



# Enhancement of Synthetic *Trichoderma*-Based Enzyme Mixtures for Biomass Conversion with an Alternative Family 5 Glycosyl Hydrolase from *Sporotrichum thermophile*

Zhuoliang Ye<sup>1‡</sup>, Yun Zheng<sup>2</sup>, Bingyao Li<sup>1</sup>, Melissa S. Borrusch<sup>1</sup>, Reginald Storms<sup>2</sup>, Jonathan D. Walton<sup>1\*</sup>

**1** Department of Energy Great Lakes Bioenergy Research Center and Department of Energy Plant Research Laboratory, Michigan State University, East Lansing, Michigan, United States of America, **2** Centre for Structural and Functional Genomics, Concordia University, Montréal, Quebec, Canada

## Abstract

Enzymatic conversion of lignocellulosic materials to fermentable sugars is a limiting step in the production of biofuels from biomass. We show here that combining enzymes from different microbial sources is one way to identify superior enzymes. Extracts of the thermophilic fungus *Sporotrichum thermophile* (synonym *Myceliophthora thermophila*) gave synergistic release of glucose (Glc) and xylose (Xyl) from pretreated corn stover when combined with an 8-component synthetic cocktail of enzymes from *Trichoderma reesei*. The *S. thermophile* extracts were fractionated and an enhancing factor identified as endo- $\beta$ 1,4- glucanase (StCel5A or EG2) of subfamily 5 of Glycosyl Hydrolase family 5 (GH5\_5). In multi-component optimization experiments using a standard set of enzymes and either StCel5A or the ortholog from *T. reesei* (TrCel5A), reactions containing StCel5A yielded more Glc and Xyl. In a five-component optimization experiment (i.e., varying four core enzymes and the source of Cel5A), the optimal proportions for TrCel5A vs. StCel5A were similar for Glc yields, but markedly different for Xyl yields. Both enzymes were active on lichenan, glucomannan, and oat  $\beta$ -glucan; however, StCel5A but not TrCel5A was also active on  $\beta$ 1,4-mannan, two types of galactomannan, and  $\beta$ 1,4-xylan. Phylogenetically, fungal enzymes in GH5\_5 sorted into two clades, with StCel5A and TrCel5A belonging to different clades. Structural differences with the potential to account for the differences in performance were deduced based on the known structure of TrCel5A and a homology-based model of StCel5A, including a loop near the active site of TrCel5A and the presence of four additional Trp residues in the active cleft of StCel5A. The results indicate that superior biomass-degrading enzymes can be identified by exploring taxonomic diversity combined with assays in the context of realistic enzyme combinations and realistic substrates. Substrate range may be a key factor contributing to superior performance within GH5\_5.

**Citation:** Ye Z, Zheng Y, Li B, Borrusch MS, Storms R, et al. (2014) Enhancement of Synthetic *Trichoderma*-Based Enzyme Mixtures for Biomass Conversion with an Alternative Family 5 Glycosyl Hydrolase from *Sporotrichum thermophile*. PLoS ONE 9(10): e109885. doi:10.1371/journal.pone.0109885

**Editor:** Alberto G. Passi, University of Insubria, Italy

**Received:** March 13, 2014; **Accepted:** September 1, 2014; **Published:** October 8, 2014

**Copyright:** © 2014 Ye et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability:** The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

**Funding:** This work was funded by the United States Department of Energy Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-07ER64494). Additional funding was provided by grant DE-FG02-91ER200021 to the MSU-Plant Research Laboratory from the United States Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences and Biosciences. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* Email: walton@msu.edu

‡ Current address: Energy Biosciences Institute, University of Illinois, Urbana, Illinois, United States of America

## Introduction

The production of liquid biofuels from lignocellulosic materials such as agricultural and forestry residues requires the conversion of the refractory plant cell wall polysaccharides to fermentable sugars by a combination of thermochemical and enzymatic treatments. Because of the complexity of plant cell walls, multiple enzymes are necessary for efficient conversion, including cellulases such as cellobiohydrolases and endo- $\beta$ 1,4-glucanases, and hemicellulases such as endo- $\beta$ 1,4-xylanases and  $\alpha$ -glucuronidases. Different pretreatment and feedstock combinations require enzyme cocktails of different composition and relative proportions [1–3].

Currently available cellulase mixtures, most of which are produced by fermentation of filamentous fungi such as *Trichoderma reesei*, contain more than 80 proteins [4]. Beyond the core cellulases and hemicellulases, the contribution of most of these proteins to lignocellulose deconstruction is not well understood. One strategy to improve our understanding of the relative and absolute importance of specific enzymes is to construct synthetic mixtures [5–8]. Synthetic mixtures containing up to 16 components have been created and optimized using response surface methodologies and robotic liquid dispensing [5]. Such mixtures present a starting point for the identification and analysis of novel enzymes that enhance biomass conversion as well as for the identification of superior versions of known enzymes. Because of the structural complexity of lignocellulose and the often-observed

synergism among different types and sources of cellulases, hemicellulases, and side-chain degrading enzymes [10,11], it is important to do the screening of novel enzymes in as realistic a context as possible. That is, tests of individual pure enzymes on model substrates do not necessarily give a realistic picture of how enzymes will behave in multi-component mixtures on natural lignocellulosic materials.

In this paper, we report the isolation and identification of an endo- $\beta$ 1,4-glucanase from the fungus *Sporotrichum thermophile* (also known as *Myceliophthora thermophila*, class Sordariomycetes) that gives higher glucose (Glc) yields than the orthologous enzyme from *T. reesei* when tested in synthetic enzyme cocktails on pretreated corn stover. Superiority was correlated with phylogenetic sequence relatedness within subfamily 5 of family 5 of glycosyl hydrolases (GH5\_5), substrate range, and several deduced structural features.

## Materials and Methods

### Biological materials

Corn stover was ground to 0.5 mm particle size with a Wiley mill and pretreated with 12.5% alkaline hydrogen peroxide (AHP) as described [12]. The 8-component synthetic enzyme mixture (8-CSM) was assembled from purified enzymes. It contained (by mass) 35% cellobiohydrolase 1 (CBH1), 7% cellobiohydrolase 2 (CBH2; GenBank P07987), 5% endo- $\beta$ 1,4-glucanase 1 (EG1, GenBank AAA34212), 4%  $\beta$ -glucosidase (BG, GenBank AAA18473), 36% AA9 (formerly GH61A, GenBank CAA71999), 2% endo- $\beta$ 1,4-xylanase 2 (EX2, GenBank AAB29346), 9% endo- $\beta$ 1,4-xylanase 3 (EX3; GenBank BAA89465), and 3%  $\beta$ -xylosidase (BX; GenBank CAA93248). CBH1 was obtained from Megazyme, Ltd. (Bray, Ireland) and the other proteins were expressed in *Pichia pastoris*, including *T. reesei* Cel5A (TrCel5A; GenBank ABA64553 or AAA34213), by the method described earlier [13]. Bulk quantities were produced by Lucigen, Inc. (Madison, WI). *S. thermophile* Cel5A (StCel5A) was produced by expression in *A. niger* [14]. Both TrCel5A and StCel5A had their native cellulose binding modules (CBM) and no purification tags.

### Chromatography

AlternaFuel CMAX (lot #MXNA2413C), prepared from *S. thermophile*, was a generous gift of Dyadic International, Inc., Jupiter, FL. Enzyme purification was performed with an Agilent HPLC 1200 system equipped with UV detector and fraction collector. All chromatography columns (Tosoh Bioscience, LLC, King of Prussia, PA.) were 8.0 mm $\times$ 7.5 cm and packed with TSKgel SuperQ-5PW (anion exchange), TSKgel SP-5PW (cation exchange), or TSKgel Phenyl-5PW (hydrophobic interaction).

Solution A for anion exchange was 25 mM Tris-HCl, pH 8.0, and solution B was the same plus 0.6 M NaCl. Solution A for cation exchange was 25 mM sodium acetate, pH 4.0, and solution B was the same plus 0.6 M NaCl. Solution A for hydrophobic interaction was 0.1 M  $\text{KH}_2\text{PO}_4$ , pH 7, plus 1.7 M  $\text{NH}_2\text{SO}_4$ , and solution B was water. In all cases, the gradients increased from 0 to 100% solution B in 30 min at a flow rate of 1 mL/min, and then held at 100% solution B another 10 min. The fractions (1 mL) were collected in micro-centrifuge tubes at a rate of 1 mL/min.

