

Highly thermostable fungal cellobiohydrolase I (Cel7A) engineered using predictive methods

Russell S.Komor, Philip A.Romero, Catherine B.Xie and Frances H.Arnold¹

Division of Chemistry and Chemical Engineering, California Institute of Technology 210-41, Pasadena, CA 91125, USA

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

Received July 13, 2012; revised July 13, 2012;
accepted August 13, 2012

Edited by Anthony Wilkinson

Building on our previous efforts to generate thermostable chimeric fungal cellobiohydrolase I (CBH I, also known as Cel7A) cellulases by structure-guided recombination, we used FoldX and a ‘consensus’ sequence approach to identify individual mutations present in the five homologous parent CBH I enzymes which further stabilize the chimeras. Using the FoldX force field, we calculated the effect on $\Delta G_{\text{Folding}}$ of each candidate mutation in a number of CBH I structures and chose those predicted to be stabilizing in multiple structures. With an alignment of 41 CBH I sequences, we also used amino acid frequencies at each candidate position to calculate predicted effects on $\Delta G_{\text{Folding}}$. A combination of mutations chosen using these methods increased the T_{50} of the most thermostable chimera by an additional 4.7°C, to yield a CBH I with T_{50} of 72.1°C, which is 9.2°C higher than that of the most stable native CBH I, from *Talaromyces emersonii*. This increased stability resulted in a 10°C increase in the optimal temperature for activity, to 65°C, and a 50% increase in total sugar production from crystalline cellulose at the optimal temperature, compared with native *T.emersonii* CBH I.

Keywords: CBH I/Cel7A/cellulase/protein engineering

Introduction

Currently, only 3% of energy for transportation in the USA comes from renewable sources (USEIA, 2010). Fossil fuels provide the rest, of which more than two-thirds are imported. The US Energy Independence and Security Act (EISA) of 2007 mandated increased use of renewable fuels to 36 billion gallons a year by 2022, nearly half of which would come from cellulosic feedstocks (USDA, 2010) such as trees, grasses or agricultural wastes. According to a recent evaluation, the EISA goals are unlikely to be met without major technological advances. Thus, development of better methods to break down biomass is even more critical than when EISA was released (NRC, 2011). One approach has focused on engineering cellulases to improve cellulose conversion into fermentable sugars. A promising strategy is to increase enzyme stability so that cellulose deconstruction can be carried at higher temperatures (above 50°C) (Viikari

et al., 2007), taking advantage of reduced contamination and viscosity as well as any increase in hydrolysis rate with temperature. Thermostable cellulases also tend to have longer lifetimes at lower temperatures and can possibly be recovered after use, which can translate to lower enzyme costs if the stability increase does not come at the cost of cellulase activity (Anbar *et al.*, 2012).

Cellobiohydrolase I (CBH I, members of the glycosyl hydrolase family 7, Cel7A) hydrolyzes cellulose chains processively from the reducing end and accounts for ~40% of the total protein and ~70% of the cellulytic activity in the industrially relevant fungus *Hypocrea jecorina* (Suominen *et al.*, 1993). CBH I has proven difficult to engineer for improved performance, a fact that stems in part from its low expression level in heterologous hosts (Laymon *et al.*, 1996; Jeoh *et al.*, 2008) and the complex structure of the enzyme: CBH I contains 8–10 disulfide bonds, depending on the fungal source, and at least three glycosylation sites, which lead to hyperglycosylation in commonly used yeast expression systems (Boer *et al.*, 2000). Only limited success has been reported for increasing the thermostability of fungal CBH I by random mutagenesis and screening (Voutilainen *et al.*, 2007). Most works have focused on more ‘rational’ protein engineering approaches such as engineering disulfide bonds (Voutilainen *et al.*, 2010) or appending carbohydrate binding modules (Voutilainen *et al.*, 2009) to increase activity on cellulose at elevated temperatures.

In our previous work (Heinzelman *et al.*, 2010), we used five enzymes from thermophilic fungi (*Chaetomium thermophilum*, *Thermoascus aurantiacus*, *H. jecorina*, *Acremonium thermophilum* and *Talaromyces emersonii*) as parents with which to construct CBH I chimeras by SCHEMA structure-guided recombination. From measurements of the thermostabilities of a sample set of the chimeras expressed in a hypoglycosylating strain of *Saccharomyces cerevisiae*, we identified stabilizing and neutral sequence blocks, which we used to construct a new set of diverse, thermostable chimeric fungal cellulases. These stable chimeras were fully active and had T_{50} values (temperature at which half of the enzyme is inactivated after a 10-min incubation) of up to 67.4°C (Heinzelman *et al.*, 2010; Komor, 2012) compared with 62.9°C for the most thermostable of the five parents, the CBH I from *T.emersonii*.

Sequence blocks that contribute positively to protein stability can be found by analyzing the sequence–stability relationship for a set of chimeras. Further analysis of the mutations within stabilizing blocks enabled discovery of individual mutations which stabilize chimeras that do not already have them (Heinzelman *et al.*, 2009a). In an alternative approach to identifying such stabilizing mutations present in native CBH I sequences, we evaluated two computational methods for their ability to screen hundreds of homologous amino acid substitutions: (i) physicochemical modeling of the mutation effects in the protein 3D structure using FoldX (Guerois *et al.*, 2002) and (ii) ‘consensus’

analysis of a multiple sequence alignment (MSA) of evolutionarily related sequences (Steipe et al., 1994). Upon experimental testing of 43 mutations predicted to be stabilizing by one or both of these methods, we identified several that when combined produce a highly stable CBH I that is also highly active on crystalline cellulose at 70°C.

