A family of thermostable fungal cellulases created by structure-guided recombination

Pete Heinzelman[a], Christopher D. Snow[a], Indira Wu[a], Catherine Nguyen[a], Alan Villalobos[b], Sridhar Govindarajan[b], Jeremy Minshull[b], and Frances H. Arnold[a,1]

[a]Division of Chemistry and Chemical Engineering 210-41, California Institute of Technology, Pasadena, CA 91125; and [b]DNA2.0, Menlo Park, CA 94025

Contributed by Frances H. Arnold, February 9, 2009 (sent for review January 19, 2009)

SCHEMA structure-guided recombination of 3 fungal class II cellobiohydrolases (CBH II cellulases) has yielded a collection of highly thermostable CBH II chimeras. Twenty-three of 48 genes sampled from the 6,561 possible chimeric sequences were secreted by the *Saccharomyces cerevisiae* heterologous host in catalytically active form. Five of these chimeras have half-lives of thermal inactivation at 63 °C that are greater than the most stable parent, CBH II enzyme from the thermophilic fungus *Humicola insolens*, which suggests that this chimera collection contains hundreds of highly stable cellulases. Twenty-five new sequences were designed based on mathematical modeling of the thermostabilities for the first set of chimeras. Ten of these sequences were expressed in active form; all 10 retained more activity than *H. insolens* CBH II after incubation at 63 °C. The total of 15 validated thermostable CBH II enzymes have high sequence diversity, differing from their closest natural homologs at up to 63 amino acid positions. Selected purified thermostable chimeras hydrolyzed phosphoric acid swollen cellulose at temperatures 7 to 15 °C higher than the parent enzymes. These chimeras also hydrolyzed as much or more cellulose than the parent CBH II enzymes in long-time cellulose hydrolysis assays and had pH/activity profiles as broad, or broader than, the parent enzymes. Generating this group of diverse, thermostable fungal CBH II chimeras is the first step in building an inventory of stable cellulases from which optimized enzyme mixtures for biomass conversion can be formulated.

biofuels | cellobiohydrolase | cellulose hydrolysis | *Trichoderma reesei* | CBH II

The performance of cellulase mixtures in biomass conversion processes depends on many enzyme properties including stability, product inhibition, synergy among different cellulase components, productive binding versus nonproductive adsorption and pH dependence, in addition to the cellulose substrate physical state and composition. Given the multivariate nature of cellulose hydrolysis, it is desirable to have diverse cellulases to choose from to optimize enzyme formulations for different applications and feedstocks. Recent studies have documented the superior performance of cellulases from thermophilic fungi relative to their mesophilic counterparts in laboratory scale biomass conversion processes (1, 2), where enhanced stability leads to retention of activity over longer periods of time at both moderate and elevated temperatures. Fungal cellulases are attractive because they are highly active and can be expressed in fungal hosts such as *Hypocrea jecorina* (anamorph *Trichoderma reesei*) at levels up to 40 g/L in the supernatant. Unfortunately, the set of documented thermostable fungal cellulases is small. In the case of the processive cellobiohydrolase class II (CBH II) enzymes, <10 natural thermostable gene sequences are annotated in the CAZy database (www.cazy.org). This limited number, combined with the difficulty of using directed evolution to generate diverse thermostable cellulases, motivated our application of SCHEMA structure-guided protein recombination (3) to the challenge of creating thermostable fungal CBH II enzymes.

SCHEMA has been used to create families of hundreds of active β-lactamase (4) and cytochrome P450 (5) enzyme chimeras. SCHEMA uses protein structure data to define boundaries of contiguous amino acid "blocks" that minimize $\langle E \rangle$, the library average number of amino acid sidechain contacts that are broken when the blocks are swapped among different parents. Meyer et al. (4) found that the probability that a β-lactamase chimera was folded and active was inversely related to the value of E for that sequence. The RASPP (Recombination as Shortest Path Problem) algorithm (6) was used to identify the block boundaries that minimized $\langle E \rangle$ relative to the library average number of mutations, $\langle m \rangle$ (4). More than 20% of the ≈500 unique chimeras characterized from a β-lactamase collection comprised of 8 blocks from 3 parents (3^8 = 6,561 possible sequences) were catalytically active (4). A similar approach produced a 3-parent, 8-block cytochrome P450 chimera family containing >2,300 catalytically active enzymes (5). Chimeras from these 2 collections were characterized by high numbers of mutations, 66 and 72 amino acids on average from the closest parent, respectively. SCHEMA/RASPP thus enabled design of chimera families having significant sequence diversity and an appreciable fraction of functional members.

It has also been shown that the thermostabilities of SCHEMA chimeras can be predicted based on sequence-stability data from a small sample of the sequences (6). Linear regression modeling of thermal inactivation data for 184 cytochrome P450 chimeras showed that SCHEMA blocks made additive contributions to thermostability. More than 300 chimeras were predicted to be thermostable by this model, and all 44 that were tested were more stable than the most stable parent. It was estimated that as few as 35 thermostability measurements could be used to predict the most thermostable chimeras. Furthermore, the thermostable P450 chimeras displayed unique activity and specificity profiles, demonstrating that chimeragenesis can lead to additional useful enzyme properties. Here, we show that SCHEMA recombination of CBH II enzymes can generate chimeric cellulases that are active on phosphoric acid swollen cellulose (PASC) at high temperatures, over extended periods of time, and with broad ranges of pH.