### Enzyme activity assays

Each 0.5-ml GENPLAT reaction contained biomass at a final glucan concentration of 2 mg/mL in 20 mM sodium citrate buffer, pH 4.8, containing 10  $\mu\text{g}/\text{mL}$  cycloheximide and 10  $\mu\text{g}/\text{mL}$  tetracycline. Hydrolysis reactions were performed in 96-deep well plates at 50°C for 24, 48, or 72 hr. At each time point, aliquots were

removed and free Glc and xylose (Xyl) measured using enzyme-linked assays as described [5]. Mannose (Man) was measured using the assay kit K-MANGL (Megazyme, Wicklow, Ireland). Every reaction was performed in duplicate, sampled twice, and free Glc, Xyl, or Man measured twice from each sample.

To assay synergistic activity, 8-CSM at a loading of 6.5  $\mu\text{g}/\text{mg}$  glucan was mixed with 50  $\mu\text{L}$  of each chromatographic fraction. Controls included each fraction by itself and 8-CSM by itself. Degree of synergy is defined as (Glc released by 8-CSM and fraction together) divided by (Glc released from fraction alone plus 8-CSM alone).

Substrate range was tested using various polysaccharide substrates [15]. Sigmacell Type 20, lichenan (from *Cetraria islandica*), xylan (from birchwood), guar gum (from *Cyamopsis tetragonoloba*), locust bean gum (from *Ceratonia siliqua*), and  $\beta$ 1,3-glucan (from *Laminaria digitata*) were obtained from Sigma-Aldrich (St. Louis, MO).  $\beta$ -Glucan (from oat, high viscosity, catalog P-BGOH), xyloglucan (from tamarind, catalog P-XYGLN),  $\beta$ 1,4-mannan (galactomannan from *Ceratonia siliqua* treated with  $\beta$ -mannanase and  $\alpha$ -galactosidase; catalog P-MANCB) and arabinoxylan (from wheat, catalog P-WAXYH) were obtained from Megazyme. Glucomannan (from roots of *Amorphophallus konjac*) was obtained from a local health food store (NOW Foods, Bloomingdale, IL). Phosphoric-acid swollen cellulose (PASC) was prepared as described [16]. Assay conditions were 5 mg/mL substrate concentration, 5  $\mu\text{g}/\text{mL}$  enzyme concentration, pH 5.0, and 50°C in 1.5-mL centrifuge tubes in an incubator rotating at 10 rpm. Incubation times were varied from 10 min to 24 hr to insure that all reaction rates were in the linear range. Release of reducing sugars was detected by absorbance at 540 nm after reaction with 3,5-dinitrosalicylic acid (DNS) [17]. All reducing sugar values are expressed as Glc equivalents. Enzyme reaction rates were analyzed using Microsoft Excel and GraphPad Prism (San Diego, CA).

Lichenan (1 mg/mL) was used as the substrate for the characterization of the pH and temperature optima of StCel5A. For the pH experiment, reactions were carried out at 50°C and pH 3–8, using 50 mM sodium citrate for the pH range 3 to 6, and 50 mM Tris-HCl for pH 7 to 8. For the temperature experiment, reactions were carried out in a total volume of 0.1 mL in 25 mM sodium citrate, pH 4.8. Reducing sugars were quantitated with p-aminohydroxybenzoic acid hydrazide [18].

### Protein analysis

Chromatographic fractions showing synergistic activity were analyzed by SDS-PAGE. After staining with Coomassie Blue R-250, bands were excised, digested with trypsin, and analyzed by mass spectrometry at the MSU Research Technology Support Facility. Results were analyzed with Scaffold (version 4.0.5, Proteome Software Inc., Portland, OR).

Protein concentrations were determined by dye binding with bovine IgG as standard [19]. Protein quantitation was verified with SDS-PAGE and Coomassie staining compared to molecular weight markers of known concentration.

Protein alignments and tree construction were performed using MEGA version 6 [20]. Neighbor-Joining with 500 bootstraps, the Jones-Taylor-Thornton (JTT) substitution model, and partial deletion of gaps (to 284 sites) was used.

### Enzyme optimization for AHP-pretreated corn stover hydrolysis

Optimization of enzyme proportions utilized the high-throughput biomass digestion platform GENPLAT [3,5]. An augmented quadratic experimental design was determined with Design-Expert

software (Stat-Ease Inc., Minneapolis, MN). The hydrolysis was carried out at a substrate concentration of 2 mg/mL glucan and a fixed enzyme loading of 15 mg/gm glucan. Free Glc and Xyl were quantitated with enzyme-linked assays.

### Homology modeling and alignments

Modeling of StCel5A was performed using Swiss-Model [21] with *Thermoascus aurantiacus* Cel5A (TaCel5A; PDB:1GZJ) and TrCel5A (PDB:3QR3) as templates. Structure alignment and superposition of StCel5A and TrCel5A were performed using PyMol version 1.5.0.4 (Schrodinger, LLC; www.schrodinger.com).

## Results

### Separation and identification of an enzyme from *Sporotrichum thermophile* that enhances Glc yields

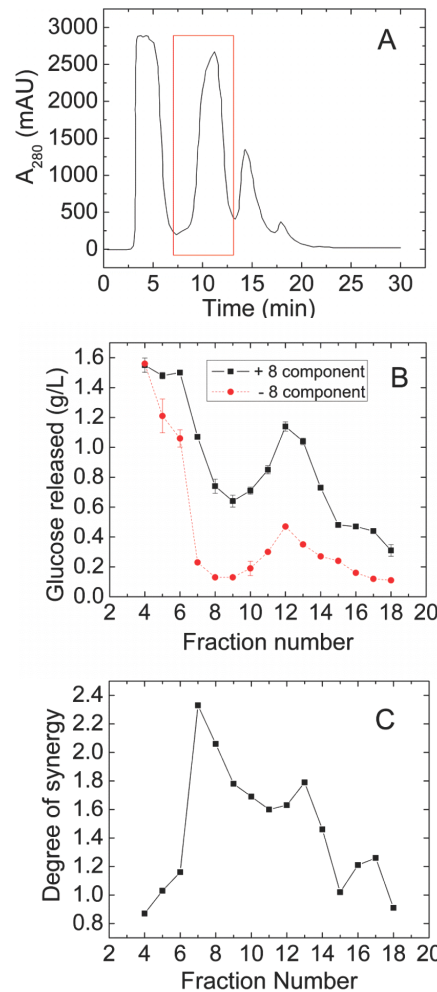
A complex mixture of the secretome of *S. thermophile* showed 1.6-fold synergy when added at a 1:2 ratio with a synthetic enzyme mixture containing CBH1, CBH2, EG1, EX2, EX3, BG, BX, and AA9 from *T. reesei* (see Materials and Methods for enzyme abbreviations and GenBank accession identifiers). The proportions of the different *T. reesei* enzymes in this synthetic mixture, called 8-CSM, were based on earlier studies [5]. The *S. thermophile* extract was fractionated in three chromatographic steps to identify the protein or proteins responsible for the synergism (Figures 1–3). Purification was guided by assaying each chromatographic fraction alone and in combination with the 8-CSM.

After the third separation (Fig. 3), fraction 37 consistently showed the highest synergy. Proteins in this fraction were analyzed by SDS-PAGE and mass-spectrometry proteomics based on the complete genome sequence of *S. thermophile* [22]. In three experiments, an endo- $\beta$ 1,4-glucanase of family 5 (GenBank accession numbers HQ163779, AEO53769, or AEM23898; Joint Genome Institute identifier Spoth2|86753, herein abbreviated StCel5A) was consistently identified as one of the most abundant proteins in the chromatographic fractions showing maximum synergism. For example, in one proteomics experiment, 113 peptides from StCel5A were observed, accounting for 48% of the total peptides with 24% protein coverage. The next most abundant proteins in this experiment were Spoth2|112050 (GH10 xylanase, 53 peptides), Spoth2|66729 (GH6  $\beta$ -glucanase, 33 peptides), and Spoth|112399 (GH131  $\beta$ -glucanase, 31 peptides). StCel5A has previously been studied and named EG2 or EG II [14,23–25]. It is endo-acting and shows synergy with CBH1 of *T. reesei* and other  $\beta$ 1,4-glucanases on filter paper, Avicel, and pretreated fir wood [14,24]. Variants of StCel5A with improved specific activity and thermostability have been developed through directed evolution [23].

