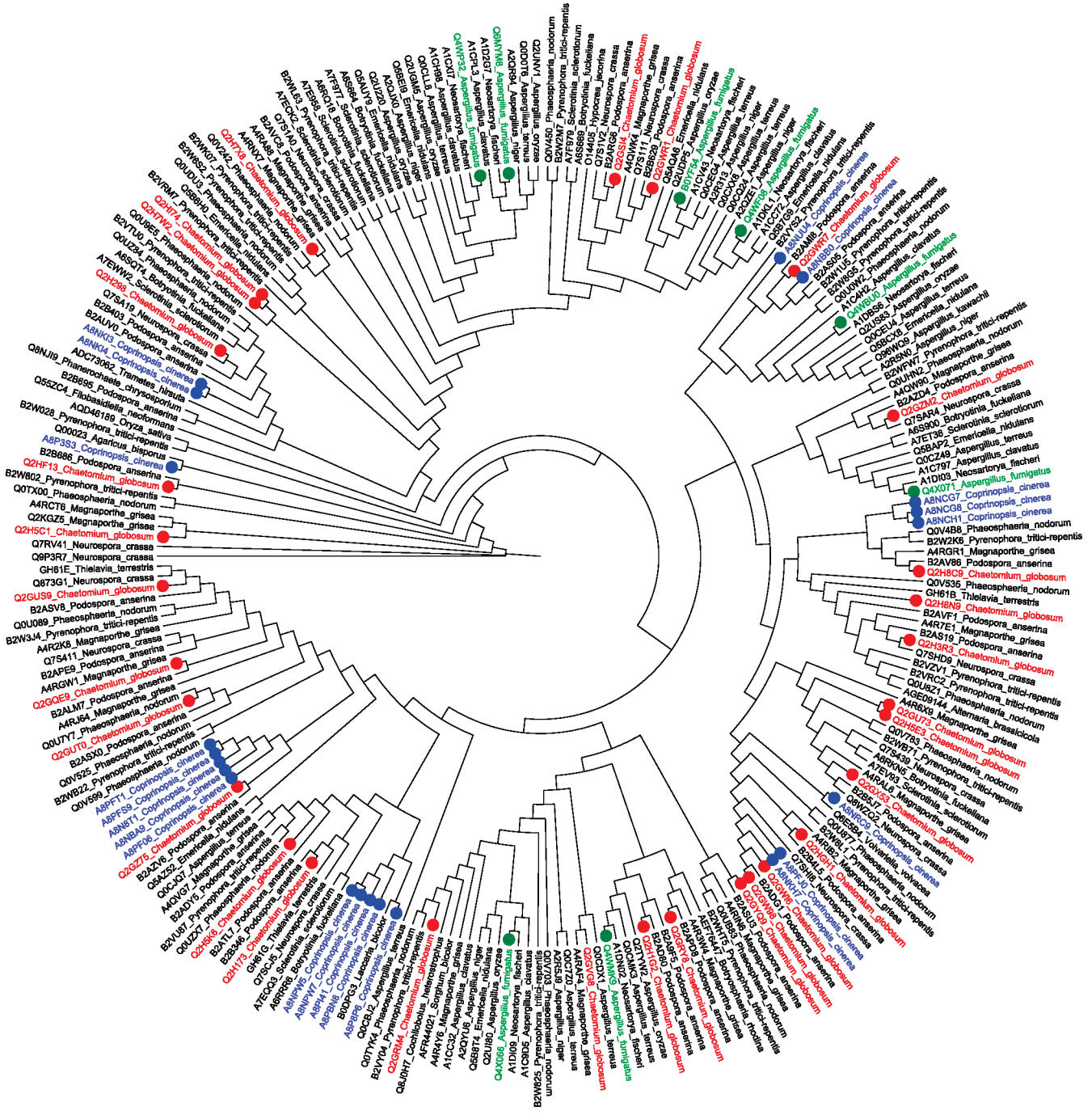

FIGURE 5: Cladogram of GH61 proteins available in the Uniprot and GeneSeqP databases. The accession number and identified species are shown for each entry. The maximum likelihood tree was constructed with PhyML as detailed in Materials and Methods. Sequences from *C. globosum*, *A. fumigatus*, and *C. cinerea* are highlighted in red, green, and blue, respectively. The cladogram was drawn by FigTree (*67*).

activity by GH61 might be limited to substrates containing other cell wall-derived material such as lignin or hemicellulose.

*Phylogenetics of the GH61 Family*. As of July 2009, the GenBank nonredundant (nr) protein database contained approximately 300 known or predicted proteins with clear sequence homology to the GH61 family based upon BLAST searches and conservation of certain key motifs, specifically an N-terminal histidine and the motifs Y-x-x-R-x-[EQ] and [EQ]-x-[FWY]-x-x-[CG] (in PROSITE format (*48*)). All of these are from filamentous fungi (a few plant cDNAs coding for GH61 proteins are probably fungal contaminants), and all but one of these fungi are likely to have the capacity to degrade plant cell wall material based on direct evidence and the presence of many known cellulase genes in their genomes. The only known exception is the presence of a single GH61-encoding gene in the sequenced genomes of *Cryptococcus neoformans* var. *neoformans* serotypes A and D. No GH61-encoding genes are recognizable either in the genome sequences of yeasts or in other noncellulolytic filamentous fungi such as *Rhizopus oryzae*, *Ustilago maydis*, or *Coccidioides immitis*. The GH61 gene family has been dramatically amplified in several fungal species for which complete or nearly complete genome sequence information is available, notably *Coprinopsis cinerea*, *Podospora anserina*, *Chaetomium globosum*, and *Staganospora nodorum*, all of which have more than 25 genes encoding GH61 proteins. A cladogram constructed from alignment of most of the GH61 proteins demonstrates that