Results

Five candidate parent genes encoding CBH II enzymes were synthesized with *S. cerevisiae* codon bias. All 5 contained identical N-terminal coding sequences, where residues 1–89 correspond to the cellulose binding module (CBM), flexible linker region and the 5 N-terminal residues of the *H. jecorina* catalytic domain. Two of the candidate CBH II enzymes, from *Humicola insolens* and *Chaetomium thermophilum*, were secreted from *S.*

Author contributions: P.H., C.D.S., J.M., and F.H.A. designed research; P.H., I.W., and C.N. performed research; P.H., C.D.S., I.W., A.V., S.G., J.M., and F.H.A. analyzed data; and P.H., C.D.S., J.M., and F.H.A. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: frances@cheme.caltech.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0901417106/DCSupplemental.

BIOCHEMISTRY

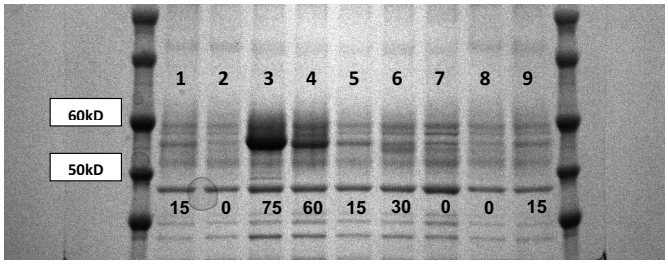

Fig. 1. SDS/PAGE gel of candidate CBH II parent gene yeast expression culture supernatants. Gel Lanes (left to right): 1, *H. jecorina*; 2, empty vector; 3, *H. insolens*; 4, *C. thermophilum*; 5, *H. jecorina* (duplicate); 6, *P. chrysosporium*; 7, *T. emersonii*; 8, empty vector (duplicate); 9, *H. jecorina* (triplicate). Numbers at bottom of gel represent concentration of reducing sugar (μg/mL) present in reaction after a 2-h 50 °C PASC hydrolysis assay. Subsequent SDS/PAGE comparison with BSA standard allowed estimation of *H. insolens* expression level of 5–10 mg/L.

cerevisiae at much higher levels than the other 3, from *Hypocrea jecorina*, *Phanerochaete chrysosporium* and *Talaromyces emersonii* (Fig. 1). Because bands in the SDS/PAGE gel for the 3 weakly expressed candidate parents were difficult to discern, activity assays in which concentrated culture supernatants were incubated with phosphoric acid swollen cellulose (PASC) were performed to confirm the presence of active cellulase. The values for the reducing sugar formed, presented in Fig. 1, confirmed the presence of active CBH II in concentrated *S. cerevisiae* culture supernatants for all enzymes except *T. emersonii* CBH II. We chose to recombine the *H. insolens* and *C. thermophilum* sequences with the most industrially relevant fungal CBH II enzyme, from *H. jecorina*. The respective sequence identities of the catalytic domains are 64% (*1*:*2*), 66% (*2*:*3*) and 82% (*1*:*3*), where *H. insolens* is parent 1, *H. jecorina* is parent 2, and *C. thermophilum* is parent 3. These respective catalytic domains contain 360, 358, and 359 amino acid residues.

Heterologous protein expression in the filamentous fungus *H. jecorina*, the organism most frequently used to produce cellulases for industrial applications, is much more arduous than in *Saccharomyces cerevisiae*. The observed secretion of *H. jecorina* CBH II from *S. cerevisiae (7)* motivated our choice of this heterologous host. To minimize hyperglycosylation, which has been reported to reduce the activity of recombinant cellulases (8,9), we expressed the recombinant CBH II genes in a glycosylation-deficient Δ*kre2 S. cerevisiae* strain. This strain is expected to attach smaller mannose oligomers to both N-linked and O-linked glycosylation sites than wild type strains (10), which more closely resembles the glycosylation of natively produced *H. jecorina* CBH II enzyme (11). The observed ≈55-kDa molecular mass of the CBH II enzyme parents can be accounted for by the ≈47-kDa molecular mass of the nonglycosylated CBH II polypeptide chains and, for the Δ*kre2* strain, the anticipated N- and O-linked glycan molecular mass contribution of 8-10 kDa (11). For wild-type *S. cerevisiae* strains that attach high mannose moieties to N-linked sites in recombinant proteins, the N-linked molecular weight contribution alone would be >10 kD (12). As such, the observed molecular weights suggest that the recombinant CBH II enzymes do not feature hyperglycosylation at N-linked sites.

The high resolution structure of *H. insolens* (13) (PDB entry 1OCN) was used as a template for SCHEMA to identify contacts that could be broken upon recombination. RASPP returned 4 candidate libraries, each with $\langle E \rangle < 15$. The candidate libraries all have lower $\langle E \rangle$ than previously constructed chimera libraries, suggesting that an acceptable fraction of folded, active chimeras could be obtained for a relatively high $\langle m \rangle$. We maximized chimera sequence diversity by selecting the block boundaries leading to the greatest $\langle m \rangle = 50$. The blocks for this design are illustrated in Fig. 2*B* and detailed in Table S1.