Materials and methods

CBH I mutagenesis

Native (parent) CBH I genes featured native codon usage and were synthesized by DNA2.0 (Menlo Park, CA, USA). Chimeric CBH I genes were constructed using splicing by overlap extension polymerase chain reaction (PCR) (Higuchi et al., 1988) with primers for each unique junction. Individual mutations were introduced using the QuikChange Lightning Site-Directed Mutagenesis Kit from Agilent Technologies (Santa Clara, CA, USA). Primers for the point mutations contained the new codon flanked on either side by 8–15 bp complementary to the parental sequence. N-terminal His₆ CBH I constructs were made via PCR using Phusion High-Fidelity DNA Polymerase from Finnzymes (Vantaa, Finland) according to the manufacturer's protocol for PCR conditions with forward primers complementary to the appropriate CBH I N-terminal sequence with NheI and His₆ overhangs. CBH I genes were cloned in a yeast expression vector Yep352/PGK91-1- α ss and transformed into expression strain YDR483W as described (Heinzelman et al., 2010).

CBH I expression

Yeast strain YDR483W BY4742 (*Mata his3 Δ 1 leu2 Δ 0 lys2 Δ 0 ura3 Δ 0 Δ kre2*, ATCC No. 4014317) containing the CBH I genes was plated on synthetic dropout-uracil medium supplemented with 16 wt% agar, and grown at 30°C for 2 days. Single colonies were picked on the second day, placed in 5 ml synthetic dextrose casamino acids (SDCAA) medium (20 g/l dextrose, 6.7 g/l Difco yeast nitrogen base, 5 g/l Bacto casamino acids, 5.4 g/l Na₂HPO₄, 8.56 g NaH₂PO₄·H₂O), and grown overnight with shaking at 250 rpm. The following morning, cultures were expanded into 40 ml of yeast peptone dextrose (YPD) medium (20 g Bacto peptone, 10 g Bacto yeast extract and 20 g dextrose) in 250 ml Tunair flasks from Sigma-Aldrich (St Louis, MO, USA) and grown for 48 h with 250 rpm shaking. Cultures were centrifuged at 4500 g for 15 min, then decanted, brought to 0.02% NaN₃ and 1/2000× Protease Inhibitor Cocktail (Sigma-Aldrich). Decanted supernatants were then stored at 4°C until characterization.

CBH I purification

Strains transformed with the N-terminal His₆ CBH I constructs were grown as described above, except after centrifugation supernatants were filtered with 0.45 μ m pore size filter units from Nalgene (Rochester, NY, USA) before being brought to 0.02% NaN₃ and 1/200× Protease Inhibitor Cocktail. The decanted supernatants were then concentrated to a volume <1 ml with Vivaspin 20 ultrafiltration spin columns with a 30 kDa MWCO PES membrane from Sartorius Stedim (Aubagne Cedex, France). The concentrated supernatants were then purified using Ni-NTA spin columns

from Qiagen (Valencia, CA, USA) as per the manufacturer's protocol and the proteins exchanged into 50 mM sodium acetate, pH 4.8, using the Vivaspin 20 spin columns. Purified protein concentration was determined using the Bradford Protein Assay from BioRad (Hercules, CA, USA) with a bovine serum albumin standard and concentrations determined by averaging readings of multiple dilutions for each sample.

Thermostability (T_{50}) measurements

We define T_{50} as the temperature at which a 10-min incubation in the absence of substrate causes loss of one-half of the cellulase activity, measured after reaction on 4-methylumbelliferyl lactopyranoside (MUL) (Sigma-Aldrich), relative to a 100% activity reference sample that does not undergo incubation. For T_{50} assays, culture supernatants were diluted using a supernatant from a negative control YPD yeast culture that does not secrete cellulase activity (it contains the Yep352/PGK91-1- α ss plasmid containing a CBH I gene with a frameshift mutation) so that approximately equivalent MUL hydrolysis rates of 2.0×10^{-8} mol MUL/l/s were obtained for samples not incubated for thermal denaturation. These diluted samples were adjusted to 1 mM DTT to ensure complete irreversible unfolding, and 125 mM sodium acetate, pH 4.8. Aliquots of 125 μ l were incubated for 10 min in a water bath across a range of temperatures bracketing the T_{50} value. Water bath temperatures were measured using two different alcohol thermometers and observed to be consistent within 0.1°C. After cooling, 25 μ l of 1.8 mM MUL dissolved in 125 mM sodium acetate, pH 4.8, 18% dimethyl sulfoxide, was added to the incubated and unheated samples, and these were reacted in a 45°C water bath for 90 min before quenching with 150 μ l of 1 M Na₂CO₃. MUL hydrolysis rates were determined using a microplate reader to measure sample fluorescence with excitation at 364 nm and emission at 445 nm and comparing values to a standard curve prepared with 4-methylumbelliferone (Sigma-Aldrich).

CBH I activity measurements

CBH I activity on MUL was measured at different temperatures. CBH I culture supernatants were diluted as above based on MUL activity measurements at 45°C for 60 min so that approximately equivalent MUL hydrolysis rates of 4.0×10^{-8} mol MUL/l/s were obtained. Duplicate samples of each dosed enzyme were reacted at several temperatures for 60 min, quenched and read as above.

CBH I temperature–activity profiles on solid cellulose were obtained by adding 5 μ g of the affinity-purified CBH I to 500 μ l of 50 mM sodium acetate, pH 4.8, containing 60 mg/ml Lattice NT cellulose from FMC (Philadelphia, PA, USA). After incubation for 20 h in a water bath at the temperature of interest, supernatant reducing sugar was determined by the Nelson–Somogyi assay as described (Heinzelman et al., 2010). Reactions were run in duplicate and repeated on different days to assess error bars.

Results

Prediction of stabilizing mutations

FoldX estimates a mutation's effect on protein stability ($\Delta\Delta G_{\text{FoldX}}$) using an atomic force field with empirically determined coefficients (Guerois et al., 2002; Schymkowitz

et al., 2005). Because the thermostable chimeras are composed of blocks from five different parent CBH Is, we expect their structures to diverge somewhat from any one parent crystal structure. In an attempt to circumvent errors that might arise from using a single, inappropriate structure for the calculation of $\Delta\Delta G_{\text{FoldX}}$, we calculated the effects of each mutation in 39 different fungal CBH I crystal structures available in the PDB (see Table SIII in supplemental information), 18 of which are of the parental CBH Is. Using multiple structures allows us to identify mutations that are predicted to be stabilizing in the majority of structures. FoldX calculations are therefore reported as mean $\Delta\Delta G_{\text{FoldX}}$ values and standard deviations for each mutation across all 39 structures.