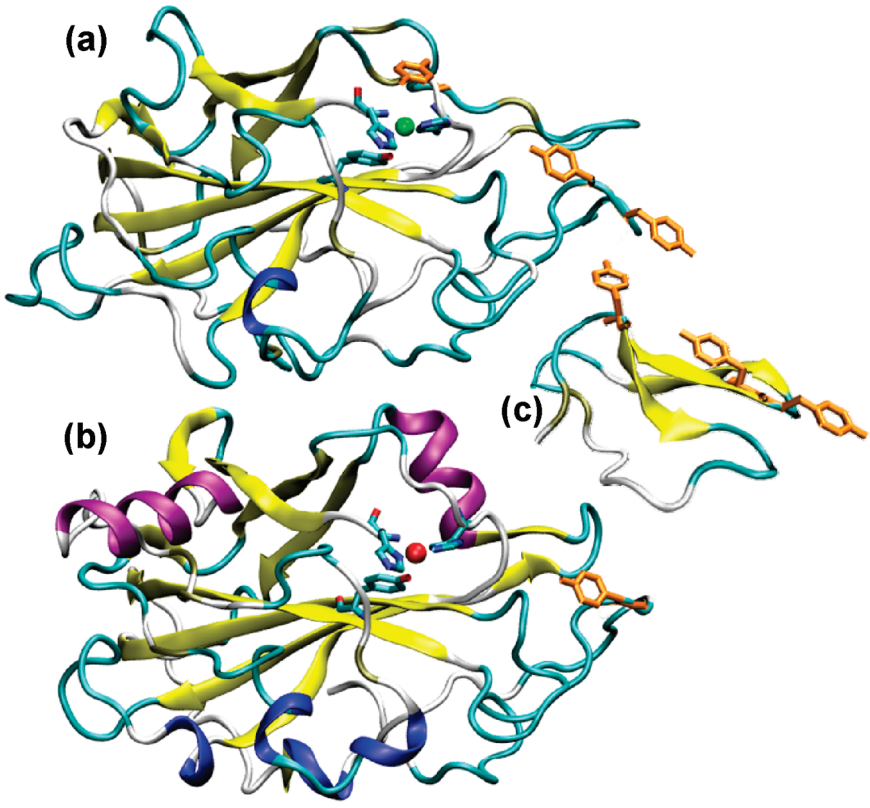


FIGURE 6: Cartoon representation of (a) GH61E, (b) TrGH61B (PDB ID 2VTC), and (c) family 1 cellulose-binding module (PDB ID 1CBH). $\beta$-Strands are in yellow, α-helices in purple, $3_{10}$-helices in blue, turns in cyan, and random coils in white, as determined by the STRIDE algorithm (*68*). Near the N-terminus of both structures is a bound metal ion, $Mg^{2+}$ (green sphere) in GH61E and $Ni^{2+}$ (red sphere) in TrGH61B. The metal ions are in close proximity to two histidines and a tyrosine conserved in GH61, which are shown in stick form. Also shown in stick form are three solvent-exposed surface tyrosines (orange) in GH61E (Tyr-67, -191, -192) that form a relatively flat planar surface potentially suitable for polysaccharide binding adjacent to the metal ion. For comparison, the similarly solvent-exposed planar tyrosines in a known cellulose-binding domain are also shown in (c). Only one of the three GH61E tyrosines (structural equivalent of Tyr-191) is conserved in TrGH61B. The longer α-helices in TrGH61B represent insertions in this sequence relative to GH61E. The structures were aligned, and the figure was prepared in VMD (*69*, *70*) and rendered with POV-ray (*71*).

sequences from two major fungal phyla, Ascomycota and Basidiomycota, are distributed throughout the tree (Figure 5). This indicates that much of the duplication and divergence of the genes encoding GH61 proteins is a fairly ancient evolutionary event, preceding the divergence of these two phyla, estimated at >600 million years ago (*49*). This diversity has been largely maintained and in some cases apparently amplified during this evolutionary span, suggesting a significant selective pressure to retain a heterogeneous collection of GH61-encoding genes in many cellulose-degrading fungal species. Further analysis shows that the amplification and divergence of GH61 genes have taken different paths in different species. For example, the basidiomycete *C. cinerea* has a large number of closely related paralogues, suggesting that much of the gene amplification in this species is a relatively recent evolutionary event. In contrast, in ascomycetes such as *C. globosum* and *Phaeosphaeria nodorum* the numerous GH61 proteins are much more dispersed throughout the tree, suggesting that these species have maintained their large collection of GH61-encoding genes over a long evolutionary span. The lesser number of GH61-encoding genes in *Aspergillus fumigatus* and other *Aspergillus* species is fairly well dispersed throughout the tree but obviously missing from certain major branches. This implies selective loss of GH61 ancestral genes during evolution of the Aspergilli.

Approximately 20% of the public GH61 sequences have obvious family 1 cellulose binding modules near the carboxy terminus. This is very likely an underestimate since many of the gene models are incomplete at the 3′ end. We find no correlation between the presence or absence of such a module and the efficacy of the GH61 protein in enhancing PCS hydrolysis. Some GH61 proteins have conserved carboxy-terminal extensions of unknown function. Again, we see no correlation between their presence and efficacy.

*Structure Determination and Hydrolytic Activity*. The GH61 proteins have been classified as glycoside hydrolases based upon reports of extremely weak endoglucanase activity observed in three family members (*50*−*53*). In an effort to clarify the function of GH61, we determined the three-dimensional structure of GH61E from *T. terrestris* by multiple isomorphous replacement (MIR) to a resolution of 1.9 Å. The structure is a compact single-domain $\beta$-sandwich consisting of two sheets in a variation of a fibronectin type III fold (Figure 6a). There is no evidence of large surface clefts, crevices, or holes that would point to a possible binding pocket for a soluble polysaccharide. Many of the most highly conserved side chain residues in the GH61 family are located in the core $\beta$-sandwich and participate in a large buried ionic network. Significantly, there is no evidence for clustering of conserved catalytic acidic residues that are present in almost all known glycoside hydrolases (*54*−*56*). The most conserved acidic residue, Glu-137, is part of the above-mentioned buried ionic network and forms an ion pair with the highly conserved and similarly solvent-inaccessible Arg-135. Mutagenesis of Glu-137 to Ala results in a complete lack of protein expression, suggesting that this salt bridge is structurally