These results indicated that StCel5A accounts for at least some of the synergism between *S. thermophile* extracts and the synthetic 8-CSM mixture of *T. reesei*, although it does not exclude the existence of other synergistic factors. The best ortholog in *T. reesei* to StCel5A is GenBank accession number ABA64553 (AAA34213), also known as EG2, referred to herein as TrCel5A.

### Enzyme mixture optimization

To further test whether StCel5A was responsible for the observed synergistic effect of the *S. thermophile* secretome, StCel5A was compared side-by-side with TrCel5A at equal molar loadings in mixture optimization experiments [5,13]. In these experiments, the lower limits of CBH1, EG1 and BG were set to 5% and the lower limit for the other enzymes including StCel5A and Tr Cel5A were set to 0%. Total enzyme loading in every reaction was fixed at 15 mg/g glucan.

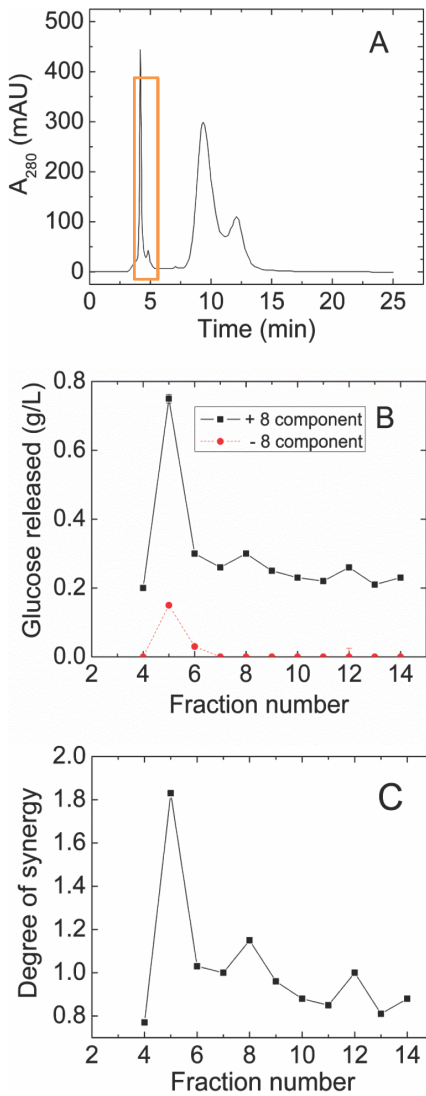


**Figure 1. First chromatographic fractionation (anion exchange) of *S. thermophile* secretome.** (A) UV trace. X-axis is retention time. The red box indicates the fractions collected for second-step separation. (B) Glc released from pretreated corn stover after incubation of an aliquot of each fraction from (A) with (black solid line) or without (red dashed line) 8-CSM. Glc released by 8-CSM alone was  $0.23 \pm 0.01$  g/L. Error bars indicate one standard deviation of the mean of triplicates. (C) Degree of synergy calculated from (B). doi:10.1371/journal.pone.0109885.g001

Results of the first optimization experiment are shown in Table S1. Only Glc release was assayed in this experiment. A subset of the results (i.e., reactions 1.43–1.55), in which all enzymes except StCel5A and TrCel5A were in the same proportions and concentrations, are shown in Fig. 4. Reactions containing StCel5A consistently gave higher Glc yields than those containing TrCel5A. The enhancement of Glc yield ranged from 27% to 156%. These results confirm that StCel5A is responsible for at least part of the synergistic activity of the *S. thermophile* secretome.

### Second optimization experiment

The first optimization experiment suggested that StCel5A was substituting for *T. reesei* endoglucanases and/or xylanases. To test this hypothesis, a second experiment was performed with varying proportions of CBH1, EG1, EX2, EX3, and either TrCel5A or StCel5A (Table S2). The proportions of these five enzymes were allowed to vary between 5% and 60%. The proportions of CBH2, BX, BG and AA9 were fixed at 5% each. In this way, a 9-

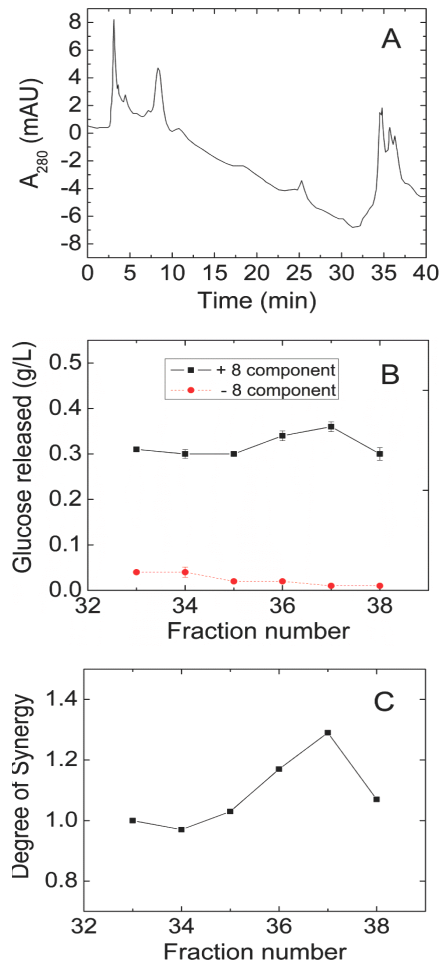


**Figure 2. Second fractionation (cation exchange) of fractions 7–10 from the separation shown in Figure 1. (A)**  $OD_{280}$ . The red box indicates the fractions collected for the third-step separation. **(B)** Activity with (black solid line) or without (red dashed line) 8-CSM as described in the legend to Fig. 1. Glc released by 8-CSM alone was  $0.24 \pm 0.01$  g/L. **(C)** Degree of synergy calculated from (B). doi:10.1371/journal.pone.0109885.g002

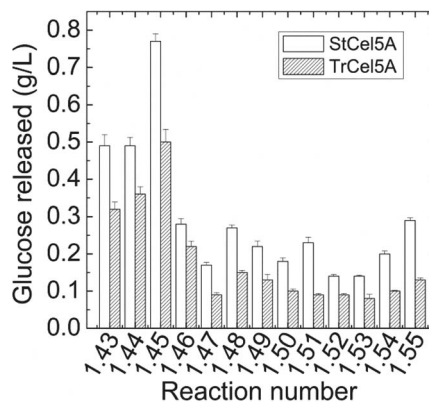
component mixture was reduced to a 5-component mixture and consequently the number of reactions needed to evaluate an augmented quadratic surface response model was reduced from 55 to 26.

Both Glc and Xyl yields were measured after 24, 48, or 72 h (Tables S3 and S4). At the best enzyme mixture, StCel5A yielded more Glc and Xyl than TrCel5A (Fig. 5). The enhancement of Glc yields by StCel5A ranged from 9% to 15%, and the enhancement of Xyl yields ranged from 9% to 10%.