Consensus analysis uses MSAs of evolutionarily related proteins to identify stabilizing mutations (Steipe *et al.*, 1994). A sufficiently large MSA approximates a canonical ensemble, and the most probable distribution of amino acids at a specific position is a Boltzmann distribution. The argument is that the frequency of a particular amino acid at a given position deviates from that expected at random according to its effect on stability, and the stability effect of any given mutation is estimated from

$$\Delta\Delta G_{\text{mut}} = -RT \ln \frac{f_{\text{mut}}}{f_{\text{WT}}} \quad (1)$$

Here f_{mut} is the frequency of the new amino acid and f_{WT} is the frequency of the original amino acid at that position (Steipe *et al.*, 1994). Although clearly not correct for all residue positions (some of which may undergo selection directly for biological function), this ‘consensus stabilization’ approach has proven useful in identifying stabilizing mutations in a variety of proteins (Pantoliano *et al.*, 1989; Steipe *et al.*, 1994; Nikolova *et al.*, 1998; Wang *et al.*, 1999), including other cellulases (Anbar *et al.*, 2012).

To assemble the CBH I MSA, we searched the NCBI non-redundant protein database and selected 41 CBH I sequences (including the five CBH I parents in the SCHEMA library) having at least 54% sequence identity to any of the parents. This lower limit on sequence identity was chosen because the more divergent sequences are more difficult to align structurally. These sequences were aligned using ClustalW2 (see Fig. S4 in supplemental information). We calculated $\Delta\Delta G_{\text{FoldX}}$ and $\Delta\Delta G_{\text{mut}}$ at each position that varies among the five native CBH Is used for recombination as well as the CBH I from the thermophilic fungus *Melanocarpus albomyces* (Szijarto *et al.*, 2008), for a total of 470 possible homologous mutations. Because promising mutations were tested in chimeras composed of blocks from five different parent CBH Is, each position can have as many as five values for f_{WT} in the consensus calculations. Only one is appropriate, however, for any given background sequence chosen for testing. Two criteria were used to select mutations to test: (i) $\Delta\Delta G_{\text{FoldX}} \leq -0.75$ kcal/mol and a low standard deviation, meaning that it was predicted to be stabilizing in many of the CBH I structures, and/or (ii) $\Delta\Delta G_{\text{mut}} < 0$, meaning the mutation appeared more frequently at that position in the CBH I MSA than the amino acid it replaced. Our thresholds for selecting mutations identified a total of 43 mutations (the top ~10% of all mutations) to test experimentally in the backgrounds of the most thermostable CBH I

Table I. Effects on thermostability of stabilizing mutations predicted by FoldX and consensus analysis

Mutation	$\Delta\Delta G_{\text{FoldX}}$ (kcal/mol) ^a	$\Delta\Delta G_{\text{mut}}$ (kcal/mol) ^b	Chimera ^c	ΔT_{50}^d
Mutations selected based on FoldX and consensus				
S57D	-0.82 ± 1.12	-1.34	131525[555555]2	-0.8
Y60I	-3.24 ± 6.33	-1.30	551522[552555]2	0.1
Y60L	-3.12 ± 5.90	-1.23	551522[552555]2	0.8
N93K	-1.12 ± 0.59	-0.54	551522[552555]2	2.1
V110L	-1.35 ± 0.83	-0.31	551522[552555]2	-0.9
T164K	-0.95 ± 0.33	-1.21	551522[552555]2	-0.6
V404A	-1.36 ± 1.94	-0.21	151525[554555]2	-1.5
Y430F	-1.68 ± 1.08	-0.06	551522[552555]2	1.5
N439G	-2.59 ± 0.73	-0.28	551522[552555]2	ND
Mutations selected based on FoldX alone				
S13P	-1.43 ± 0.28	0.26	121522[552555]2	0.7
T41V	-1.55 ± 0.47	0.036	131525[555555]2	-0.3
A201P	-0.91 ± 0.37	0.58	551522[552555]2	0.2
S222K	-1.43 ± 0.50	1.40	551522[552555]2	0.0
T257V	-1.01 ± 0.34	1.13	551522[552555]2	0.1
T257K	-1.17 ± 0.38	1.93	151525[554555]2	-2.0
T273K	-0.88 ± 0.23	0.47	551522[552555]2	-0.1
T395P	-2.29 ± 0.62	1.97	151525[554555]2	ND
T408D	-1.22 ± 0.55	1.77	551522[552555]2	0.2
Mutations selected based on consensus alone				
S5T	0.06 ± 0.22	-0.84	121522[552555]2	0.2
D52T	-0.07 ± 0.51	-1.45	131525[555555]2	-0.9
L113M	0.83 ± 1.02	-0.92	551522[552555]2	-0.5
S130T	-0.3 ± 1.03	-1.15	551522[552555]2	0.4
V217I	-0.01 ± 0.83	-1.28	551522[552555]2	-0.3
Mutations predicted by FoldX to be stabilizing in some but not all CBH I structures (high standard deviation)				
N126G	-2.46 ± 0.94	0.15	551522[552555]2	-0.8
H208Y	-1.27 ± 0.69	0.74	551522[552555]2	0.7
V227L	-1.52 ± 1.43	1.23	151525[554555]2	-0.8
T257I	-1.38 ± 0.59	1.11	151525[554555]2	-0.9
T273P	-1.17 ± 0.38	1.83	151525[554555]2	-0.9
D300K	-1.01 ± 0.34	0.57	151525[554555]2	-2.7
S324P	-2.14 ± 1.27	0.20	551522[552555]2	0.9
E325P	-1.25 ± 0.92	1.42	151525[554555]2	-1.8
A329G	-0.95 ± 1.78	0.82	151525[554555]2	-1.4
V331M	-1.47 ± 1.20	1.80	151525[554555]2	-3.5
T339P	-1.38 ± 1.19	0.82	151525[554555]2	-0.8
T339Q	-0.80 ± 0.40	0.58	151525[554555]2	-2.7
Q345M	-1.41 ± 0.78	1.20	151525[554555]2	-2.0
D354V	-2.25 ± 1.25	0.54	551522[552555]2	1.0
H358K	-2.03 ± 1.72	0.16	151525[554555]2	-1.7
H358R	-2.15 ± 1.63	1.15	151525[554555]2	-2.1
H358V	-2.08 ± 2.32	1.80	151525[554555]2	-3.2
A383Y	-0.89 ± 0.49	0.45	551522[552555]2	0.6
T392I	-0.79 ± 1.25	1.95	151525[554555]2	0.5
P399G	-1.40 ± 3.21	0.07	151525[554555]2	-1.8