The *H. insolens* CBH II catalytic domain has an α/β barrel structure in which the 8 helices define the barrel perimeter and 7 parallel β-sheets form the active site (Fig. 2*A*). Two extended loops form a roof over the active site, creating a tunnel through which the substrate cellulose chains pass during hydrolysis. Five of the 7 block boundaries fall between elements of secondary structure, whereas block 4 begins and ends in the middle of consecutive α-helices (Fig. 2*A* and *B*. The majority of interblock sidechain contacts occur between blocks that are adjacent in the primary structure (Fig. 2*C*).

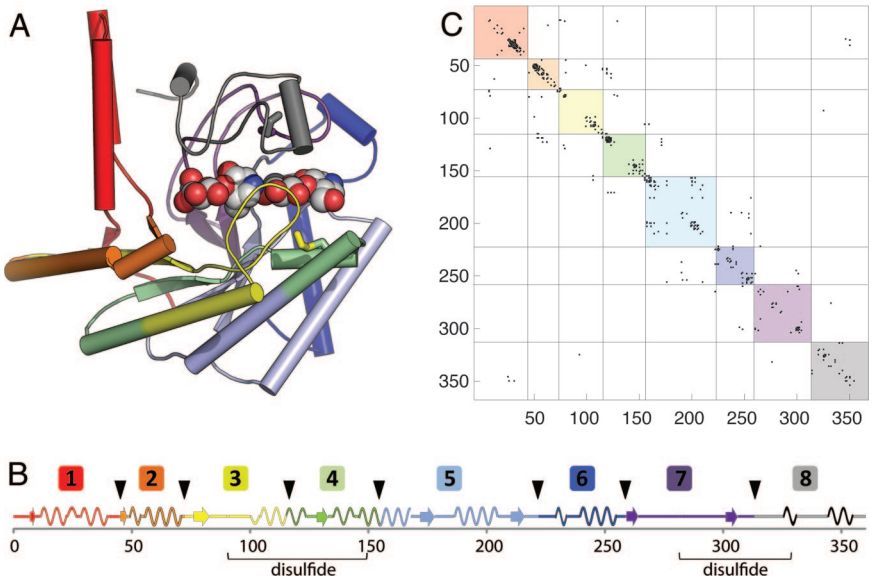

Fig. 2. Illustrations of CBH II chimera library block boundaries. (*A*) *H. insolens* CBH II catalytic domain ribbon diagram with blocks distinguished by color. CBH II enzyme is complexed with cellobio-derived isofagomine glycosidase inhibitor (13). (*B*) Linear representation of *H. insolens* catalytic domain showing secondary structure elements, disulfide bonds and block divisions denoted by black arrows. (*C*) Sidechain contact map denoting contacts (side chain heavy atoms within 4.5Å) that can be broken upon recombination. The majority of broken contacts occur between consecutive blocks.

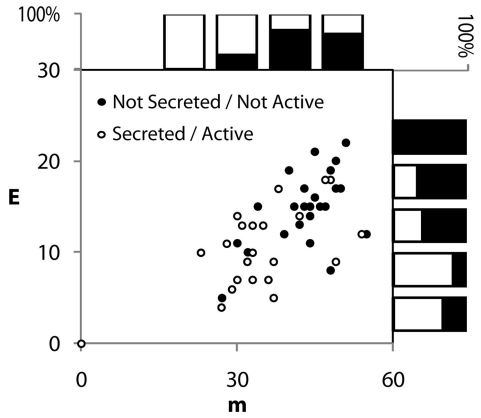

Fig. 3. Number of broken contacts (*E*) and number of mutations from closest parent (*m*) for 23 secreted/active and 15 not secreted/not active sample set chimeras.

A sample set of 48 chimera genes was designed as 3 sets of 16 chimeras having 5 blocks from one parent and 3 blocks from either one or both of the remaining 2 parents (Table S2); the sequences were selected to equalize the representation of each parent at each block position. The corresponding genes were synthesized and expressed.

Twenty-three of the 48 sample set *S. cerevisiae* concentrated culture supernatants exhibited hydrolytic activity toward PASC. These results suggest that thousands of the 6,561 possible CBH II chimera sequences encode active enzymes. The 23 active CBH II sample set chimeras show considerable sequence diversity, differing from the closest parental sequence and each other by at least 23 and 36 amino acid substitutions and as many as 54 and 123, respectively. Their average mutation level $\langle m \rangle$ is 36.

Because Meyer et al. (4) found correlations between E, m, and the probability that a chimera is folded and active, we sought to determine whether similar correlations existed for the sample set CBH II chimeras. The amount of CBH II enzyme activity in concentrated expression culture supernatants, as measured by assaying for activity on PASC, was correlated to the intensity of CBH II bands in SDS/PAGE gels (Fig. 1). As with the *H. jecorina* CBH II parent, activity could be detected for some CBH II chimeras with undetectable gel bands. There were no observations of CBH II chimeras presenting gel bands but lacking activity. The probability of a CBH II chimera being secreted in active form was inversely related to both E and m (Fig. 3).

Half-lives of thermal inactivation ($t_{1/2}$) were measured at 63 °C for concentrated culture supernatants of the parent and active chimeric CBH II enzymes. The *H. insolens*, *H. jecorina* and *C. thermophilum* CBH II parent half-lives were 95, 2 and 25 min, respectively (Table 1). The active sample set chimeras exhibited a broad range of half-lives, from <1 min to >3,000. Five of the 23 active chimeras had half-lives greater than that of the most thermostable parent, *H. insolens* CBH II.