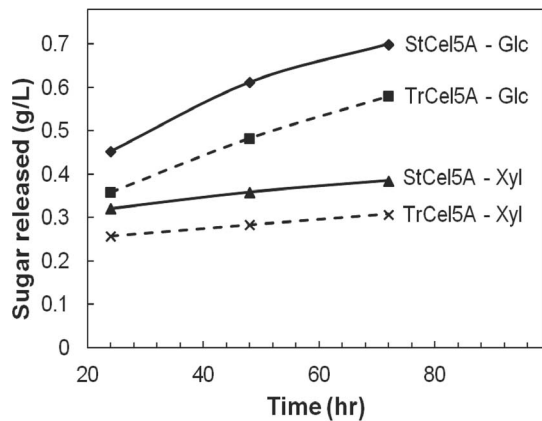
The model predictions of optimized proportions of component enzymes after 72 h, based on the data in Tables S3 and S4, are shown in Fig. 6. For Glc, the optimal proportions for mixtures containing either TrCel5A or StCel5A were fairly similar for EX2, EX3, and CBH1. The largest difference was in the optimal proportions of EG1 (50% for TrCel5A vs. only 31% for StCel5A), and in the proportions of Cel5A itself (6% for TrCel5A and 19%



**Figure 3. Third fractionation (hydrophobic interaction) of fraction 5 from the separation shown in Figure 2. (A)**  $OD_{280}$ . **(B)** Activity with (black solid line) or without (red dashed line) 8-CSM, as described in the legend to Fig. 1. Glc released by 8-CSM alone was  $0.26 \pm 0.01$  g/L. **(C)** Degree of synergy calculated from (B). doi:10.1371/journal.pone.0109885.g003



**Figure 4. Comparison of StCel5A and TrCel5A on Glc yields from corn stover in selected reactions in the first optimization experiment.** In each pair of reactions, the proportions of all of the enzymes except StCel5A and TrCel5A were identical. Table S1 shows the complete experimental results. doi:10.1371/journal.pone.0109885.g004



**Figure 5. Yields of Glc and Xyl as a function of time and source of Cel5A for 8-CSM plus either StCel5A or TrCel5A.** The experimental data are shown in Tables S3 and S4. Error bars (standard deviations) are smaller than the data symbols. doi:10.1371/journal.pone.0109885.g005

for StCel5A). This reciprocal relationship between Cel5A and EG1 is consistent with both of them being endo- $\beta$ 1,4-glucanases.

For optimal Xyl release, however, the enzyme proportions showed much greater differences depending on whether TrCel5A or StCel5A was in the mixture (Fig. 6). Whereas the optimal proportions of CBH1 and EX2 did not change appreciably, switching from TrCel5A to StCel5A decreased the optimal level of EX3 from 23% to 6%, while Cel5A increased from 6% to 61%, and EG1 decreased from 47% to 6%. This result suggests a particularly complex interaction between xylanases and glucanases for the efficient release of Xyl, with marked differences in behavior between TrCel5A and StCel5A.

### Comparisons of StCel5A and TrCel5A

The physical properties of StCel5A and TrCel5A are similar, with predicted  $M_r$ 's of 42,385 and 44,227, pI's of 5.15 and 4.85, three and one N-glycosylation sites, and signal peptides of 16 and 21 amino acids, respectively. Therefore, none of these properties are likely to contribute to their different abilities to digest biomass.

**pH and temperature.** The two enzymes showed similar pH profiles (Fig. 7A). Although StCel5A was earlier shown to have greater thermostability [14], our results indicated that the two enzymes were equally active at 50°C (Fig. 7B). Furthermore, when both enzymes were preincubated at 50°C for 48 hr and then assayed on lichenan for 48 hr, neither enzyme showed any detectable loss of activity (data not shown). Thus, neither pH nor temperature response appear to account for the differential effectiveness of TrCel5A and StCel5A.

**Kinetic parameters.** Tambor et al. [14] determined that the specificity constants ( $k_{cat}/K_m$ ) for TrCel5A and StCel5A were 1.8 and 1.5  $\text{mL s}^{-1} \text{mg}^{-1}$  on PASC and 6.1 and 15  $\text{mL s}^{-1} \text{mg}^{-1}$  on CM-cellulose. In a standardized side-by-side assay, 1.8 pmol of either TrCel5A or StCel5A produced 0.17 and 0.22 nmol reducing sugars, respectively, from Avicel. Hydrolysis of 2% PASC in 20 h required 1.43 pmol of TrCel5A and 2.05 pmol of StCel5A [14]. Although not conclusive, these values do not seem sufficiently different to account for the superiority of StCel5A at releasing Glc from corn stover when combined with other glycosyl hydrolases, especially considering the intrinsic uncertainty in kinetic measurements. Neither enzyme was tested on lichenan, xylan, or mannan in the prior study [14].

**Substrate specificities.** GH5 is a large family of glycosyl hydrolases, and members have a range of activities, including  $\beta$ 1,4-glucanase, lichenanase,  $\beta$ 1,4-xylanase, xyloglucanase,  $\beta$ 1,4-mannanase,  $\beta$ 1,6-galactanase,  $\beta$ 1,3-glucosidase, and  $\beta$ -fucosidase [26]. To determine if the superiority of StCel5A might be related to substrate specificity, it was compared against TrCel5A on an assortment of polysaccharide substrates.

Both enzymes were active on lichenan and  $\beta$ -glucan (mixed-linked glucan), and glucomannan, and slightly active on laminarin ( $\beta$ 1,3-glucan), crystalline cellulose (SigmaCell), and PASC (Fig. 8). Note that “glucomannan” is a mainly unbranched polymer with alternating  $\beta$ 1,4-linked Glc and Man within the backbone, whereas “galactomannan” has a  $\beta$ 1,4-linked mannan backbone with  $\alpha$ -linked Gal side chains. StCel5A but not TrCel5A showed activity on the two galactomannans (locust bean gum and guar gum) and on the two xylans, one of which is substituted (arabinoxylan) and one which is not (xylan) (Fig. 8). Neither enzyme was active on xyloglucan. Therefore, StCel5A and TrCel5A differ qualitatively in their substrate specificities, and this might be a clue to their differential effectiveness at releasing Glc and Xyl from a complex biomass substrate in combination with other cell-wall active enzymes (Figs. 4–6).

We attempted to measure StCel5A-catalyzed release of Man from corn stover in an optimized 9-component enzyme cocktail. No Man could be detected (data not shown).

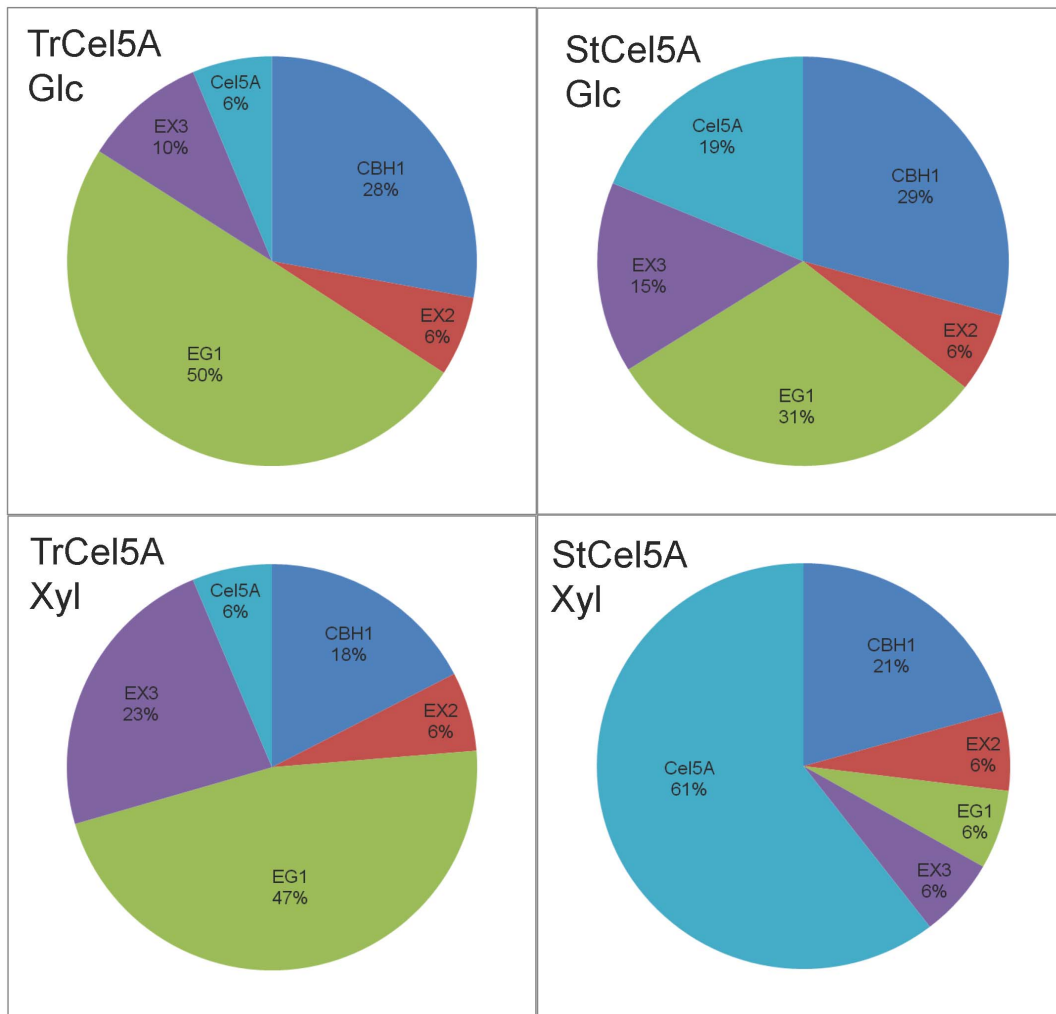
### Phylogenetic analysis of StCel5A, TrCel5A, and related GH5 cellulases