^aAverage of the FoldX calculations performed on multiple CBH I structures and the standard deviation of the values in different structures.

^bValue predicted by Equation (1).

^cChimera sequences, block numbering and breakpoint locations are provided in our previous work (Heinzelman *et al.*, 2010; Komor, 2012).

^dExperimentally observed effect on T_{50} of the mutation in the CBH I chimera. Error in T_{50} measurements is $< 1^\circ\text{C}$. ND indicates insufficient protein expression for determining a T_{50} value. Values in bold indicate a significantly thermostabilizing effect.

chimeras. Seventy-two mutations that were predicted to be stabilizing in the parent sequences were already present in the thermostable chimeras and were therefore not tested. The 43 tested mutations and their effects on the T_{50} of the most thermostable chimeras are reported in Table I.

Nine of the 43 tested mutations satisfied the criteria for both $\Delta\Delta G_{\text{FoldX}}$ and $\Delta\Delta G_{\text{mut}}$. Three of these were stabilizing

(Y60L, N93K, Y430F, shown in bold in Table I) in an already-thermostable chimera, one was essentially neutral, four were destabilizing and one resulted in loss of enzyme expression. We tested nine other mutations that satisfied the criterion for $\Delta\Delta G_{\text{FoldX}}$ but not $\Delta\Delta G_{\text{mut}}$. One of these (S13P) was stabilizing, six were essentially neutral, one was destabilizing and one caused loss of enzyme expression. Of the five tested mutations that satisfied the criterion for $\Delta\Delta G_{\text{mut}}$ but not $\Delta\Delta G_{\text{FoldX}}$, three were essentially neutral and two were destabilizing. Interestingly, the two mutations that resulted in loss of expression (N439G and T395P) are both located in a chimera block that does not abolish expression when substituted as a whole. Mutation codons were chosen to match those already present in the parental sequences to avoid DNA level effects on expression. This suggests that the loss of expression caused by these mutations is context-dependent and offset by other mutations contained in the block from which they originated.

We also wished to test the hypothesis that mutations predicted to be stabilizing in multiple CBH I structures were stabilizing more often than those predicted to be stabilizing in only a few structures. We thus tested 20 mutations with $\Delta\Delta G_{\text{FoldX}}$ of -0.75 kcal/mol or lower and with large standard deviations, at least 50% of the $\Delta\Delta G_{\text{FoldX}}$ value (Table I). This set yielded five stabilizing mutations (H208Y, S324P, D354Y, A383Y, T392I) and 15 destabilizing mutations, essentially the same success rate as when only a few structures were used.

Overall, 24% of the tested mutations that were predicted by FoldX to be stabilizing actually increased T_{50} of a thermostable chimera, while 21% of the consensus mutations did so.

Figure 1 summarizes the predicted stability effects according to each method for all 470 mutations. Also shown (in color) are the measured effects on T_{50} of the 43 tested mutations. Overall, FoldX predicts that a full 43% (202) of the 470 possible homologous mutations are at least slightly

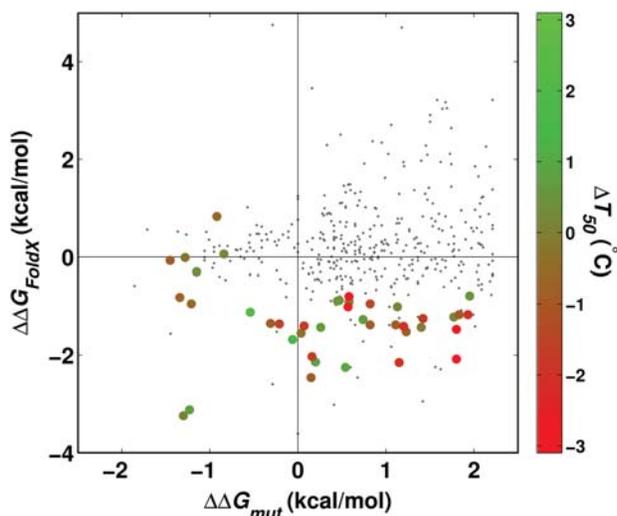


Fig. 1. Stability effects predicted by FoldX and consensus sequence analysis for all 470 possible mutations between six naturally occurring CBH Is. The mutations tested experimentally in thermostable chimeras are shown color-coded according to their measured effects on T_{50} (red is destabilizing, green is stabilizing). Several mutations predicted by both methods to be significantly stabilizing (lower and left-hand quadrants) were already present in the thermostable chimeras and were therefore not tested

stabilizing (although most of these did not meet the stringent selection criterion), and 86% are either stabilizing or essentially neutral. In contrast, a similar calculation using FoldX predicts that only 19% of all possible random mutations are stabilizing (in the same backgrounds) and that 44% are either stabilizing or neutral. Thus, FoldX also predicts that the homologous substitutions are significantly more conservative than random mutations, as has been observed previously (Drummond *et al.*, 2005; Jochens and Bornscheuer, 2010), although the FoldX calculations overestimate these fractions for both homologous and random mutations.