In attempting to construct a predictive quantitative model for CBH II chimera half-life, we experimented with 5 different linear regression data modeling algorithms (Table S3) (14). Each algorithm was used to construct a model relating the block compositions of each sample set CBH II chimera and the parents to the $\log(t_{1/2})$. These models produced thermostability weight values that quantified a block's contribution to $\log(t_{1/2})$. For all 5 modeling algorithms, this process was repeated 1,000 times, with 2 randomly selected sequences omitted from each calculation, so that each algorithm produced 1,000 weight values for each of the 24 blocks. The mean and standard deviation (SD) were calculated for each block's thermostability weight. The predictive accuracy of each model algorithm was assessed by measuring how well each model predicted the $t_{1/2}$ values of the 2 omitted sequences. The correlation between measured and predicted values for the 1,000 algorithm iterations is the model algorithm's cross-validation score. For all 5 models, the cross-validation scores (X-val) were less than or equal to 0.57 (Table S3), indicating that linear regression modeling could not be applied to this small, 23 chimera $t_{1/2}$ dataset for quantitative CBH II chimera half-life prediction.

Table 1. Sequences and half-lives of thermal inactivation for active CBH II enzyme sample set chimeras at 63°C, pH 4.8

Chimera	$t_{1/2}$ (min)	$t_{1/2}$ (min)
H. insolens (P1)	90	100
H. jecorina (P2)	2	2
C. thermophilum (P3)	30	20
11113132	2,800	3,600
21333331	500	630
21311131	460	500
22232132	280	330
33133132	200	200
33213332	150	130
13333232	100	130
12133333	70	110
13231111	60	40
11313121	50	45
11332333	40	40
12213111	40	40
23311333	35	30
13111313	20	20
31311112	15	15
23231222	10	10
33123313	10	10
22212231	5	15
21223122	5	10
21131311	3	3
23233133	3	2
31212111	2	3
32333113	<1	<1

Results for two independent trials are shown.

We therefore used linear regression modeling to qualitatively classify blocks as stabilizing, destabilizing or neutral. Each block's impact on chimera thermostability was characterized using a scoring system that accounts for the thermostability contribution determined by each of the regression algorithms. For each algorithm, blocks with a thermostability weight value >1 SD above neutral were scored "+1," blocks within 1 SD of neutral were assigned zero and blocks 1 or more SD below neutral were scored "−1." A "stability score" for each block was obtained by summing the 1, 0, −1 stability scores from each of the 5 models. Table S4 summarizes the scores for each block. Block 1/parent 1 (B1P1), B6P3, B7P3, and B8P2 were identified as having the greatest stabilizing effects, whereas B1P3, B2P1, B3P2, B6P2, B7P1, B7P2, and B8P3 were found to be the most strongly destabilizing blocks.

We synthesized a second set of genes encoding CBH II enzyme chimeras to validate the predicted stabilizing blocks and identify cellulases more thermostable than the most stable parent. The 24 chimeras included in this validation set (Table 2) were devoid of the 7 blocks predicted to be most destabilizing and enriched in the 4 most stabilizing blocks, where represen-

Table 2. Sequences of 24 validation set CBH II enzyme chimeras, 9 of which were expressed in active form

Inactive	Active
12122132	12111131
12132332	12132331
12122331	12131331
12112132	12332331
13122332	13332331
13111132	13331332
13111332	13311331
13322332	13311332
22122132	22311331
22322132	
22311332	
23111332	
23321131	
23321332	
23321331	

tation was biased toward higher stability scores. Additionally, the "HJPlus" 12222332 chimera was constructed by substituting the predicted most stabilizing blocks into the *H. jecorina* CBH II enzyme (parent 2).

Concentrated supernatants of *S. cerevisiae* expression cultures for 9 of the 24 validation set chimeras, and the HJPlus chimera, showed activity toward PASC (Table 2). Of the 15 chimeras for which activity was not detected, 9 contained block B4P2. Of the 16 chimeras containing B4P2 in the initial sample set, only one showed activity toward PASC. Summed over both chimera sets and HJPlus, just 2 of 26 chimeras featuring B4P2 were active, indicating that this particular block is highly detrimental to expression of active cellulase in *S. cerevisiae*.

We evaluated the stabilities of the 10 functional chimeric CBH II enzymes from the validation set. Because the stable enzymes already had half-lives of >50 h, we chose residual hydrolytic activity toward PASC after a 12-h thermal inactivation at 63 °C as the metric for preliminary evaluation. This 12-h incubation produced a measurable decrease in the activity of the sample set's most thermostable chimera, 11113132, and completely inactivated the thermostable *H. insolens* parent CBH II. All 10 of the functional validation set chimeras retained a greater fraction of their activities than the most stable parent, *H. insolens* CBH II (Fig. S1).

We next investigated the activities of selected thermostable chimeras, using purified enzymes. The parent CBH II enzymes and 3 thermostable chimeras, the most thermostable sample set chimera 11113132, the most thermostable validation set chimera 13311332 and the HJPlus chimera 12222332, were expressed with C-terminal His_6 purification tags and purified. To minimize thermal inactivation of CBH II enzymes during the activity test, we used a shorter, 2-h incubation with the PASC substrate at 50 °C, pH 4.8. As shown in Table 3, the parent and chimera CBH II specific activities were within a factor of 4 of the most active parent CBH II enzyme, from *H. jecorina.* The specific activity of HJPlus was greater than all other CBH II enzymes tested, except for *H. jecorina* CBH II.