Table 2: Activity of Two Partially Purified GH61 Proteins on Polysaccharide Substrates[a]

| substrate | GH61B (%) | GH61E (%) |
|---|---|---|
| PCS | 0.71 ± 0.26 | 0.28 ± 0.31 |
| Avicel | 0.13 ± 0.02 | 0.08 ± 0.01 |
| carboxymethylcellulose | 0.09 ± 0.02 | 0.08 ± 0.03 |
| arabinan | 0.40 ± 0.03 | 0.24 ± 0.01 |
| $\beta$-glucan | 0.88 ± 0.22 | 0.45 ± 0.13 |
| xyloglucan | 0.46 ± 0.11 | 0.38 ± 0.08 |
| arabinoxylan | 0.66 ± 0.09 | 0.52 ± 0.04 |
| mannan | 0.18 ± 0.05 | 0.17 ± 0.03 |
| galactomannan | 0.04 ± 0.04 | 0.09 ± 0.02 |
| galactan | 0.27 ± 0.33 | 0.59 ± 0.12 |
| amylose | 0.38 ± 0.06 | 0.27 ± 0.04 |
| amylopectin | 0.90 ± 0.12 | 0.51 ± 0.09 |
| PASC | 0.46 ± 0.16 | 0.44 ± 0.13 |
| xylan | 0.36 ± 0.19 | 0.79 ± 0.21 |
| pectin | 0.55 ± 0.21 | 0.82 ± 0.09 |
| chitin | 0.38 ± 0.09 | 1.19 ± 0.74 |

[a]Activity was measured as described in Materials and Methods using partially purified recombinantly expressed GH61B from *T. terrestris* or partially purified recombinantly expressed GH61E from *T. terrestris*. The values are the amount of reducing sugar measured as a percentage of the total theoretical amount of reducing sugar present in the substrate and are means of triplicate determinations ± SEM and are corrected for background values in the absence of added GH61 protein.

important. During the course of this research, the structure of another GH61 protein, GH61B from *T. reesei* (TrGH61B), became available and the PDB coordinates (PDB ID 2VTC) released (Figure 6b) (*57*). TrGH61B shares only 29% structure-based sequence identity with GH61E, with an rmsd of 1.22 Å over 169 C-α atoms aligned (for the A chains). No biochemical information on the function of TrGH61B was reported, though the authors also cast serious doubts on the assignment as a glycoside hydrolase.

We tested the hydrolytic activity of two partially purified GH61 proteins on 16 polysaccharide substrates and found negligible activity as measured by production of reducing sugars even after 72 h of incubation (Table 2). In particular, activity on carboxymethylcellulose, a traditional endoglucanase substrate, was statistically indistinguishable from background. More thorough analysis by HPLC of the reaction products produced on phosphoric acid-swollen cellulose showed no detectable release of glucose, cellobiose, cellotriose, or cellotetraose after 72 h incubation with *T. terrestris* GH61B and 0.06% conversion to cellobiose by GH61E. In comparison, the Cel7B endoglucanase from *T. reesei* converts approximately 41% of the cellulose to soluble reducing sugars at the same protein dosing and time of incubation. We also assayed *T. aurantiacus* GH61A for transglycosidase activity with either glucose or cellobiose as substrate. Based on mass spectrometry analysis, no higher mass products were formed.

*The Metal Ion-Binding Site and Metal Ion Requirement for GH61 Function*. Potentially the most interesting feature in the GH61E and TrGH61B structures is the clustering of three highly conserved histidines at the surface near the N-terminus. In the first GH61E structure obtained, there appeared to be a metal ion in this region, identified as magnesium from the crystallization buffer, coordinated by two of the histidines. The lack of liganding carboxylates and the presence of the three histidines suggested a potential zinc ion-binding site. We thus soaked the crystals in a solution containing zinc sulfate in order to obtain a

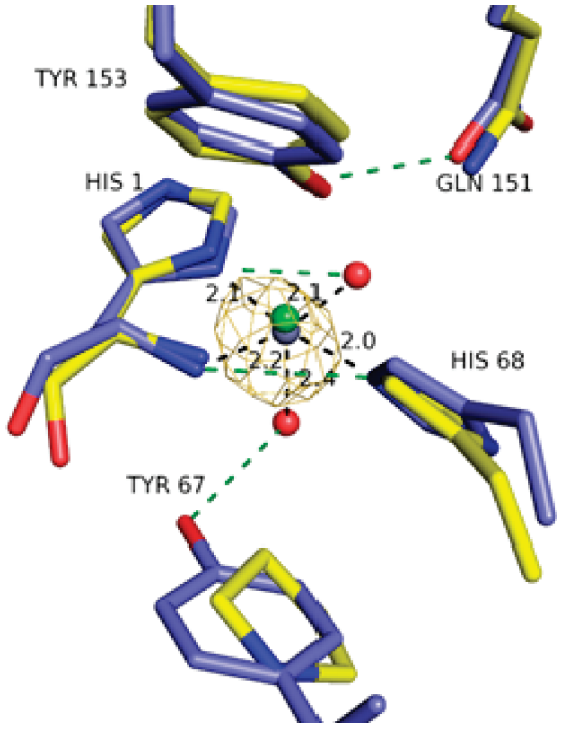


FIGURE 7: Structure of the metal ion-binding site of GH61E (carbon atoms in blue) compared with TrGH61B (carbon atoms in yellow). The tetragonal pyramidal geometry of the zinc ion (gray sphere) bound to chain A of GH61E is shown by black dashed lines with the corresponding metal ion coordination distances. Two waters in GH61E are shown as red spheres, and the nickel ion in TrGH61B is shown as a green sphere. Distances of less than 3.0 Å between potential hydrogen-bonding pairs are indicated by green dashed lines. Tyr-67 in GH61E is not highly conserved but is often a polar amino acid, although it is a proline in TrGH61B. The hydroxyl oxygen of Tyr-153 in GH61E is positioned to potentially provide an additional coordination site for octahedral geometry, but the distance (2.97 Å) is substantially greater than typically observed for such sites (*58*). The corresponding tyrosine in TrGH61B is closer to the nickel ion (2.4−2.5 Å). The anomalous difference Fourier map is shown as a yellow net contoured at 5$\sigma$ level in a 8 Å radius region around the zinc ion, providing a positive identification for its chemical nature. Alignment of the two structures was performed using the program O (*42*) by considering only the two coordinating histidines and the α-carbon of the conserved tyrosine. The figure was made in PYMOL (*72*).