Consensus analysis predicts that fully 22% (103) of the possible homologous substitutions are stabilizing (although again most have predicted effects that do not meet the selection criterion). The two methods, however, agree on only 10%. Nine of these satisfying our selection criteria were tested in the context of a thermostable CBH I chimera, of which only three were actually stabilizing. The two methods disagree on predictions for the remaining mutations. Of the 28 tested mutations from the 155 (33%) additional mutations predicted to be stabilizing by FoldX alone, only six were actually stabilizing. None of the five tested mutations from the 56 (12%) that were predicted to be stabilizing by consensus, but not FoldX, conferred additional stability to the thermostable chimeras.

Combining thermostabilizing mutations

Five stabilizing mutations (S13P, Y60L, S324P, A383Y and Y430F) were chosen based on ease of combination with PCR and introduced into the most stable chimera with the appropriate series of sequence blocks from the five parental CBH Is to accommodate all five mutations (CBH I 'TS0', whose $T_{50} = 67.4 \pm 0.2^\circ\text{C}$, see Table II). This generated CBH I TS5, whose $T_{50} = 70.0 \pm 0.2^\circ\text{C}$. Three of the four remaining stabilizing mutations (N93K, D354V and T392I) were added to CBH I TS5 (Tyr was already present at position 208 in this chimera, so the ninth stabilizing mutation was redundant) to produce CBH I TS8, whose $T_{50} = 72.1 \pm 0.3^\circ\text{C}$. CBH I TS8 is 4.7°C more stable than the most stable native parent CBH I. All the stabilizing mutations and their effects on the stability of CBH I TS0 are summarized in Table II.

Cellulase activity of thermostable CBH I

To assess whether the stabilizing mutations had a negative impact on the catalytic activity and to determine the extent to which increased ability to tolerate incubation at high temperature translates into increased activity at high temperatures, we measured the temperature-dependent activities of the most stable native CBH I (*Temersonii* CBH I), most stable chimera (CBH I TS0), and the most stable chimera with five and eight stabilizing mutations (CBH I TS5 and CBH I TS8). Activities were measured for yeast secretion culture acting on the soluble MUL substrate over a 60-min period at temperatures from 45°C to 70°C ; the results are reported in Fig. 2a. Activities were also measured for purified enzymes acting on microcrystalline cellulose for a period of 20 h at the same temperatures (Fig. 2b). Under both conditions, the *Temersonii* CBH I has an optimal temperature of $\sim 55^\circ\text{C}$, which increases for the stabilized enzymes up to $\sim 65^\circ\text{C}$ for CBH I TS8. The more stable

Table II. Summary of the properties of stabilizing mutations in the CBH I TS0 background

Enzyme ^a	$\Delta\Delta G_{\text{FoldX}}$ (kcal/mol) ^b	$\Delta\Delta G_{\text{mut}}$ (kcal/mol) ^c	T_{50} (°C)	ΔT_{50} (°C) ^d	ΔT_{50} (°C) ^e
CBH I TS0 (151525[554555]2)			67.4 ± 0.2		
CBH I TS0 S13P	-1.43	0.26	67.8 ± 0.5	0.4	0.7
CBH I TS0 Y60L	-3.12	-1.23	67.8 ± 0.2	0.4	0.8
CBH I TS0 S324P	-2.14	0.20	67.6 ± 0.7	0.2	0.9
CBH I TS0 A383Y	-0.89	0.45	67.8 ± 0.3	0.4	0.6
CBH I TS0 Y430F	-1.68	-0.06	68.1 ± 0.1	0.7	1.5
CBH I TS5 (CBH I TS0 S13P, Y60L, S324P, A383Y, Y430F)	-9.26	-0.35	70.0 ± 0.2	2.6	
CBH I TS5 N93K	-1.12	-0.54	70.8 ± 0.1	0.8	2.1
CBH I TS5 D354V	-2.25	0.54	70.6 ± 0.1	0.6	1.0
CBH I TS5 T392I	-0.79	1.94	70.5 ± 0.4	0.5	0.5
CBH I TS8 (CBH I TS5 N93K, D354V, T392I)	-4.16	1.94	72.1 ± 0.3	2.1	

^aChimera sequences, block numbering and breakpoint locations are explained in our previous work (Heinzelman *et al.*, 2010; Komor, 2012).

^bAverage of the FoldX calculations performed on multiple CBH I structures and the standard deviation of the values in different structures.

^cValue predicted by equation 1.

^dExperimentally calculated effect on T_{50} of the mutation compared to its 'parent' enzyme (the enzyme listed before the mutations in the 'enzyme' column). Error in T_{50} measurements is $<1^\circ\text{C}$.

^eExperimentally calculated effect on T_{50} of the mutation in the chimera in which the mutation was first tested. Error in T_{50} measurements is $<1^\circ\text{C}$.

enzymes produce more products at their optimal temperatures than the less stable enzymes at theirs. This effect is most pronounced on the solid substrate and the longer reaction time, with CBH I TS8 producing $\sim 50\%$ more sugar equivalents than the *T.emersonii* CBH I, at their respective optimal temperatures.