The pH dependence of cellulase activity is also important, because a broad pH/activity profile would allow the use of a CBH II chimera under a wider range of potential cellulose hydrolysis conditions. *H. jecorina* CBH II has been observed to have optimal activity in the pH range 4 to 6, with activity markedly reduced outside these values (16). Fig. 4 shows that the *H. insolens* and *C. thermophilum* CBH II enzymes and all 3 purified thermostable CBH II chimeras have pH/activity profiles that are considerably broader than that of *H. jecorina* CBH II. Although

Table 3. Specific activity values (μg glucose reducing sugar equivalent/(μg CBH II enzyme × min)) for parent and selected thermostable CBH II enzyme chimeras

CBH II Enzyme	μg of reducing sugar/ (μg of enzyme) × min) × 10^2
Humicola insolens (Parent 1)	4.0 ± 0.5
Hypocrea jecorina (Parent 2)	13 ± 2
Chaetomium thermophilum (Parent 3)	5.0 ± 0.5
HJPlus (Chimera 12222332)	10.0 ± 0.8
Chimera 11113132	4.5 ± 0.5
Chimera 13311332	6.7 ± 0.3

Error bars show standard errors, where standard error is defined as standard dev/sqrt (*n*), for between 5 and 8 replicates per CBH II enzyme. Two-hour reaction, 3 mg enzyme/g PASC, 50°C, 25 mM sodium acetate, pH 4.8.

Liu et al. (16) report an optimal pH of 4 for *C. thermophilum* CBH II, we observed the optimal pH of the recombinant enzyme to be near 7. Native *H. insolens* CBH II has a broad pH/activity profile, with maximum activity around pH 9 and ≈60% of this maximal activity at pH 4 (17). A similarly broad profile was observed for the recombinant enzyme. The HJPlus chimera has a much broader pH/activity profile than *H. jecorina* CBH II, showing a pH dependence similar to the other 2 parent CBH II enzymes.

Achieving activity at elevated temperature and retention of activity over extended time intervals are two primary motivations for engineering highly stable CBH II enzymes. We therefore measured the performance of thermostable CBH II chimeras in cellulose hydrolysis across a range of temperatures over a 40-h time interval. As shown in Fig. 5, all 3 thermostable chimeras were active on PASC at higher temperatures than the parent CBH II enzymes. The chimeras retained activity at 70 °C, whereas the *H. jecorina* CBH II did not hydrolyze PASC above 57 °C and the stable *H. insolens* enzyme showed no hydrolysis above 63 °C. The activity of HJPlus in long-time cellulose hydrolysis assays exceeded that of all of the parents at their respective optimal temperatures.

Discussion

We have constructed a diverse family of CBH II enzymes by swapping blocks of sequence from 3 fungal CBH II enzymes. Twenty-three of 48 chimeric sequences sampled from this set were secreted in active form by *S. cerevisiae*, and 5 have half-lives at 63 °C that were greater than the most stable parent. Given that this 48-member sample set represents <1% of the total possible 6,561 sequences, we predict that this one collection of chimeras

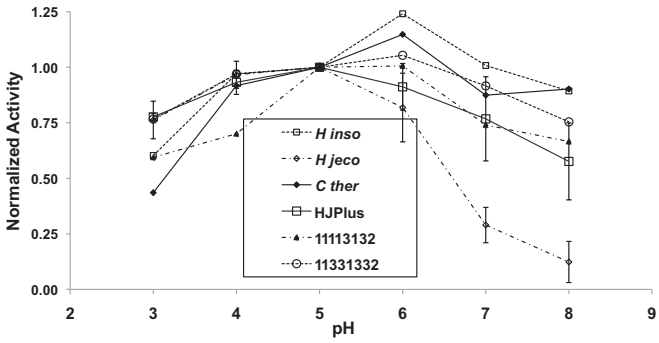

Fig. 4. Specific activity, normalized to pH 5.0, as a function of pH for parent CBH II enzymes and 3 thermostable chimeras. Data presented are averages for 2 replicates. Error bars for HJPlus and *H. jeco* denote values for 2 independent trials. Sixteen-hour reaction, 300 μg of enzyme per gram of PASC, 50 °C, 12.5 mM sodium citrate/12.5 mM sodium phosphate buffer at pH as shown.

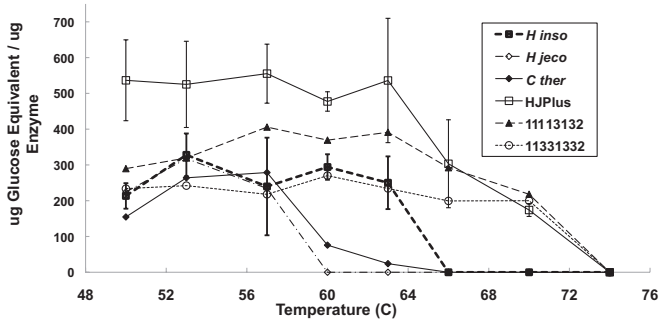

Fig. 5. Long-time cellulose hydrolysis assay results (micrograms of glucose reducing sugar equivalent per microgram of CBH II enzyme) for parents and thermostable chimeras across a range of temperatures. Error bars indicate standard errors for 3 replicates of HJPlus and *H. insolens* CBH II enzymes. 40-h reaction, 100 μg of enzyme per gram of PASC, 50 mM sodium acetate, pH 4.8.

already contains hundreds of active, thermostable CBH II enzymes, a number that dwarfs the approximately twenty fungal CBH II enzymes in the CAZy database (www.cazy.org).