To explore further the reasons for superior performance of StCel5A and the possible involvement of substrate specificity, additional fungal members of GH5 were analyzed. By the classification of Aspeborg et al. [27], StCel5A and TrCel5A are both in subfamily 5 of GH5, denoted herein as GH5\_5. Both have canonical CBM1 modules and associated linker regions following the signal peptides at the N termini. The full-length proteins share 35.3% amino acid identity and the catalytic domains share 32.3% identity. In the tree of GH5\_5, there are 17 nodes separating StCel5A and TrCel5A [27]. All of the intervening proteins are of fungal origin with the exception of a single plant protein (BAK01092), which is probably an artifact. All of these proteins are known or predicted to belong to EC 3.2.1.4.

A phylogenetic tree of just the fungal sequences in GH5\_5 was constructed. Within GH5\_5, CBM's are common but not universal, and some have C-terminal CBM's. Additional fungal members of GH5\_5 that were present in the JGI database but not CAZy or GenBank (as of August, 2013), and therefore not in Aspeborg et al. [27], were added. On the other hand, not all of the sequences between StCel5A and TrCel5A in the GH5\_5 tree nor all of the additional high scoring sequences in GenBank and JGI were included. Duplicates (i.e., sharing >95% amino acid identity) were excluded, and only proteins with N-terminal CBM's, like StCel5A and TrCel5A themselves, were included to remove this feature as a potentially confounding variable. The exception to this exclusion rule was TaCel5A from *Thermoascus aurantiacus*, which lacks a native CBM but was nonetheless included because it has a solved crystal structure (GenBank accession AAL88714, PDB ID 1GZJ) [28,29].

Before alignment, the signal peptides, CBM's, and linkers (amounting to ~95 amino acids) were removed, on the assumption that the basis of enzymatic superiority resides in the catalytic domain. The alignment of the resulting protein set, composed of 30 proteins of broad taxonomic distribution within the higher fungi, revealed two subfamilies within GH5\_5, which we herein designate GH5\_5\_1 and GH5\_5\_2 (Fig. 9). TrCel5A falls into

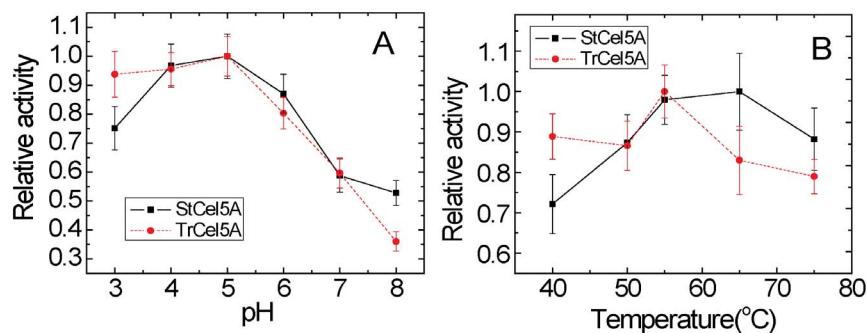




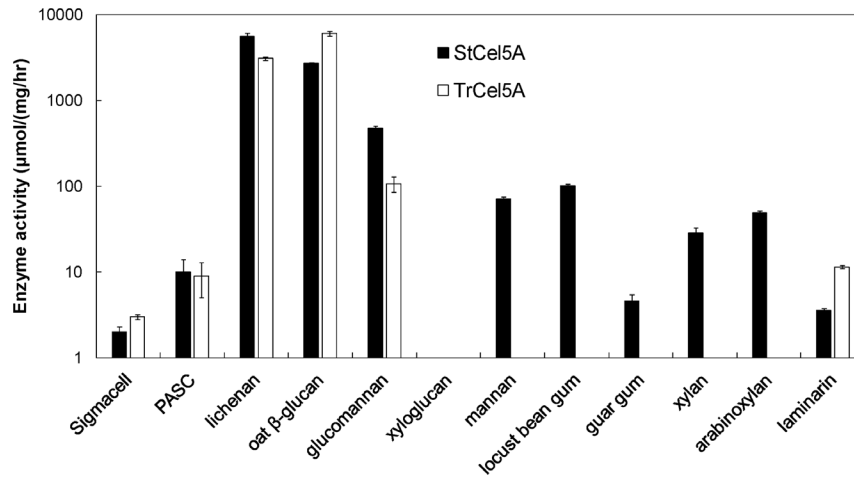
**Figure 6. Optimal enzyme proportions for release of Glc or Xyl from corn stover after 72 h digestion with either TrCel5A or StCel5A.** The other enzymes in the experiment (AA9, CBH2, BG, and BX) were fixed at 5% each. Glc and Xyl yields were 58% and 31% with TrCel5A and 70% and 39% with StCel5A, respectively. The data and statistical analyses are shown in Tables S3–S5. doi:10.1371/journal.pone.0109885.g006

family GH5\_5\_1 whereas StCel5A falls into subfamily GH5\_5\_2 (Fig. 9). The separation of the fungal members of GH5\_5 into two clades is also apparent in the phylogenetic tree of Aspeborg et al. [27], and all of the sequences in common between the two trees are congruent in regard to their membership in the two GH5\_5

subfamilies. Examples from both Ascomycota and Basidiomycota are present in both GH5\_5\_1 and GH5\_5\_2, which suggests that the resolution into these two subfamilies reflects a functional specialization that predates the split between these two phyla, and not merely the taxonomic relatedness of the source organisms.



**Figure 7. Hydrolytic activities of StCel5A or TrCel5A with lichenan as a function of (A) pH or (B) temperature.** doi:10.1371/journal.pone.0109885.g007



**Figure 8. Activities of pure TrCel5A and StCel5A on defined substrates.** Sigmacell is crystalline cellulose; PASC is phosphoric-acid swollen cellulose; lichenan and  $\beta$ -glucan are two types of mixed-linkage  $\beta$ 1,3- $\beta$ 1,4 glucan, one from Iceland moss and the other from oats; xyloglucan is  $\beta$ 1,4-linked glucan with  $\alpha$ 1,6-linked Xyl side chains from tamarind; mannan is insoluble  $\beta$ 1,4-mannan; locust bean gum and guar gum are two types of  $\beta$ 1,4-linked mannan with side chains of  $\alpha$ 1,6-linked galactose; xylan is  $\beta$ 1,4-linked xylan; arabinoxylan is  $\beta$ 1,4-linked xylan with  $\alpha$ 1,6-linked arabinose side chains; laminarin is predominantly  $\beta$ 1,3-linked glucan. Shaded bars, StCel5A; open bars, TrCel5A. TrCel5A had no detectable activity against mannan, either galactomannan, xylan, or arabinoxylan. Neither enzyme had detectable activity against xyloglucan. doi:10.1371/journal.pone.0109885.g008

In regard to whether GH5\_5 subfamily is correlated with enzymatic superiority and/or substrate specificity, biochemical information is not available for most of the proteins in GH5\_5. However, TaCel5A, which shares 65% amino acid identity with StCel5A and 34% identity with TrCel5A, does not use xylan as a substrate, unlike StCel5A [30,31]. EG1 of *Volvariella volvacea* (accession ACE06751), which is also in GH5\_5\_2, is also inactive on xylan [32]. Based on these two counter-examples, the trait of substrate specificity does not strictly correlate with the subfamilies shown in Fig. 9. It therefore remains to be determined whether the trait of superiority, as defined by our particular assays, correlates with subfamily within GH5\_5.