Discussion

Although random mutagenesis and screening can be used to identify stabilizing mutations, a high percentage of random mutations are expected to be neutral or destabilizing (Bloom *et al.*, 2005). The low expression levels of CBH I in heterologous hosts, however, limit the number of mutations that can be effectively screened. Recently, CBH I expression levels of up to 0.3 g/l in *S.cerevisiae* have been reported (Ilmen *et al.*, 2011). We instead limited consideration to those amino acids present in the parent sequences and screened potential stabilizing mutations computationally. Homologous mutations are significantly more conservative than random mutations, as they occur largely on the protein surface and are known to be compatible with the protein overall fold and function, in at least one background (Drummond *et al.*, 2005; Romero and Arnold, Submitted 2012). Homologous mutations are also less likely to be detrimental to catalytic activity (Jochens and Bornscheuer, 2010). Limiting the search to mutations in homologs therefore enriches the search in neutral and stabilizing mutations. This is reflected in the FoldX calculations, where the predicted rate of stabilizing homologous mutations is approximately twice the rate when all possible mutations are considered. However, the 470 possible mutations from

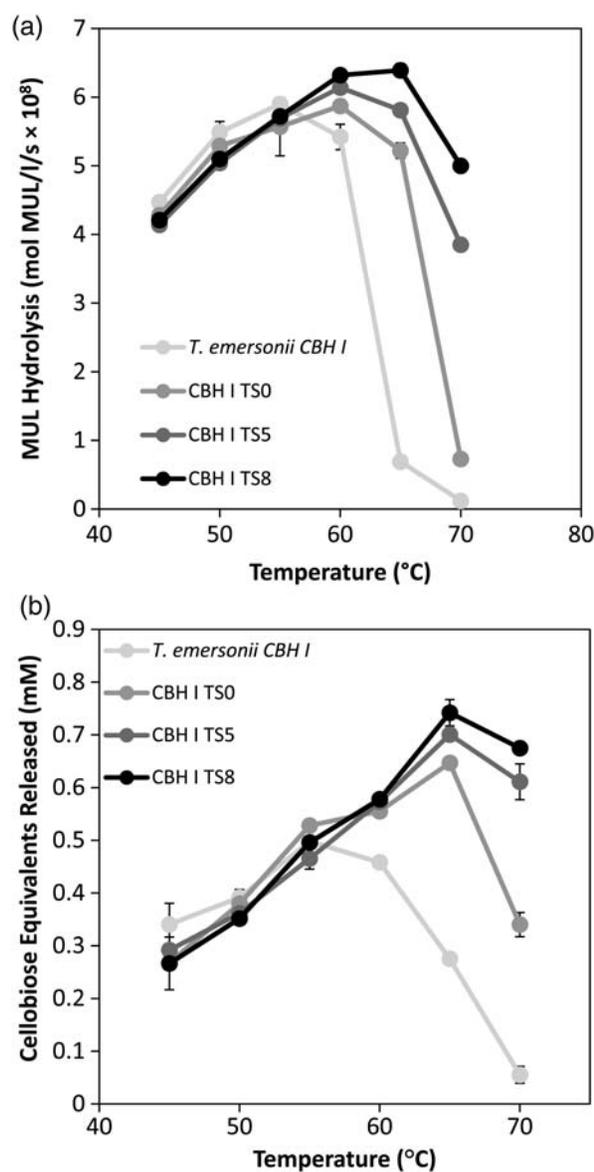


Fig. 2. Activities over a range of temperatures measured for *T.emersonii* CBH I, the most stable chimera (CBH I TS0), TS0 with five thermostabilizing mutations (CBH I TS5) and TS0 with eight thermostabilizing mutations (CBH I TS8) on the (a) soluble MUL substrate at pH 4.8 and (b) solid microcrystalline cellulose (60 mg/Lattice MT cellulose, pH 4.8).

the six CBH I homologs are still too many to test experimentally. As a further filter, we used FoldX and consensus sequence analysis, individually and in combination.

Both FoldX and consensus sequence analysis identified mutations that would further stabilize chimeras that were already highly stable, but only with $\sim 20\%$ accuracy. The criterion for mutation selection using FoldX was a $\Delta\Delta G_{\text{FoldX}} < -0.75$ kcal/mol; as can be seen by examination of the data summarized in Fig. 1, moving this cutoff to < -1.75 kcal/mol increases the reliability to about 45%. There are, however, only a handful of mutations with predicted effects at this level. Dropping the requirement of being predicted to be stabilizing in multiple structures (low standard deviations in $\Delta\Delta G_{\text{FoldX}}$) does not lower the success rate of finding thermostabilizing mutations. Increasing the cutoff level for $\Delta\Delta G_{\text{mut}}$

also does not improve the prediction accuracy for consensus analysis, which remains between 20 and 30%. Both methods are useful for enriching stabilizing mutations when many mutations can be tested in parallel, but the qualitative and quantitative reliability is very limited for individual mutations. These results are consistent with previously published work evaluating the effectiveness of FoldX and similar computational methods, which also concluded that the methods are reasonably good on average but not at predicting the effects of individual mutations (Potapov *et al.*, 2009). Even with the modest success rate, however, we were able to identify several thermostabilizing mutations and use them to construct a CBH I enzyme with increased sugar production at elevated temperatures.

The modest success rate in identifying stabilizing mutations could reflect the fact that they were tested only in highly thermostable chimera backgrounds; it is possible that some mutations that are neutral or even destabilizing in these chimeras become stabilizing when tested in the parent sequences, which are all less stable. Furthermore, many mutations that were predicted to be stabilizing in the parent CBH Is were not tested here, because they were already present in the chimeras. A larger fraction of these mutations could be stabilizing in the parent backgrounds. It is also possible that the thermodynamic stability predictions made by these computational methods, particularly FoldX, do not fully align with the T_{50} measurements of kinetic stability in the irreversibly unfolding CBH I. Thermodynamic stability is the difference in free energy between folded and unfolded state(s) that are in equilibrium (Privalov, 1979). Kinetic stability reflects the difference in free energy between the folded state and a transition state for unfolding and is a measure of the activation energy of irreversible unfolding (Rodriguez-Larrea *et al.*, 2006). Because most industrial enzymes, including cellulases, unfold at least partially irreversibly, increasing kinetic stability is the target for enzyme engineering (Sanchez-Ruiz, 1992). However, it is difficult for computational methods to predict kinetic stability, since the structure of the transition state is unknown. It is therefore assumed that kinetic unfolding goes through a partially unfolded state whose energy is similar to or at least proportional to that of the unfolded state.