Our approach of using the sample set sequence-stability data to identify blocks that contribute positively to chimera thermostability was validated by finding that all 10 catalytically active chimeras in the second CBH II validation set were more thermostable than the most stable parent, a naturally-thermostable CBH II from the thermophilic fungus, *H. insolens*. This work has thus far generated a total of 33 new CBH II enzymes that are expressed in catalytically active form in *S. cerevisiae*, 15 of which are more thermostable than the most stable parent from which they were constructed. These 15 thermostable enzymes are diverse in sequence, differing from each other and their closest natural homologs at as many as 94 and 58 amino acid positions, respectively.

Analysis of the thermostabilities of CBH II chimeras in the combined sample and validation sets indicates that the 4 thermostabilizing blocks identified, B1P1, B6P3, B7P3, and B8P2, make cumulative contributions to thermal stability when present in the same chimera. Four of the 5 sample set chimeras that are more thermostable than the *H. insolens* CBH II contain either 2 or 3 of these stabilizing blocks (Table 1). The 10 active members of the validation set, all of which are more stable than the *H. insolens* enzyme, contain at least 2 stabilizing blocks, with 5 of the 6 most thermostable chimeras in this group containing either 3 or 4 stabilizing blocks. With additional sequence-stability data for model construction, it will be possible to test whether SCHEMA blocks contribute in a fully additive manner, similar to that for P450 chimeras (3).

Minimizing the number of broken contacts upon recombination (Fig. 2*C*) allows the blocks to be approximated as decoupled units that make independent contributions to the stability of the entire protein, thus leading to cumulative or even additive contributions to chimera thermostability. For this CBH II enzyme recombination, SCHEMA was effective in minimizing such broken contacts: whereas there are 303 total interblock contacts defined in the *H. insolens* parent CBH II crystal structure, the CBH II SCHEMA library design results in only 33 potential broken contacts. Given that the CBH II enzyme parents do not feature obvious structural subdomains, and only 4 of the 8 blocks (1, 5, 7, and 8) resemble compact structural units or modules (18), the low number of broken contacts demonstrates that the SCHEMA/RASPP algorithm is effective for cases in which the number of blocks appears greater than the number of structural subdivisions. As observed for β-lactamase (4) and cytochrome P450 (5) chimeras, low E values were predictive of chimera folding and activity. Although not used here, this relationship should be valuable for designing chimera sample sets that contain a high fraction of active members.

We investigated whether recombination might also lead to improvements in other enzyme properties, such as pH dependence, specific activity and the ability to hydrolyze cellulose over long time intervals and at elevated temperatures. Whereas the specific activity of *H. jecorina* CBH II declines sharply as pH increases above the optimum value of 5, HJPlus, created by substituting stabilizing blocks onto the most industrially relevant *H. jecorina* CBH II enzyme, retains significantly more activity at these higher pHs (Fig. 4). The thermostable 11113132 and 13311332 chimeras, the *H. insolens* and *C. thermophilum* CBH II cellulase parents, have even broader pH/activity profiles than HJPlus. The narrow pH/activity profile of *H. jecorina* CBH II has been attributed to the deprotonation of several carboxyl-carboxylate pairs, which destabilizes the protein above pH $\approx$ 6 (19). The substitution of parent 3 in block 7 in HJPlus changes aspartate 277 to histidine, eliminating the carboxyl-carboxylate pair between D277 and D316 (of block 8). Replacing D277 with the positively charged histidine may prevent destabilizing charge repulsion at nonacidic pH, allowing HJPlus to retain activity at higher pH than *H. jecorina* CBH II. The even broader pH/activity profiles of the remaining 2 thermostable chimeras and the *H. insolens* and *C. thermophilum* parent CBH II enzymes may be due to the absence of acidic residues at positions corresponding to the E57-E119 carboxyl-carboxylate pair of HJPlus and *H. jecorina* CBH II.

HJPlus exhibits both relatively high specific activity (Table 3) and high thermostability (Fig. S2). Fig. 5 shows that these properties lead to good performance in long-time hydrolysis experiments: HJPlus hydrolyzed cellulose at temperatures 7–15 °C higher than the parent CBH II enzymes had a significantly increased long-time activity relative to all of the parents at their temperature optima, bettering *H. jecorina* CBH II by a factor of 1.7. Given that the specific activity of the HJPlus chimera is less than that of the *H. jecorina* CBH II parent, this increased long-time activity can be attributed to the ability of the thermostable HJPlus to retain activity at optimal hydrolysis temperatures over longer reaction timer.

The other two thermostable chimeras shared HJPlus's broad temperature range. This observation supports a positive correlation between $t_{1/2}$ at elevated temperature and maximum operating temperature, and suggests that many of the thermostable chimeras among the 6,561 CBH II chimera sequences will also be capable of degrading cellulose at elevated temperatures. Although this ability to hydrolyze the amorphous PASC substrate at elevated temperatures bodes well for the potential utility of thermostable fungal CBH II chimeras, studies with more challenging crystalline substrates and substrates containing lignin will provide a more complete assessment of this CBH II enzyme family's relevance to biomass degradation applications.