structure with bound zinc. The structure obtained was better resolved in the metal ion-binding region, and the anomalous difference map revealed a bound zinc ion in all molecules in the asymmetric unit (Figure 7); however, only two of the His are involved in binding. Metal ion-donor atom distances to the side chain and backbone nitrogens of His-1 and the side chain nitrogen of His-68 are very typical of $Zn^{2+}$ coordination (*58*), and two additional ligands (water and/or glycerol) are usually present within 0.5 Å of preferred coordination distance. The site has a tetragonal pyramidal geometry.

Similarly, the TrGH61B structure contains $Ni^{2+}$ (from the crystallization conditions) bound at approximately the same site where $Mg^{2+}$ and $Zn^{2+}$ are bound in GH61E (Figure 7). Apparently, at sufficiently high concentrations, GH61 metal ion-binding sites can accommodate different divalent metal ions. The high degree of sequence conservation in this region coupled with good coordinating distances with the protein ligands strongly suggested that this was a functional metal ion-binding site and prompted a series of experiments which show that a divalent

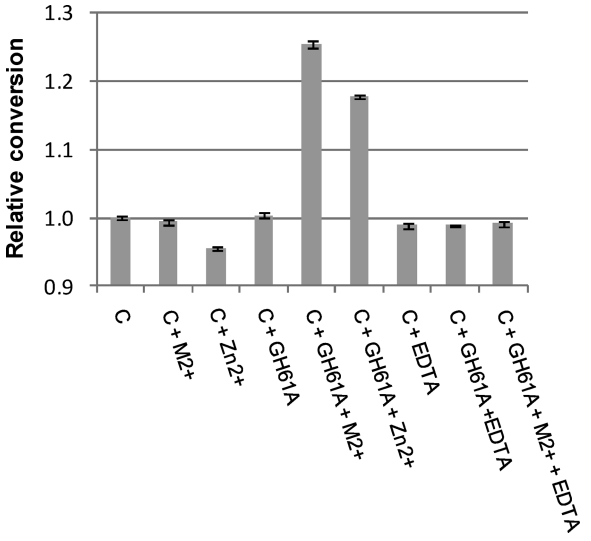


FIGURE 8: Metal ion requirement for GH61 stimulation of PCS hydrolysis. Hydrolysis was carried out as described in Materials and Methods with *T. reesei* cellulase (C) with or without *T. aurantiacus* GH61A, 1 mM metal ions, and/or EDTA. Data for divalent Ca, Co, Mg, Mn, and Ni were combined ($M^{2+}$) since the results were nearly identical for each. $Zn^{2+}$ was somewhat inhibitory to cellulase action and is shown separately. Relative conversion of cellulose to glucose is shown relative to hydrolysis for cellulase alone set to 1.0. Error bars are SEM with $n = 15$ for combined $M^{2+}$ data and $n = 3$ for all others.

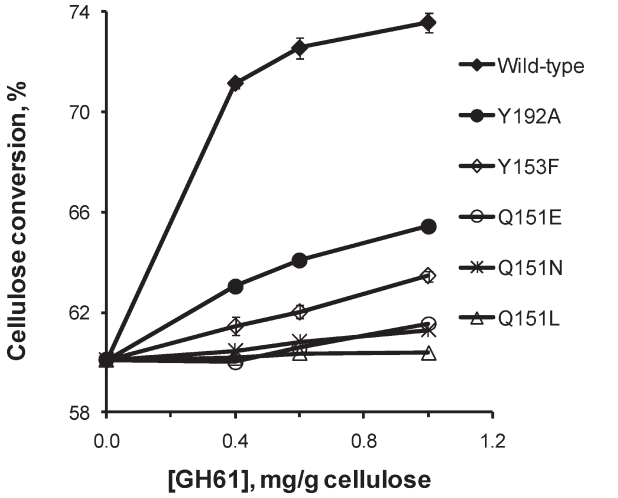


FIGURE 9: Effect of site-directed mutation on GH61E activity. Each mutant protein was added at the indicated concentration to a cellulase preparation from *T. reesei* strain SMA-135 loaded at 4 mg/g of cellulose. PCS hydrolysis was assayed as described in Materials and Methods with an incubation time of 72 h. Each point is the mean of a duplicate determination. The SEM was in most cases approximately the size of the symbol and for clarity is shown only for the wild-type and Y153F mutant. Data for the H1N and H68A mutations overlay the data for Q151L and for clarity are not included.

metal ion is essential for the ability of GH61 to stimulate PCS hydrolysis.

The metal ion requirement was investigated by adding GH61A from *T. aurantiacus* to a desalted cellulase mixture from *T. reesei* and assaying hydrolysis enhancement in the presence of the metal ion chelator EDTA or in the presence of added divalent metal ion. The PCS substrate in this case was washed with EDTA to remove sequestered metal ions. The results (Figure 8) show clearly that enhancement of PCS hydrolysis by GH61A is metal ion-dependent and that several divalent metal ions are functional in this regard. Similar results were obtained with GH61E. These results point to a metal ion-binding site with considerable plasticity, as already suggested by the structural data.

The importance of the metal ion-binding site was further investigated by mutagenizing those residues directly or indirectly involved in metal binding (Figure 9). Mutation of the directly interacting His-1 to Asn or His-68 to Ala resulted in a completely inactive protein. Mutation of the closely interacting Tyr-153 to Phe substantially reduced but did not eliminate activity. Mutation of Gln-151 (H-bonded to Tyr-153) to Leu was completely inactivating whereas the more conservative substitutions Asn or Glu retained a small residual activity.