#### Comparisons of the known and deduced structures of TrCel5A, StCel5A, and TaCel5A

Conservation of amino acids were compared between StCel5A, TrCel5A, and other members of GH5\_5. The following amino acids were 100% conserved in all of the sequences (numbering is based on the Met start of TrCel5A): Glu259 and Glu148 (active site nucleophile and general base, respectively), Arg60, His104, Asn105, Phe124, Asn147, His218, Tyr220, and Trp292. These are all known or thought to be important for catalysis and/or substrate binding in GH5 [28,29,33,34].

The tertiary structure of StCel5A was modeled on the known structure of TrCel5A (Fig. 10). When superimposed on TrCel5A, the RMSD of the model calculated using the Align function in PyMol was 0.950 Å. TrCel5A has three loops that are predicted to be absent from StCel5A. Whereas loops 2 and 3 are far from the active site, loop 1 is predicted to be near the active site cleft (Fig. 10). This loop is also missing in TaCel5A (PDB ID: 1GZJ). The absence of this loop in both TaCel5A and StCel5A is consistent with their classification in subfamily 2 of GH5\_5 (Fig. 9). However, the two Cys residues that form a disulfide bond in loop 1 of TrCel5A are present in some but not all members of the GH5\_5\_1 family.

Potentially significant amino acids that are conserved within a subfamily, but not between subfamilies, include Gly189, which is always Trp in family 2 but Gly, Ala, or Glu in family 1; Gly291,

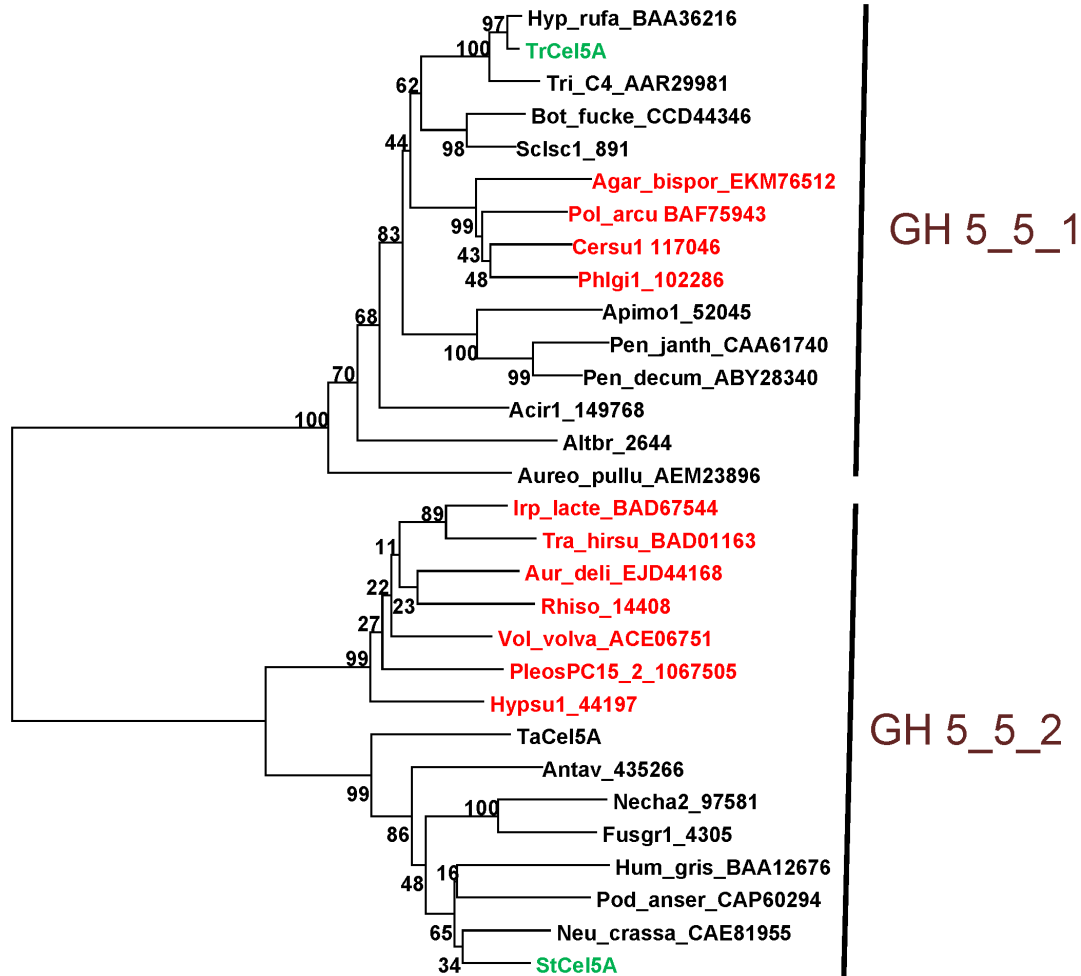
which is always Trp in family 2 but Gly or Ser in family 1; Phe297, which is always Trp in family 2 and always Phe in family 1; and Asp298, which is always Trp in family 2 but never in family 1 (the most common substitutions are Asp or Gln). That is, four Trp residues are conserved in GH5\_5\_2 but not GH5\_5\_1. This might be significant in regard to the different behaviors (superiority and substrate range) of StCel5A and TrCel5A because Trp residues are involved in substrate binding through ring stacking in many glycosyl hydrolases [35]. For example, Trp170, Trp174, Trp273, Trp 278, and Trp279 line the binding cleft of TaCel5A [28,29]. The corresponding amino acids in StCel5A are Trp169, Trp173, Trp272, Trp277, and Trp278, and in TrCel5A the corresponding amino acids are Trp185, Gly189, Trp292, Phe297, and Asp298. The conserved and subfamily-unique Trp residues in the binding cleft of GH5\_5\_1 and GH 5\_5\_2 are illustrated in Fig. 11.

#### Discussion

In this paper, we report the re-identification of an endo- $\beta$ 1,4-glucanase from *S. thermophile* based on its synergistic activity at releasing fermentable sugars when tested in combination with a core set of eight enzymes from *T. reesei* on a realistic biomass substrate. In side-by-side comparisons, StCel5A was consistently superior at releasing both Glc and Xyl. Optimization experiments suggested that superiority was related to xylanase activity, and StCel5A, but not TrCel5A, was shown to utilize xylan as a substrate.

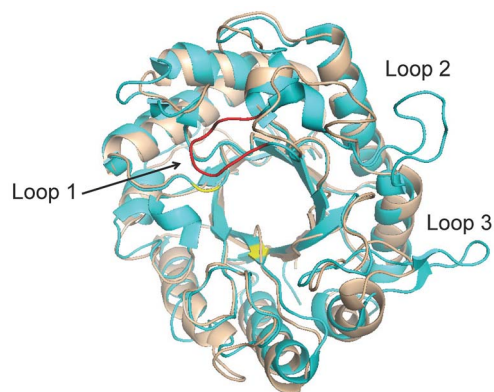
There are many previous reports of synergism between the cell-wall active enzymes of different micro-organisms [7,11,36], and this type of synergism can have multiple explanations. For example, some micro-organisms make enzymes that others do not [37–40], and enzymes from different sources with nominally the same catalytic activity can have different and superior pH optima, thermal stability, or resistance to denaturation.