The majority of biomass conversion processes use mixtures of fungal cellulases (primarily CBH II, cellobiohydrolase class I (CBH I), endoglucanases and β-glucosidase) to achieve high levels of cellulose hydrolysis. Generating a diverse group of thermostable CBH II enzyme chimeras is the first step in building an inventory of stable, highly active cellulases from which enzyme mixtures can be formulated and optimized for specific applications and feedstocks.

Materials and Methods

CBH II Expression Plasmid Construction. Parent and chimeric genes encoding CBH II enzymes were cloned into yeast expression vector YEp352/PGK91-1-αss (Fig. S2). DNA sequences encoding parent and chimeric CBH II catalytic domains were designed with *S. cerevisiae* codon bias, using GeneDesigner software (DNA2.0), and synthesized by DNA2.0. The CBH II catalytic domain genes were digested with XhoI and KpnI, ligated into the vector between the

BIOCHEMISTRY

XhoI and KpnI sites and transformed into *E. coli* XL-1 Blue (Stratagene). CBH II genes were sequenced using the following primers: *CBH2L* (5′-*GCTGAACGTGTCATCGGTTAC*-3′) and *RSQ3080* (5′-*GCAACACCTGGCAATTCCTTACC*-3′). C-terminal His_6 parent and chimera CBH II constructs were made by amplifying the CBH II gene with forward primer *CBH2LPCR* (5′-*GCTGAACGTGTCATCGTTACTTAG*-3′) and reverse primers complementary to the appropriate CBH II gene with His_6 overhangs and stop codons. PCR products were ligated, transformed and sequenced as above.

CBH II Enzyme Expression in *S. cerevisiae*. *S. cerevisiae* strain YDR483W BY4742 (*Matα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 Δkre2*, ATCC No. 4014317) was made competent using the EZ Yeast II Transformation Kit (Zymo Research), transformed with plasmid DNA and plated on synthetic dropout −uracil agar. Colonies were picked into 5 mL overnight cultures of synthetic dextrose casamino acids (SDCAA) media (20 g/L dextrose, 6.7 g/L Difco yeast nitrogen base, 5 g/L Bacto casamino acids, 5.4 g/L Na_2HPO_4, 8.56 g/L $NaH_2PO_4{\cdot}H_2O$) supplemented with 20 μg/mL tryptophan and grown overnight at 30 °C, 250 rpm. Five-milliliter cultures were expanded into 40 mL of SDCAA in 250-mL Tunair flasks (Shelton Scientific) and shaken at 30 °C, 250 rpm for 48 h. Cultures were centrifuged, and supernatants were concentrated to 500 μL, using an Amicon ultrafiltration cell fitted with 30-kDa PES membrane, for use in $t_{1/2}$ assays. Concentrated supernatants were brought to1 mM phenylmethylsulfonylfluoride and 0.02% NaN_3. His_6-tagged CBH II proteins were purified using Ni-NTA spin columns (Qiagen) per the manufacturer's protocol and the proteins exchanged into 50 mM sodium acetate, pH 4.8, using Zeba-Spin desalting columns (Pierce). Purified protein concentration was determined using Pierce Coomassie Plus protein reagent with BSA as standard. SDS/PAGE analysis was performed by loading either 20 μL of concentrated culture supernatant or ≈5 μg of purified CBH II enzyme onto a 7.5% Tris·HCl gel (Bio-Rad) and staining with SimplyBlue safe stain (Invitrogen). CBH II supernatants or purified proteins were treated with EndoH (New England Biolabs) for 1 h at 37 °C per the manufacturer's instructions. CBH II enzyme activity in concentrated yeast culture supernatants was measured by adding 37.5 μL of concentrated culture supernatant to 37.5 μL of PASC and incubating for 2 h at 50 °C. Reducing sugar equivalents formed were determined via Nelson-Somogyi (20) assay as described below.

Half-Life, Specific Activity, pH-Activity and Long-Time Cellulose Hydrolysis Measurements. Phosphoric acid swollen cellulose (PASC) was prepared as described by Zhang et al. (21) To enhance CBH II enzyme activity on the substrate, PASC was preincubated at a concentration of 10 g/L with 10 mg/mL *A. niger* endoglucanase (Sigma) in 50 mM sodium acetate, pH 4.8 for 1 h at 37 °C. Endoglucanase was inactivated by heating to 95 °C for 15 min, PASC was washed twice with 50 mM acetate buffer and resuspended at 10 g/L in deionzed water.

CBH II enzyme $t_{1/2}$ values were measured by adding concentrated CBH II expression culture supernatant to 50 mM sodium acetate, pH 4.8, at a concentration giving A_{520} of 0.5 as measured in the Nelson–Somogyi reducing sugar assay (20) after incubation with treated PASC as described below. The 37.5-μL CBH II enzyme/buffer mixtures were inactivated in a water bath at 63 °C. After inactivation, 37.5 μL of endoglucanase-treated PASC was added and hydrolysis was carried out for 2 h at 50 °C. Reaction supernatants were filtered through Multiscreen HTS plates (Millipore). Nelson–Somogyi assay $\log(A_{520})$ values, obtained using a SpectraMax microplate reader (Molecular Devices) corrected for background absorbance, were plotted versus time and CBH II enzyme half-lives obtained from linear regression, using Microsoft Excel.