A mutation was also made in one of three solvent-exposed tyrosines (Tyr-192) that form a relatively flat planar surface on GH61E that resembles the polysaccharide binding surface present in family 1 cellulose binding modules (Figure 6c). Mutation to Ala reduced activity substantially but far from completely (Figure 9), suggesting an important role for this residue, which is conserved in TrGH61B and moderately well conserved in other GH61 proteins.

## DISCUSSION

The cellulases produced by *T. reesei* have been the primary system for all commercial and most academic attempts to create an enzymatic system to degrade lignocellulose in an economic fashion. Despite many years of research, improving the specific activity of this cellulase system has proven extremely challenging, and the only substantial commercially relevant progress has come from supplementation with $\beta$-glucosidase and, depending on substrate, hemicellulases (e.g., ref *59*). Our discovery of the GH61 family as a major stimulatory factor is a significant breakthrough in reducing the enzyme loading required to hydrolyze lignocellulosic biomass to fermentable sugars. Enhancement of cellulase activity by GH61 is not limited to dilute acid-pretreated corn stover but also occurs with steam-exploded wheat and rice straw and organosolv-pulped lodgepole pine. Somewhat surprisingly, enhancement does not occur with any of the relatively pure cellulose substrates tested.

The structure of GH61E from *T. terrestris* and the recently determined structure of TrGH61B (*57*) fail to support the classification of these proteins as glycoside hydrolases since there is no clustering of conserved catalytic acidic side chains that constitute the canonical catalytic machinery for this class of proteins. We cannot rule out an unusual mechanism of hydrolytic catalysis, but two of the most highly purified and active GH61 proteins in our assays (GH61B and GH61E from *T. terrestris*) produce negligible amounts of reducing ends after prolonged incubation with a wide variety of polysaccharide substrates. The very small amount of reducing ends produced with some substrates could easily reflect residual contamination of the proteins with carbohydrases of the expression host *A. oryzae*. Also, few of these commercial substrates are 100% pure, and contaminating carbohydrases could be generating reducing ends from the impurities. It is always difficult to prove a negative, but our preparation of *T. terrestris* GH61B produces no detectable release of glucose, cellobiose, cellotriose, or cellotetraose after 72 h incubation with phosphoric acid-swollen cellulose, and thus it is clearly not an endoglucanase capable of degrading this substrate to the level of soluble oligosaccharides. Furthermore, since the cellulase-enhancing effect of GH61 is not seen with

relatively pure cellulosic substrates such as phosphoric acid-swollen cellulose, the effect of GH61 cannot be explained by simple endo-1,4-$\beta$-D-glucanase synergy with *Trichoderma* cellulases. These results leave unresolved the molecular mechanism of GH61 action but greatly lower the probability that GH61 proteins act by a classic hydrolytic mechanism.

Inspired by the presence of a metal ion in the structure, we have further shown that metal ions are necessary to GH61 function. Many ions seem to be able to fulfill this role, and this might be correlated to a structural plasticity of the site. In particular, the distance between the metal ion and the mostly conserved Tyr (153 in our structure) is considerably shorter for TrGH61B (about 2.4 Å) than in our structure (about 2.9 Å with Zn), but both are much longer than the normal distances for coordination (*58*). Also, in some GH61 sequences the equivalent residue is a Phe. However, mutation of this residue to Phe in GH61E causes a substantial but not complete reduction in activity, suggesting that the hydroxyl moiety is important for activity, at least in GH61E. Also important is Gln-151, a highly conserved residue in the GH61 family that in our structure and that of *T. reesei* GH61 forms a hydrogen bond with the hydroxyl of Tyr-153. Glutamate occurs at this position in a small subset of GH61 proteins, but mutation to glutamate in GH61E causes a substantial reduction in activity. Subtle differences in the binding sites might modulate the yet unclarified molecular function of GH61 proteins.

Karkehabadi et al. (*57*) noted a structural similarity between TrGH61B and chitin-binding protein CBP21 from *Serratia marcescens*, a protein that stimulates the chitin-degrading activity of chitinases while having no chitinase activity itself (*60*, *61*). We similarly found that CBP21 gives a highly significant *Z*-score of 8.8 when doing a search for structural homologues of GH61E based on the DaliLite method (*62*). Furthermore, the most highly conserved residues in homologues of CBP21 correspond structurally to the metal-binding site in GH61, and mutation of these residues significantly impacts the ability of CBP21 to either bind chitin or to stimulate chitinase activity (*60*, *61*). For example, mutation of His-114 in CBP21 to Ala eliminates activity, and mutation of the structurally equivalent His-68 in GH61E to Ala similarly eliminates activity. Interestingly, the equivalent of GH61E Tyr-153 in CPB21 is Phe, a substitution which partly inactivates GH61E. The apparent structural and functional homology between GH61 and CBP21, if truly the result of common evolutionary origin and not fortuitous convergent evolution, suggests an ancient origin for a divergent and previously unrecognized superfamily of proteins that could function to enhance the degradation of polysaccharide substrates by a common but unknown mechanism. Recently, a CBP21 homologue from *Thermobifida fusca* was shown to boost cellulase hydrolysis of filter paper, albeit at a high concentration relative to the cellulases (*63*), suggesting that some members of this family (carbohydrate-binding module family 33) may possess an activity similar to GH61.