To investigate the effects of the selected mutations on the enzyme structure and the basis for the observed changes in stability we used the crystal structures of the parent CBH Is from *H.jecorina* (Divne *et al.*, 1994) and *T.emersonii* (Grassick *et al.*, 2004). Figure 3 shows the locations of the nine stabilizing mutations with the *T.emersonii* CBH I structure used to illustrate the distribution of mutations. The stabilizing mutations occur mainly on the surface of the protein, which is not unexpected since most homologous mutations are also on the surface. Buried mutations are more likely to be destabilizing and lead to loss of core catalytic function; they tend to accumulate more in proteins that are more divergent in their sequences. The only stabilizing mutation that is not directly on the surface is Y430F, which is located in a surface beta sheet but with its side chain facing into a hydrophobic pocket in both the *T.emersonii* CBH I structure (tyrosine) and the *H.jecorina* CBH I structure (phenylalanine). The more-hydrophobic phenylalanine is apparently more compatible than tyrosine within this highly hydrophobic pocket.



Fig. 3. Locations of the nine verified stabilizing mutations, using the *T.emersonii* CBH I crystal structure (PDB ID 1Q9H).

Two other stabilizing mutations, S13P and N93K, are located on different strands of the same beta sheet, with the side chain of lysine 93 facing the solvent in the *T.emersonii* CBH I structure and that of S93 facing outward in the *H.jecorina* CBH I structure. The other stabilizing mutations are all in surface loops. In the *H.jecorina* CBH I structure, mutation A383Y appears to place an aromatic amino acid in the correct orientation for a favorable pi-stacking interaction with tyrosine 247, which is also present in the thermostable chimeras. Mutations S324P, D354V and T392I all substitute hydrophobic amino acids near other hydrophobic residues (phenylalanine 275, tyrosine 254 and proline 397, respectively) present in the thermostable chimeras that could render the folding free energy more favorable. In addition, when serine is present at position 324, its side chain points toward serine 254, leading to the unfavorable electrostatic repulsion between the two polar oxygen atoms which is relieved by the substitution. Mutation N93K replaces a polar side chain with a positively charged one close to several polar residues. A threonine and a tyrosine are within 6 Å of residue 93 in the *T.emersonii* CBH I structure, and the negatively polarized oxygen atom of each could have favorable electrostatic interactions with the positively charged nitrogen atom of the arginine, depending on the orientation of the side chain. From examining the crystal structures, it is not clear why mutation Y60L increases stability.

Both S13P and S324P result in substitution with proline in loops on the surface of the protein. Residue 13 is near the N-terminus, and 324 is in the middle of a long loop. Prolines have restricted conformations compared to other residues and thus lower the entropy of the denatured state, which is thermodynamically stabilizing (Matthews *et al.*, 1987). This effect is pronounced in loops that are less structured and more flexible.

The goal of this work is to generate stable fungal CBH Is and enhance the rate of cellulase-catalyzed cellulose degradation by allowing the enzyme to operate at higher temperature. Because sugar production by a CBH involves multiple steps that includes cellulase binding and diffusion along the cellulose chain (Igarashi *et al.*, 2009) along with the

elementary catalytic step, the temperature dependence of catalytic activity is difficult to predict. Cellulase specific activities have nonetheless been observed to increase with temperature up to the point of enzyme denaturation (Voutilainen *et al.*, 2008; Heinzelman *et al.*, 2009b; Mingardon *et al.*, 2011). Therefore, if activity at moderate temperatures can be maintained while the enzyme optimum operating temperature is increased, higher activity can be expected.

The benefits of enzyme thermostabilization are illustrated in the CBH I activity–temperature profiles shown in Fig. 2. At moderate temperatures (45–55°C), where the enzyme remains active during the entire assay period, the enzymes all have similar specific activities. Thus these enzymes stabilized with homologous substitutions indeed retained full catalytic function. Furthermore, their specific activities increase at higher temperatures (60–70°C). The magnitude of the increase for the CBH I (~50%) is less than the ~100% or more increase we observed for similarly stabilized CBH II enzymes (Heinzelman *et al.*, 2009a), but is nonetheless highly significant for industrial applications. The increase in specific activity with temperature, enabled by the enzyme's greater kinetic stability, illustrates the benefit of increasing an enzyme's ability to tolerate and retain function at elevated temperatures.

Conclusions

Overall, we were able to expand on our previous CBH I engineering work by using predictive methods to identify individual mutations that further stabilize thermostable fungal CBH I chimeras. Using $\Delta\Delta G$'s predicted by both FoldX and consensus sequence analysis, we identified eight individual mutations that further increased the T_{50} of the most stable chimera by 4.7°C. While these methods helped to identify stabilizing mutations, they also predicted increased stability for a larger number of mutations that proved to be either neutral or destabilizing. These results show that both methods are somewhat effective in selecting mutations for testing, but are not very reliable in predicting the effects of individual mutations. Compared with the most stable natural CBH I, this nearly 10°C increase in T_{50} translated into a corresponding 10°C increase in optimal reaction temperature and a 50% increase in total sugar production at the optimal temperature. This enzyme can be used with thermostable cellulases of other classes for biomass conversion at increased temperatures, lowering the cost of fuel production from cellulosic sources.

Supplementary data

Supplementary data are available at *PEDS* online.

Acknowledgements

The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

Funding

This work was supported by the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office.