For specific activity measurements, purified CBH II enzyme was added to PASC to give a final reaction volume of 75 μL of 25 mM sodium acetate, pH 4.8, with 5 g/L PASC and CBH II enzyme concentration of 3 mg of enzyme per gram of PASC. Incubation proceeded for 2 h in a 50 °C water bath and the reducing sugar concentration determined. For pH/activity profile measurements, purified CBH II enzyme was added at a concentration of 300 μg per gram of PASC in a 75-μL reaction volume. Reactions were buffered with 12.5 mM sodium citrate/12.5 mM sodium phosphate, run for 16 h at 50 °C and reducing sugar determined. Long-time cellulose hydrolysis measurements were performed with 300-μL volumes of 1 g/L treated PASC in 100 mM sodium acetate, pH 4.8, 20 mM NaCl. Purified CBH II enzyme was added at 100 μg per gram of PASC and reactions carried out in water baths for 40 h before reducing sugar determination.

ACKNOWLEDGMENTS. We thank Iogen, Inc. (Ottawa, ON) for providing the YEp352/PGK91–1-αss vector and recommending use of the Δ*kre2 S. cerevisiae* strain. This work was supported by the Army-Industry Institute for Collaborative Biotechnologies and the Caltech Innovation Institute.

1. Viikari L, Alapuranen M, Puranen T, Vehmaanpera J, Siika-aho M (2007) Thermostable enzymes in lignocellulose hydrolysis. *Adv Biocheml Eng Biotechnol* 108:121–145.
2. Szijártó N, et al. (2008) Hydrolysis of amorphous and crystalline cellulose by heterologously produced cellulases of Melanocarpus albomyces. *J Biotechnol* 156:140–147.
3. Li Y, et al. (2007) A diverse family of thermostable cytochrome P450s created by recombination of stabilizing fragments. *Nat Biotechnol* 25:1051–1056.
4. Meyer MM, Hochrein L, Arnold FH (2006) Structure-guided SCHEMA recombination of distantly related beta-lactamases. *Protein Eng Des Sel* 19:563–570.
5. Otey CR, et al. (2006) Structure-guided recombination creates an artificial family of cytochromes P450. *PLoS Biol* 4:e112.
6. Endelman JB, Silberg JJ, Wang ZG, Arnold FH (2004) Site-directed protein recombination as a shortest-path problem. *Protein Eng Des Sel* 17:589–594.
7. Penttilä M, et al. (1988) Efficient secretion of two fungal cellobiohydrolases by *Saccharomyces cerevisiae. Gene* 63:103–102.
8. Godbole S, et al. (1999) Cloning and expression of Trichoderma reesei cellobiohydrolase I in Pichia pastoris. *Biotechnol Prog* 15:828–833.
9. Jeoh T, Michener W, Himmel ME, Decker SR, Adney WS (2008) Implications of cellobiohydrolase glycosylation for use in biomass conversion. *Biotechnol Biofuels* 1:10.
10. Lussier M, Sdicu AM, Bussey H (1999) The KTR and MNN1 mannosyltransferase families of Saccharomyces cerevisiae. *Biochim Biophys Acta* 1426:323–334.
11. Hui JP, White TC Thibault P (2002) Identification of glycan structure and glycosylation sites in cellobiohydrolase II and endoglucanases I and II from Trichoderma reesei. *Glycobiology* 12:837–849.
12. Innis MA (1989) Glycosylation of heterologous proteins in Saccharomyces cerevisiae. *Biotechnology* 13:233–246.
13. Varrot A, et al. (2003) Distortion of a cellobio-derived isofagomine highlights the potential conformational itinerary of inverting beta-glucosidases. *Chem Commun* 8:946–947.
14. Liao J, et al. (2007) Engineering proteinase K using machine learning and synthetic genes. *BMC Biotechnol* 26:7–16.
15. Koivula A, et al. (1996) The active site of Trichoderma reesei cellobiohydrolase II: The role of tyrosine 169. *Protein Eng* 9:691–699.
16. Liu SA, Li DC, Zhang Y (2005) Cloning and expressing of cellulase gene (CBH2) from thermophilic fungi Chaetomium thermophilum CT2. *Sheng Wu Gong Cheng Xue Bao* 21:892–899.
17. Schulein M (1997) Enzymatic properties of cellulases from Humicola insolens. *J Biotechnol* 57:71–81.
18. Go M, Nosaka M (1987) Protein architecture and the origin of introns. *Cold Spring Harbor Symp Quant Bio* 52:915–924.
19. Wohlfahrt G, Pelikka T, Boer H, Teeri TT, Koivula A (2003) Probing pH-dependent functional elements in proteins: Modification of carboxylic acid pairs in *Trichoderma reesei* cellobiohydrolase Cel6A. *Biochemistry* 42:10095–10103.
20. Nelson HA (1944) Photometric adaptation of the Somogyi method for the determination of glucose. *J Biol Chem* 153:375–380.
21. Zhang YHP, Cui XB, Lynd LR, Huang L (2006) A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: Evidences from supramolecular structures and enzymatic hydrolysis. *Biomacromolecules* 7:644–648.