A number of hypotheses could be put forward to explain the action of GH61 in enhancing lignocellulose degradation. As previously mentioned there are literature reports of extremely weak endoglucanase activity in some members of the family, and although we do not concur with the classification of GH61 proteins as canonical $\beta$-1,4-endoglucanases, it is conceivable that GH61 proteins cleave a relatively rare bond found in lignocellulose and that this bond is particularly effective at impeding cellulase activity. Currently, we have no evidence to support or reject this hypothesis other than the lack of significant activity on the substrates that we have tested. The possibility of an "expansin-like" hydrogen-bond disrupting activity seems unlikely given the lack of GH61 enhancement of cellulase activity on relatively pure forms of cellulose. A disruption of hemicellulose structure is conceivable if such disruption influences accessibility of the cellulases to cellulose; however, we see little effect of GH61 on release of xylose from PCS, suggesting any effect of GH61 on the xylan component of PCS has little or no influence on the hydrolyzability of xylan by hemicellulases. The requirement for metal ion in GH61 action has precedence in various carbohydrate-active enzymes. For example, proteins of carbohydrate esterase family 4 typically have a conserved metal-binding triad (His, His, Asp), and the structures of several family members suggest that a bound metal ion functions as a Lewis acid and facilitates the nucleophilic activity of a metal-bound water molecule at the active site. As with GH61 proteins, several divalent metal ions can function in this regard, although with significantly different efficacies depending on the protein (*64*, *65*). A metal ion could also be involved in substrate binding as in the $Ca^{2+}$-dependent xylan-binding domain (CBM36) of *Paenibacillus polymyxa* xylanase 43A (*66*).

Any explanation of GH61 function must accommodate the existence of multiple GH61 genes in cellulolytic fungi and particularly the dramatic amplification of the family in certain species. We have evidence that not all GH61 proteins are created equal and many are completely inactive as enhancers of PCS hydrolysis with *T. reesei* cellulases whereas others stimulate PCS hydrolysis by cellulases from dozens of different species. This could be the result of many factors including temperature stability and pH optimum; however, the proteins may also have subtle differences in substrate specificity that will only become apparent when looking at a broader array of plant cell wall materials, particularly those in a more native state.

Regardless of the mechanism of GH61 action, we have demonstrated that relatively low amounts of these proteins can dramatically stimulate the hydrolysis of lignocellulosic substrates such as PCS. Coexpression of highly active GH61 proteins in the cellulase producer *T. reesei* can increase the apparent specific activity of the *Trichoderma* cellulases by approximately a factor of 2 when targeting high cellulose conversion levels. *T. reesei* expresses its own GH61 proteins, but none of the three present in the genome sequence is expressed at significant levels in our strain under the culture conditions employed. The *T. reesei* GH61B protein, which appears to be expressed at the highest level, is only slightly active in stimulating PCS hydrolysis. This relative lack of GH61 activity no doubt accounts for the ability of exogenous or heterologously expressed GH61 protein to significantly stimulate *T. reesei* cellulase activity on PCS. Current models for economical conversion of lignocellulosic biomass to biofuels typically assume conversion levels of 80−90%, although the optimum value with respect to lowering the product selling price will depend on a number of variables, particularly the relative cost of enzymes and biomass (*9*). At the 80−90% conversion level, the reduction in protein loading afforded by GH61 translates directly to a 1.7−1.9-fold cost reduction and further enables the enzymatic saccharification platform for production of cheap sugars and biofuels from abundant and renewable lignocellulosic biomass.

## ACKNOWLEDGMENT

We thank Flemming Hansen and the MAX-lab staff for help with data collection. Valuable technical assistance was provided

by Derek Akerhielm, William Albano, David Dotson, Elizabeth Zaretsky, Randall Kramer, James Langston, Alfredo Lopez de Leon, Suchindra Maiyuran, Jason Quinlan, and Matt Sweeney. Mark Wogulis provided partially purified GH61 proteins for carbohydrase analysis and critiqued the manuscript.