References

- Anbar,M., Gul,O., Lamed,R., Sezerman,U.O. and Bayer,E.A. (2012) *Appl. Environ. Microbiol.*, **78**, 3458–3464.
- Bloom,J.D., Silberg,J.J., Wilke,C.O., Drummond,D.A., Adami,C. and Arnold,F.H. (2005) *Proc. Natl Acad. Sci. USA*, **102**, 606–611.
- Boer,H., Teeri,T.T. and Koivula,A. (2000) *Biotechnol. Bioeng.*, **69**, 486–494.
- Divne,C., Stahlberg,J., Reinikainen,T., Ruohonen,L., Pettersson,G., Knowles,J.K., Teeri,T.T. and Jones,T.A. (1994) *Science*, **265**, 524–528.
- Drummond,D.A., Silberg,J.J., Meyer,M.M., Wilke,C.O. and Arnold,F.H. (2005) *Proc. Natl Acad. Sci. U S A*, **102**, 5380–5385.
- Grassick,A., Murray,P.G., Thompson,R., Collins,C.M., Byrnes,L., Birrane,G., Higgins,T.M. and Tuohy,M.G. (2004) *Eur. J. Biochem.*, **271**, 4495–4506.
- Guerois,R., Nielsen,J.E. and Serrano,L. (2002) *J. Mol. Biol.*, **320**, 369–387.
- Heinzelman,P., Komor,R., Kanaan,A., Romero,P., Yu,X.L., Mohler,S., Snow,C. and Arnold,F. (2010) *Protein Eng. Des. Sel.*, **23**, 871–880.
- Heinzelman,P., Snow,C.D., Smith,M.A., *et al.* (2009a) *J. Biol. Chem.*, **284**, 26229–26233.
- Heinzelman,P., Snow,C.D., Wu,I., Nguyen,C., Villalobos,A., Govindarajan,S., Minshull,J. and Arnold,F.H. (2009b) *Proc. Natl Acad. Sci. USA*, **106**, 5610–5615.
- Higuchi,R., Krummel,B. and Saiki,R.K. (1988) *Nucleic Acids Res.*, **16**, 7351–7367.
- Igarashi,K., Koivula,A., Wada,M., Kimura,S., Penttila,M. and Samejima,M. (2009) *J. Biol. Chem.*, **284**, 36186–36190.
- Ilmen,M., den Haan,R., Brevnova,E., *et al.* (2011) *Biotechnol. Biofuels*, **4**, 30.
- Jeoh,T., Michener,W., Himmel,M.E., Decker,S.R. and Adney,W.S. (2008) *Biotechnol. Biofuels*, **1**.
- Jochens,H. and Bornscheuer,U.T. (2010) *ChemBioChem*, **11**, 1861–1866.
- Komor,R.S. (2012) *Chemistry and Chemical Engineering*. California Institute of Technology, Pasadena, CA.
- Laymon,R.A., Adney,W.S., Mohagheghi,A., Himmel,M.E. and Thomas,S.R. (1996) *Appl. Biochem. Biotechnol.*, **57–8**, 389–397.
- Matthews,B.W., Nicholson,H. and Becktel,W.J. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 6663–6667.
- Mingardon,F., Bagert,J.D., Maisonnier,C., Trudeau,D.L. and Arnold,F.H. (2011) *Appl. Environ. Microbiol.*, **77**, 1436–1442.
- Nikolova,P.V., Henckel,J., Lane,D.P. and Fersht,A.R. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 14675–14680.
- NRC (2011) *Renewable Fuel Standard: Potential Economic and Environmental Effects of U.S. Biofuel Policy*. The National Academies Press, Washington, DC.
- Pantoliano,M.W., Whitlow,M., Wood,J.F., Dodd,S.W., Hardman,K.D., Rollence,M.L. and Bryan,P.N. (1989) *Biochemistry*, **28**, 7205–7213.
- Potapov,V., Cohen,M. and Schreiber,G. (2009) *Protein Eng. Des. Sel.*, **22**, 553–560.
- Privalov,P.L. (1979) *Adv. Protein Chem.*, **33**, 167–241.
- Rodriguez-Larrea,D., Minning,S., Borchert,T.V. and Sanchez-Ruiz,J.M. (2006) *J. Mol. Biol.*, **360**, 715–724.
- Romero,P.A. and Arnold,F.H. (Submitted 2012) *PLoS Comp. Biol.*
- Sanchez-Ruiz,J.M. (1992) *Biophys. J.*, **61**, 921–935.
- Schymkowitz,J., Borg,J., Stricher,F., Nys,R., Rousseau,F. and Serrano,L. (2005) *Nucleic Acids Res.*, **33**, W382–W388.
- Steipe,B., Schiller,B., Pluckthun,A. and Steinbacher,S. (1994) *J. Mol. Biol.*, **240**, 188–192.
- Suominen,P.L., Mantyla,A.L., Karhunen,T., Hakola,S. and Nevalainen,H. (1993) *Mol. Gen. Genet.*, **241**, 523–530.
- Szjarto,N., Siika-aho,M., Tenkanen,M., Alapuranen,M., Vehmaanpera,J., Reczeya,K. and Viikari,L. (2008) *J. Biotechnol.*, **136**, 140–147.
- USDA (2010) *USDA Biofuels Strategic Production Report*. U.S. Department of Agriculture, Washington, D.C, pp. 1–23.
- USEIA (2010) *Annual Energy Review*. U.S. Energy Information Administration, Washington, D.C, pp. 1–447.
- Viikari,L., Alapuranen,M., Puranen,T., Vehmaanpera,J. and Siika-Aho,M. (2007) *Adv. Biochem. Eng. Biotechnol.*, **108**, 121–145.
- Voutilainen,S.P., Boer,H., Alapuranen,M., Janis,J., Vehmaanpera,J. and Koivula,A. (2009) *Appl. Microbiol. Biotechnol.*, **83**, 261–272.
- Voutilainen,S.P., Boer,H., Linder,M.B., Puranen,T., Rouvinen,J., Vehmaanpera,J. and Koivula,A. (2007) *Enzyme Microb. Technol.*, **41**, 234–243.
- Voutilainen,S.P., Murray,P.G., Tuohy,M.G. and Koivula,A. (2010) *Protein Eng. Des. Sel.*, **23**, 69–79.
- Voutilainen,S.P., Puranen,T., Siika-Aho,M., Lappalainen,A., Alapuranen,M., Kallio,J., Hooman,S., Viikari,L., Vehmaanpera,J. and Koivula,A. (2008) *Biotechnol. Bioeng.*, **101**, 515–528.
- Wang,Q., Buckle,A.M., Foster,N.W., Johnson,C.M. and Fersht,A.R. (1999) *Protein Sci.*, **8**, 2186–2193.