The current work suggests that altered substrate specificity might be an additional contributor to synergistic activity. The involvement of broader substrate specificity in the superiority of StCel5A over TrCel5A is supported by the optimization experiments, in which the optimal proportions of the nine component



**Figure 9. Alignment of the catalytic domains of fungal proteins in subfamily 5 of GH5 (GH5\_5) showing its further division into two subclades, GH5\_5\_1 and GH5\_5\_2.** All of the sequences have N-terminal CBM's except TaCel5A. StCel5A and TrCel5A are shown in green. Sequences from Basidiomycota are shown in red; all other sequences are from Ascomycota. Species abbreviations and taxonomic affiliations are shown in Table S6.

doi:10.1371/journal.pone.0109885.g009

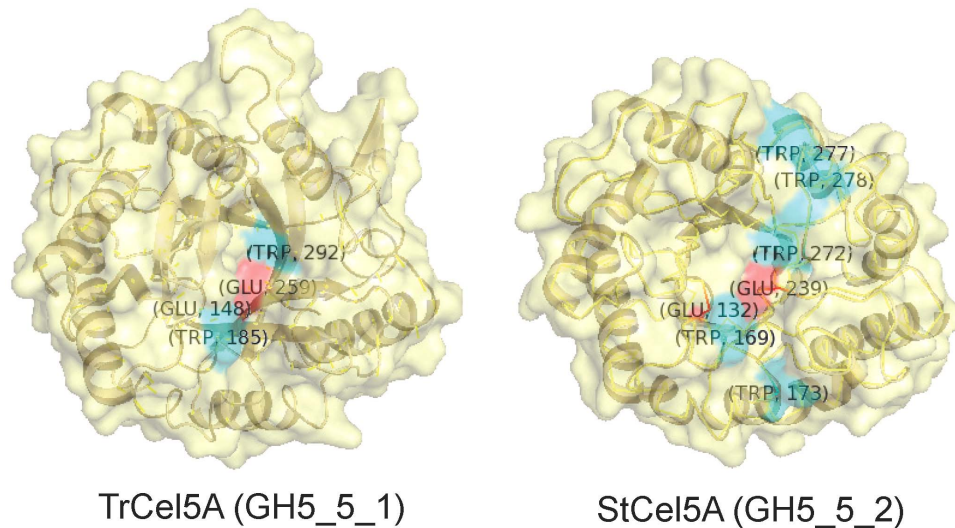


**Figure 10. Modeled structure of StCel5A (grey) based on the structure of TaCel5A (PDB ID: 1GZJ) superimposed on TrCel5A (PDB ID: 3QR3, shown in blue).** The two active-site Glu residues are shown in yellow. Loops 1, 2, and 3 are present in TrCel5A but not StCel5A; loop1 is shown in red.

doi:10.1371/journal.pone.0109885.g010

enzymes was dependent on whether TrCel5A or StCel5A were included in the mixtures. The difference in optimal proportions was especially pronounced for yields of Xyl compared to Glc. All of our enzyme mixtures contained endo- $\beta$ 1,4-xylanases of GH10 (EX3) and GH11 (EX2), but xylanases in GH5 are known to have a different, generally broader, substrate range than either of these canonical xylanase families. Notably, GH5 xylanases have the ability to hydrolyze xylans that are substituted with glucuronic acid [26]. Therefore, StCel5A, which has xylanase activity in addition to  $\beta$ 1,4-glucanase activity, may be superior to TrCel5A for Xyl release because it can degrade xylans that are not substrates for either EX2 or EX3 [41–43]. The xylanase activity of StCel5A might also explain why StCel5A also promotes Glc yields, because xylanases are known to enhance the accessibility of cellulose to cellulases [11]. StCel5A also had better activity than TrCel5A against various  $\beta$ 1,4-mannans, including two types of galactomannan and unsubstituted mannan (Fig. 8). Because corn stover contains small amounts of galactose or Man compared to Glc and Xyl [3], it is unlikely that mannanase activity instead of xylanase activity accounts for the superiority of StCel5A, and, in fact, no release of Man from corn stover could be measured with StCel5A. Furthermore, we earlier showed in a 16-component optimization





TrCel5A (GH5\_5\_1)

StCel5A (GH5\_5\_2)

**Figure 11. Surface views of StCel5A and TrCel5A showing Trp residues (blue) that are conserved in all members of GH5\_5\_1 (e.g., TrCel5A) or GH5\_5\_2 (e.g., StCel5A).** The active-site Glu residues are shown in red. For orientation, the loop at the top of TrCel5A is the loop 2 shown in Fig. 10.

doi:10.1371/journal.pone.0109885.g011

experiment that  $\beta$ -mannanase makes no contribution to Glc or Xyl release from pretreated corn stover (although it was one of the most important enzymes for release of Glc from dried distillers' grains) [3].

A structural understanding of the basis of substrate specificity in glycosyl hydrolases has been challenging. Chen et al. [44] analyzed the basis of substrate discrimination between glucan and galactomannan across the entire GH5 family. They identified a motif of six amino acids that when mutated altered substrate specificity, using Cel5A from the bacterium *Thermotoga maritima* (TmCel5A, AAD36816) as the model. TmCel5A is in subfamily 25 of GH5 [27] and can hydrolyze both cellulose and galactomannan, like StCel5A. Four of the six critical amino acids identified by Chen et al. [44] could be unambiguously identified in StCel5A and TrCel5A despite low overall (<20%) amino acid identity. Among these six critical amino acids, His95 (TmCel5A numbering) was invariant in all three sequences, while the other residues varied but in a pattern that did not correlate with the known substrate specificity of the three enzymes. For example, in TmCel5A, TrCel5A, or StCel5A, respectively, amino acid 53 was Pro, Pro, or Asp; amino acid 96 was His, Asn, or Asn; and amino acid 287 was Asp, Ala, or Gly. Thus, substrate discrimination appears to follow different rules in subfamily 5 of GH5 than in subfamily 25.

The superiority of StCel5A over TrCel5A might also be due to features of the two enzymes that are unrelated to substrate specificity. This is supported by the fact that two other members of GH5\_5\_2 are not active against xylan, unlike StCel5A. Another consideration is that the critical feature(s) of StCel5A that cause superiority might manifest themselves only in combination with other necessary biomass-degrading enzymes (i.e., the 8-component synthetic mixture) and a real biomass substrate (i.e., pretreated corn stover). Therefore, the critical differences might not be observable in any experiments using pure individual enzymes and pure substrates, including the enzyme kinetic [14] and substrate specificity experiments (Fig. 8). Potentially critical properties of an enzyme that might be apparent in experiments involving enzyme mixtures and real biomass but not in isolation on pure substrates include nonproductive binding to one or more biomass compo-

nents such as lignin, enhanced resistance to inhibitors in biomass, or inhibition by the products of other enzymes.

Further studies, including biochemical analysis of additional members of GH5\_5, should resolve the structural features that are responsible for superiority, substrate specificity, and phylogenetic relatedness within this subfamily.

## Supporting Information

**Table S1 Design and results of first optimization experiment.**

(DOCX)

**Table S2 Design and results of first optimization experiment.**

(DOCX)

**Table S3 Glc yields (g/L) in the second optimization experiment based on the experimental design shown in Table S2.**

(DOCX)

**Table S4 Xyl yields (g/L) from the second optimization experiment.**

(DOCX)

**Table S5 Statistical analyses of the optimization experiments shown in Tables S3 and S4 as generated by DesignExpert.**

(DOCX)

**Table S6 Species abbreviations and taxonomic affiliations for the fungi shown in Fig. 9.**

(DOCX)

## Acknowledgments

We thank Dyadic, Inc., for supplying the CMAX; Lucigen, Inc., for supplying enzymes; and Doug Whitten of the MSU Research Technology Support Facility for the proteomics analyses. The authors thank Drs. Taichi E. Takasuka, Christopher M. Bianchetti, and Brian G. Fox (GLBRC and Department of Biochemistry, University of Wisconsin-Madison) for performing an earlier version of the assays shown in Fig. 8

that established the broad substrate specificity of StCel5A, and for their input on the possible role of broadened substrate specificity in identifying superior enzymes for biomass deconstruction.