## REFERENCES


1. Farrell, A. E., Plevin, R. J., Turner, B. T., Jones, A. D., O'Hare, M., and Kammen, D. M. (2006) Ethanol can contribute to energy and environmental goals. *Science 311*, 506–508.
2. van Wyk, J. P. (2001) Biotechnology and the utilization of biowaste as a resource for bioproduct development. *Trends Biotechnol. 19*, 172–177.
3. Wyman, C. E. (2003) Potential synergies and challenges in refining cellulosic biomass to fuels, chemicals, and power. *Biotechnol. Prog. 19*, 254–262.
4. Perlack, R., Wright, L., Turhollow, A., Graham, R., Stokes, B., and Erbach, D. (2005) Biomass as feedstock for a bioenergy and bioproducts industry: The technical feasibility of a billion-ton annual supply (U.S. Department of Energy and U.S. Department of Agriculture, Eds.) U.S. Department of Energy, Washington, DC.
5. National Academy of Sciences, National Academy of Engineering, National Research Council (2009) Liquid Transportation Fuels from Coal and Biomass: Technological Status, Costs, and Environmental Impacts, National Academy of Sciences, Washington, DC.
6. Clements, L. D., and van Dyne, D. L. (2006) The lignocellulosic biorefinery—a strategy for returning to a sustainable source of fuels and industrial organic chemicals, in Biorefineries—Industrial Processes and Products (Kamm, B., Gruber, P. R., and Kamm, M., Eds.) pp 115−127, Wiley-VCH, Weinheim.
7. Kamm, B., and Kamm, M. (2007) Biorefineries—multi product processes. *Adv. Biochem. Eng./Biotechnol. 105*, 175–204.
8. Galbe, M., Sassner, P., Wingren, A., and Zacchi, G. (2007) Process engineering economics of bioethanol production. *Adv. Biochem. Eng./Biotechnol. 108*, 303–327.
9. Aden, A. (2008) Biochemical production of ethanol from corn stover: 2007 state of technology model, National Renewable Energy Laboratory.
10. Aden, A., and Foust, T. (2009) Technoeconomic analysis of the dilute sulfuric acid and enzymatic hydrolysis process for the conversion of corn stover to ethanol. *Cellulose 16*, 535–545.
11. Lynd, L. R., Weimer, P. J., van Zyl, W. H., and Pretorius, I. S. (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev. 66*, 506–577.
12. Foreman, P. K., Brown, D., Dankmeyer, L., Dean, R., Diener, S., Dunn-Coleman, N. S., Goedegebuur, F., Houfek, T. D., England, G. J., Kelley, A. S., Meerman, H. J., Mitchell, T., Mitchinson, C., Olivares, H. A., Teunissen, P. J., Yao, J., and Ward, M. (2003) Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus Trichoderma reesei. *J. Biol. Chem. 278*, 31988–31997.
13. Kubicek, C. P., Eveleigh, D. E., Esterbauer, H., Steiner, W., and Kubicek-Pranz, E. M. (1990) Trichoderma reesei Cellulases, Royal Society of Chemistry, Cambridge, U.K.
14. Fryksdale, B. G., Jedrzejewski, P. T., Wong, D. L., Gaertner, A. L., and Miller, B. S. (2002) Impact of deglycosylation methods on two-dimensional gel electrophoresis and matrix assisted laser desorption/ionization-time of flight-mass spectrometry for proteomic analysis. *Electrophoresis 23*, 2184–2193.
15. Markov, A. V., Gusakov, A. V., Dzedziulia, E. I., Ustinov, B. B., Antonov, A. A., Okunev, O. N., Bekkarevich, A. O., and Sinitsyn, A. P. (2006) Properties of hemicellulases of the enzyme complex from *Trichoderma longibrachiatum*. *Prikl. Biokhim. Mikrobiol. 42*, 654–664.
16. Martinez, D., Berka, R. M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S. E., Chapman, J., Chertkov, O., Coutinho, P. M., Cullen, D., Danchin, E. G., Grigoriev, I. V., Harris, P., Jackson, M., Kubicek, C. P., Han, C. S., Ho, I., Larrondo, L. F., de Leon, A. L., Magnuson, J. K., Merino, S., Misra, M., Nelson, B., Putnam, N., Robbertse, B., Salamov, A. A., Schmoll, M., Terry, A., Thayer, N., Westerholm-Parvinen, A., Schoch, C. L., Yao, J., Barbote, R., Nelson, M. A., Detter, C., Bruce, D., Kuske, C. R., Xie, G., Richardson, P., Rokhsar, D. S., Lucas, S. M., Rubin, E. M., Dunn-Coleman, N., Ward, M., and Brettin, T. S. (2008) Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat. Biotechnol. 26*, 553–560.
17. Vinzant, T. B., Adney, W. S., Decker, S. R., Baker, J. O., Kinter, M. T., Sherman, N. E., Fox, J. W., and Himmel, M. E. (2001) Fingerprinting *Trichoderma reesei* hydrolases in a commercial cellulase preparation. *Appl. Biochem. Biotechnol. 91−93*, 99–107.
18. Bayer, E. A., Chanzy, H., Lamed, R., and Shoham, Y. (1998) Cellulose, cellulases and cellulosomes. *Curr. Opin. Struct. Biol. 8*, 548–557.
19. Maheshwari, R., Bharadwaj, G., and Bhat, M. K. (2000) Thermophilic fungi: their physiology and enzymes. *Microbiol. Mol. Biol. Rev. 64*, 461–488.
20. Mosier, N. S., Hall, P., Ladisch, C. M., and Ladisch, M. R. (1999) Reaction kinetics, molecular action, and mechanisms of cellulolytic proteins. *Adv. Biochem. Eng./Biotechnol. 65*, 23–40.
21. Tomme, P., Warren, R. A., and Gilkes, N. R. (1995) Cellulose hydrolysis by bacteria and fungi. *Adv. Microb. Physiol. 37*, 1–81.
22. Xu, F. (2004) Enhancing biomass conversion to fermentable sugars: A progress report of a joint government-industrial project, in Biotechnology of Lignocellulose Degradation and Biomass Utilization (Ohmiya, K., Sakka, K., Karita, S., Kimura, T., Sakka, M., and Onishi, Y., Eds.) pp 793−804, Uni Publishers, Tokyo.
23. Zhang, Y. H., and Lynd, L. R. (2004) Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. *Biotechnol. Bioeng. 88*, 797–824.
24. Merino, S. T., and Cherry, J. (2007) Progress and challenges in enzyme development for biomass utilization. *Adv. Biochem. Eng./Biotechnol. 108*, 95–120.
25. Lamsa, M., Fidantsef, A., and Gorre-Clancy, B. (2004) Variants of beta-glucosidases, Patent Number WO2004099228A2, World Intellectual Property Organization.
26. Merino, S. (2008) Methods of increasing secretion of polypeptides having biological activity, Patent Number WO2008057637A2, World Intellectual Property Organization.
27. Dotson, W. D., Greenier, J., and Ding, H. (2007) Polypeptides having cellulolytic enhancing activity and polynucleotides encoding same, Patent Number US 7,271,244, USPTO.
28. Brown, K., Harris, P., Zaretsky, E., Re, E., Vlasenko, E., McFarland, K., and Lopez de Leon, A. (2008) Polypeptide from a cellulolytic fungus having cellulolytic enhancing activity, Patent Number US 11/046,124, USPTO.
29. Schell, D. J., Farmer, J., Newman, M., and McMillan, J. D. (2003) Dilute-sulfuric acid pretreatment of corn stover in pilot-scale reactor: investigation of yields, kinetics, and enzymatic digestibilities of solids. *Appl. Biochem. Biotechnol. 105−108*, 69–85.
30. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol. 215*, 403–410.
31. Altschul, S. F., Wootton, J. C., Gertz, E. M., Agarwala, R., Morgulis, A., Schaffer, A. A., and Yu, Y. K. (2005) Protein database searches using compositionally adjusted substitution matrices. *FEBS J. 272*, 5101–5109.
32. Jain, E., Bairoch, A., Duvaud, S., Phan, I., Redaschi, N., Suzek, B. E., Martin, M. J., McGarvey, P., and Gasteiger, E. (2009) Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinf. 10*, 136.
33. Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res. 33*, 511–518.
34. Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res. 30*, 3059–3066.
35. Guindon, S., Delsuc, F., Dufayard, J. F., and Gascuel, O. (2009) Estimating maximum likelihood phylogenies with PhyML. *Methods Mol. Biol.* (*Clifton, NJ*) *537*, 113–137.
36. Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol. 52*, 696–704.
37. Gascuel, O. (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol. 14*, 685–695.
38. Hordijk, W., and Gascuel, O. (2005) Improving the efficiency of SPR moves in phylogenetic tree search methods based on maximum likelihood. *Bioinformatics* (*Oxford, England*) *21*, 4338–4347.
39. Le, S. Q., and Gascuel, O. (2008) An improved general amino acid replacement matrix. *Mol. Biol. Evol. 25*, 1307–1320.
40. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, in Methods in Enzymology (Carter, C. W. J., and Sweet, R. M., Eds.) pp 307−326, Academic Press, New York.
41. Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr. 55*, 849–861.
42. Jones, T. A., Zou, J. Y., Cowan, S. W., 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* (Part 2), 110–119.
43. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. 60*, 2126–2132.
44. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. 54*, 905–921.
45. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. 53*, 240–255.
46. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr. 26*, 283–291.
47. Lever, M. (1972) A new reaction for colorimetric determination of carbohydrates. *Anal. Biochem. 47*, 273–279.
48. Bairoch, A. (1993) The PROSITE dictionary of sites and patterns in proteins, its current status. *Nucleic Acids Res. 21*, 3097–3103.
49. Berney, C., and Pawlowski, J. (2006) A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. *Proc. Biol. Sci. 273*, 1867–1872.
50. Bauer, S., Vasu, P., Persson, S., Mort, A. J., and Somerville, C. R. (2006) Development and application of a suite of polysaccharide-degrading enzymes for analyzing plant cell walls. *Proc. Natl. Acad. Sci. U.S.A. 103*, 11417–11422.
51. Karlsson, J., Saloheimo, M., Siika-Aho, M., Tenkanen, M., Penttila, M., and Tjerneld, F. (2001) Homologous expression and characterization of Cel61A (EG IV) of *Trichoderma reesei*. *Eur. J. Biochem. 268*, 6498–6507.
52. Koseki, T., Mese, Y., Fushinobu, S., Masaki, K., Fujii, T., Ito, K., Shiono, Y., Murayama, T., and Iefuji, H. (2008) Biochemical characterization of a glycoside hydrolase family 61 endoglucanase from *Aspergillus kawachii*. *Appl. Microbiol. Biotechnol. 77*, 1279–1285.
53. Saloheimo, M., Nakari-Setala, T., Tenkanen, M., and Penttila, M. (1997) cDNA cloning of a *Trichoderma reesei* cellulase and demonstration of endoglucanase activity by expression in yeast. *Eur. J. Biochem. 249*, 584–591.
54. Davies, G., and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure 3*, 853–859.
55. Holmquist, M. (2000) Alpha/beta-hydrolase fold enzymes: structures, functions and mechanisms. *Curr. Protein Pept. Sci. 1*, 209–235.
56. Ly, H. D., and Withers, S. G. (1999) Mutagenesis of glycosidases. *Annu. Rev. Biochem. 68*, 487–522.
57. Karkehabadi, S., Hansson, H., Kim, S., Piens, K., Mitchinson, C., and Sandgren, M. (2008) The first structure of a glycoside hydrolase family 61 member, Cel61B from *Hypocrea jecorina*, at 1.6 Å resolution. *J. Mol. Biol. 383*, 144–154.
58. Zheng, H., Chruszcz, M., Lasota, P., Lebioda, L., and Minor, W. (2008) Data mining of metal ion environments present in protein structures. *J. Inorg. Biochem. 102*, 1765–1776.
59. Berlin, A., Maximenko, V., Gilkes, N., and Saddler, J. (2007) Optimization of enzyme complexes for lignocellulose hydrolysis. *Biotechnol. Bioeng. 97*, 287–296.
60. Vaaje-Kolstad, G., Horn, S. J., van Aalten, D. M., Synstad, B., and Eijsink, V. G. (2005) The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. *J. Biol. Chem. 280*, 28492–28497.
61. Vaaje-Kolstad, G., Houston, D. R., Riemen, A. H., Eijsink, V. G., and van Aalten, D. M. (2005) Crystal structure and binding properties of the *Serratia marcescens* chitin-binding protein CBP21. *J. Biol. Chem. 280*, 11313–11319.
62. Holm, L., and Park, J. (2000) DaliLite workbench for protein structure comparison. *Bioinformatics* (*Oxford, England*) *16*, 566–567.
63. Moser, F., Irwin, D., Chen, S., and Wilson, D. B. (2008) Regulation and characterization of *Thermobifida fusca* carbohydrate-binding module proteins E7 and E8. *Biotechnol. Bioeng. 100*, 1066–1077.
64. Blair, D. E., Schuttelkopf, A. W., MacRae, J. I., and van Aalten, D. M. (2005) Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor. *Proc. Natl. Acad. Sci. U.S.A. 102*, 15429–15434.
65. Taylor, E. J., Gloster, T. M., Turkenburg, J. P., Vincent, F., Brzozowski, A. M., Dupont, C., Shareck, F., Centeno, M. S., Prates, J. A., Puchart, V., Ferreira, L. M., Fontes, C. M., Biely, P., and Davies, G. J. (2006) Structure and activity of two metal ion-dependent acetylxylan esterases involved in plant cell wall degradation reveals a close similarity to peptidoglycan deacetylases. *J. Biol. Chem. 281*, 10968–10975.
66. Jamal-Talabani, S., Boraston, A. B., Turkenburg, J. P., Tarbouriech, N., Ducros, V. M., and Davies, G. J. (2004) Ab initio structure determination and functional characterization of CBM36; a new family of calcium-dependent carbohydrate binding modules. *Structure 12*, 1177–1187.
67. Rambaut, A. (2009) FigTree version 1.2.2 (http://tree.bio.ed.ac.uk/software/figtree/).
68. Frishman, D., and Argos, P. (1995) Knowledge-based protein secondary structure assignment. *Proteins 23*, 566–579.
69. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graphics 14* (33−38), 27–38.
70. Russell, R. B., and Barton, G. J. (1992) Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins 14*, 309–323.
71. The POV-Ray Team (2004) Persistence of vision Raytracer (Version 3.6), Persistence of Vision Pty. Ltd., Williamstown, Victoria, Australia.
72. DeLano, W. L. (2008) The PyMOL molecular graphics system, DeLano Scientific LLC, Palo Alto, CA.