## References

- Yang B, Dai Z, Ding S-Y, Wyman CE (2011) Enzymatic hydrolysis of cellulosic biomass. *Biofuels* 2: 421–450.
- Mansfield SD, Mooney C, Saddler JN (1999) Substrate and enzyme characteristics that limit cellulose hydrolysis. *Biotechnol Prog* 15: 804–816.
- Banerjee G, Car S, Scott-Craig JS, Borrusch MS, Walton JD (2010) Rapid optimization of enzyme mixtures for deconstruction of diverse pretreatment/biomass feedstock combinations. *Biotechnol Biofuels* 3: 22.
- Nagendran S, Hallen-Adams HE, Paper JM, Aslam N, Walton JD (2009) Reduced genomic potential for secreted plant cell-wall-degrading enzymes in the ectomycorrhizal fungus *Amanita bisporigera*, based on the secretome of *Trichoderma reesei*. *Fung Genet Biol* 46: 427–435.
- Banerjee G, Car S, Scott-Craig JS, Borrusch MS, Bongers M, et al. (2010) Synthetic multi-component enzyme mixtures for deconstruction of lignocellulosic biomass. *Bioresour Technol* 101: 9097–9105.
- Gao D, Chundawat SP, Liu T, Hermanson S, Gowda K, et al. (2010) Strategy for identification of novel fungal and bacterial glycosyl hydrolase hybrid mixtures that can efficiently saccharify pretreated lignocellulosic biomass. *BioEnergy Res* 3: 67–81.
- Berlin A, Maximenko V, Gilkes N, Saddler J (2007) Optimization of enzyme complexes for lignocellulose hydrolysis. *Biotechnol Bioeng* 97: 287–296.
- Kim E, Irwin DC, Walker LP, Wilson DB (1998) Factorial optimization of a six-cellulase mixture. *Biotechnol Bioengineer* 58: 494–501.
- Falls M, Shi J, Ebrink MA, Redmond T, Yang B, et al. (2011) Investigation of enzyme formulation on pretreated switchgrass. *Bioresour Technol* 102: 11072–11079.
- Kostylev M, Wilson D (2012) Synergistic interactions in cellulose hydrolysis. *Biofuels* 3: 61–70.
- Selig MJ, Knoshaug EP, Adney WS, Himmel ME, Decker SR (2008) Synergistic enhancement of cellobiohydrolase performance on pretreated corn stover by addition of xylanase and esterase activities. *Bioresour Technol* 99: 4997–5005.
- Banerjee G, Car S, Scott-Craig JS, Hodge DB, Walton JD (2011) Alkaline peroxide pretreatment of corn stover: effects of biomass, peroxide, and enzyme loading and composition on yields of glucose and xylose. *Biotechnol Biofuels* 4: 1–15.
- Banerjee G, Car S, Scott-Craig JS, Borrusch MS, Aslam N, et al. (2010) Synthetic enzyme mixtures for biomass deconstruction: production and optimization of a core set. *Biotechnol Bioengineer* 106: 707–720.
- Tambor JH, Ren H, Ushinsky S, Zheng Y, Riemens A, et al. (2012) Recombinant expression, activity screening and functional characterization identifies three novel endo-1,4- $\beta$ -glucanases that efficiently hydrolyse cellulosic substrates. *Appl Microbiol Biotechnol* 93: 203–214.
- Takasuka TE, Walker JA, Bergeman LF, Vander Meulen KA, Makino SI, et al. (2012) Cell-free translation of biofuels enzymes. In: *Methods in Molecular Biology*, in press. Humana, Totowa, NJ.
- Weimer PJ, Lopez-Guisa JM, French AD (1990) Effect of cellulose fine structure on kinetics of its digestion by mixed ruminal microorganisms in vitro. *Appl Env Microbiol* 56: 2421–2429.
- Miller GL (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Chem* 31: 426–428.
- Lever M (1972) A new reaction for colorimetric determination of carbohydrates. *Anal Biochem* 47: 273–279.
- Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248–254.
- Tamura K, Stecher G, Peterson D, Filipiski A, Kumar S (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30: 2725–2729.
- Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195–201.
- Berka RM, Grigoriev IV, Ottillar R, Salamov A, Grimwood J, et al. (2011) Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*. *Nat Biotechnol* 29: 922–927.
- Zhang X, Patil S, Yang J, Dhawan IK (2012) Endoglucanase variants. US Patent Application 0208235 A1.
- Gusakov AV, Salanovich TN, Antonov AI, Ustinov BB, Okunev ON, et al. (2011) Construction of highly efficient cellulase compositions for enzymatic hydrolysis of cellulose. US Patent 7,883,872 B2.
- Henrissat B, Teeri TT, Warren RAJ (1998) A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. *FEBS Lett* 425: 352–354.
- Pollet A, Delcour JA, Courtin CM (2010) Structural determinants of the substrate specificities of xylanases from different glycoside hydrolase families. *Crit Rev Biotechnol* 30: 176–191.
- Aspeborg H, Coutinho PM, Wang Y, Brumer H 3rd, Henrissat B (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol Biol* 12: 186.
- Lo Leggio L, Larsen S (2002) The 1.62 Å structure of *Thermoascus aurantiacus* endoglucanase: completing the structural picture of subfamilies in glycoside hydrolase family 5. *FEBS Lett* 523: 103–108.
- Van Petegem F, Vandenberghe I, Bhat MK, Van Beeumen J (2002) Atomic resolution structure of the major endoglucanase from *Thermoascus aurantiacus*. *Biochem Biophys Res Comm* 296: 161–166.
- Srikrishnan S, Randall A, Baldi P, Da Silva NA (2012) Rationally selected single-site mutants of the *Thermoascus aurantiacus* endoglucanase increase hydrolytic activity on cellulosic substrates. *Biotechnol Bioengineer* 109: 1595–1599.
- Shepherd MG, Tong CC, Cole AL (1981) Substrate specificity and mode of action of the cellulases from the thermophilic fungus *Thermoascus aurantiacus*. *Biochem J* 193: 67–74. Lee TM, Farrow MF, Arnold FH, Mayo SL (2011) A structural study of *Hypocrea jecorina* Cel5A. *Protein Sci* 20: 1935–1940.
- Ding S-J, Ge W, Buswell JA (2001) Endoglucanase I from the edible straw mushroom, *Volvariella volvacea*. *Eur J Biochem* 268: 5687–5695.
- Lee TM, Farrow MF, Arnold FH, Mayo SL (2011) A structural study of *Hypocrea jecorina* Cel5A. *Protein Sci* 20: 1935–1940.
- Zechel DL, Withers SG (2000) Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Chem Res* 33: 11–18.
- Boraston AB, Bolam DN, Gilbert HJ, Davies GJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* 382: 769–781.
- Hu J, Arantes V, Pribowo A, Saddler JN (2013) The synergistic action of accessory enzymes enhances the hydrolytic potential of a “cellulase mixture” but is highly substrate specific. *Biotechnol Biofuels* 6: 112.
- Scott-Craig JS, Borrusch MS, Banerjee G, Harvey CM, Walton JD (2011) Biochemical and molecular characterization of secreted  $\alpha$ -xylosidase from *Aspergillus niger*. *J Biol Chem* 286: 42848–42854.
- Eastwood DC, Floudas D, Binder M, Majcherczyk A, Schneider P, et al. (2011) The plant cell wall-decomposing machinery underlies the functional diversity of forest fungi. *Science* 333: 762–765.
- Floudas D, Binder M, Riley R, Barry K, Blanchette RA, et al. (2012) The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* 336: 1715–1719.
- Hori C, Gaskell J, Igarashi K, Samejima M, Hibbett D, et al. (2013) Genomewide analysis of polysaccharides degrading enzymes in 11 white- and brown-rot Polyporales provides insight into mechanisms of wood decay. *Mycologia* 105: 1412–1427.
- Cruys-Bagger N, Elmerdahl J, Praestgaard E, Tatsumi H, Spodsberg N, et al. (2012) Pre-steady-state kinetics for hydrolysis of insoluble cellulose by cellobiohydrolase Cel7A. *J Biol Chem* 287: 18451–18458.
- Fox JM, Levine SE, Clark DS, Blanch HW (2011) Initial- and processive-cut products reveal cellobiohydrolase rate limitations and the role of companion enzymes. *Biochemistry* 51: 442–452.
- Ye Z, Berson RE (2011) Kinetic modeling of cellulose hydrolysis with first order inactivation of adsorbed cellulase. *Bioresour Technol* 102: 11194–11199.
- Chen Z, Friedland GD, Pereira JH, Revoco SA, Chan R, et al. (2012) Tracing determinants of dual substrate specificity in glycoside hydrolase family 5. *J Biol Chem* 287: 25335–25343.

## Author Contributions

Conceived and designed the experiments: ZY YZ MSB RS JDW. Performed the experiments: ZY YZ MSB RS JDW BL. Analyzed the data: ZY MSB JDW BL. Contributed reagents/materials/analysis tools: ZY YZ RS. Wrote the paper: ZY YZ MSB RS JDW.