8-31-2017

Improving the Thermal Stability of Cellobiohydrolase Cel7A from *Hypocrea jecorina* by Directed Evolution

Frits Goedegebuur
DuPont Industrial Biosciences, The Netherlands

Lydia Dankmeyer
DuPont Industrial Biosciences, The Netherlands

Peter Gualfetti
DuPont Industrial Biosciences

Saeid Karkehabadi
Swedish University of Agricultural Sciences, Sweden

Henrik Hansson
Swedish University of Agricultural Sciences, Sweden

See next page for additional authors

Authors
Frits Goedegebuur, Lydia Dankmeyer, Peter Gualfetti, Saeid Karkehabadi, Henrik Hansson, Suvamay Jana, Vicky Huynh, Bradley R. Kelemen, Paulien Kruithof, Edmund A. Larenas, Pauline J. M. Teunissen, Jerry Ståhlberg, Christina M. Payne, Colin Mitchinson, and Mats Sandgren

Improving the Thermal Stability of Cellobiohydrolase Cel7A from *Hypocrea jecorina* by Directed Evolution

Notes/Citation Information
Published in *The Journal of Biological Chemistry*, v. 292, no. 42, p. 17418-17430.

Digital Object Identifier (DOI)
https://doi.org/10.1074/jbc.M117.803270

Improving the thermal stability of cellobiohydrolase Cel7A from *Hypocrea jecorina* by directed evolution

Received for publication, June 26, 2017, and in revised form, August 24, 2017 Published, Papers in Press, August 31, 2017, DOI 10.1074/jbc.M117.803270

Frits Goedegebuur[‡1], Lydia Dankmeyer[‡], Peter Gualfetti[§], Saeid Karkehabadi[¶], Henrik Hansson[¶], Suvamay Jana[‖], Vicky Huynh[§], Bradley R. Kelemen[§], Paulien Kruithof[‡], Edmund A. Larenas[§], Pauline J. M. Teunissen[§], Jerry Ståhlberg[¶2], Christina M. Payne[‖3], Colin Mitchinson[§], and Mats Sandgren[¶4]

From [‡]DuPont Industrial Biosciences, Archimedesweg 30, Leiden 2333CN, The Netherlands, [§]DuPont Industrial Biosciences, Palo Alto, California 94304, the [¶]Department of Molecular Sciences, Swedish University of Agricultural Sciences, PO Box 7015, Uppsala SE-75007, Sweden, and the [‖]Department of Chemical and Materials Engineering, University of Kentucky, Lexington, Kentucky 40506

Edited by Joseph Jez

Secreted mixtures of *Hypocrea jecorina* cellulases are able to efficiently degrade cellulosic biomass to fermentable sugars at large, commercially relevant scales. *H. jecorina* Cel7A, cellobiohydrolase I, from glycoside hydrolase family 7, is the workhorse enzyme of the process. However, the thermal stability of Cel7A limits its use to processes where temperatures are no higher than 50 °C. Enhanced thermal stability is desirable to enable the use of higher processing temperatures and to improve the economic feasibility of industrial biomass conversion. Here, we enhanced the thermal stability of Cel7A through directed evolution. Sites with increased thermal stability properties were combined, and a Cel7A variant (FCA398) was obtained, which exhibited a 10.4 °C increase in T_m and a 44-fold greater half-life compared with the wild-type enzyme. This Cel7A variant contains 18 mutated sites and is active under application conditions up to at least 75 °C. The X-ray crystal structure of the catalytic domain was determined at 2.1 Å resolution and showed that the effects of the mutations are local and do not introduce major backbone conformational changes. Molecular dynamics simulations revealed that the catalytic domain of wild-type Cel7A and the FCA398 variant exhibit similar behavior at 300 K, whereas at elevated temperature (475 and 525 K), the FCA398 variant fluctuates less and maintains more native contacts over time. Combining the structural and dynamic investigations, rationales were developed for the stabilizing effect at many of the mutated sites.

Enzymatic conversion of cellulosic biomass, a renewable feedstock rich in carbon, to fermentable sugars is an important process in the production of sustainable alternative fuels and chemicals (1). Biomass-utilizing microorganisms employ a variety of enzymes to degrade the complex carbohydrate cellulose, a polymer of β-1,4-glucose (2). One such microorganism is the filamentous fungus *Hypocrea jecorina* (also known by the name of its anamorph, *Trichoderma reesei*), which is among the most prolific producers of biomass-degrading enzymes (3). The cellulolytic system expressed by *H. jecorina* includes a complex of endoglucanases (Cel7B/EG[5] I, Cel5A/EG II, Cel5B, Cel12A/EG III, Cel45A/EG V, and Cel74A) and two cellobiohydrolases (CBHs), which processively hydrolyze from either the reducing end (Cel7A/CBH I) or the non-reducing end (Cel6A/CBH II) of cellulose chains. Together, these enzymes synergistically break down cellulose into soluble sugars (4–10). Beyond these primary cellulose-degrading enzymes, *H. jecorina* is known (10–12) to produce seven β-glucosidases (Cel1A, Cel1B, Cel3A Cel3B, Cel3C, Cel3D, and Cel3E), at least two lytic polysaccharide monooxygenases (LPMO9A/Cel61A and LPMO9B/Cel61B), a swollenin, a glucuronyl esterase, and a lyase-like protein (13, 14), all of which play a role in cellulose degradation. The four most abundant components of the cellulolytic system of *H. jecorina* (Cel7A, Cel6A, Cel7B, and Cel5A) account for >75% of the protein produced by this fungus under inducing conditions and can be secreted in excess of 50 g/liter (15). The major component, Cel7A, constitutes nearly half of total secreted protein, and it has been shown to be a major rate-limiting factor in cellulose degradation (2, 16), suggesting that GH7 enzymes play a key role in biomass-degrading fungi (17). Given its importance in biomass deconstruction, Cel7A has long been a primary component of industrial biomass-degrading enzyme mixtures (18), and *H. jecorina* is the predominant host for such applications as a result of its high-titer production of cellulases (19).

This work was supported, in part, by a subcontract from the Office of Biomass Program, within the Department of Energy Office of Energy Efficiency and Renewable Energy; the Swedish Research Council for Environment, Agricultural Sciences, and Spatial Planning (Formas; Grant 213-2013-1607; principal investigator: J. S.); and the National Science Foundation (NSF) under Grant 1552355 (principal investigator: C. M. P.). F. G., L. D., B. R. K., P. K., and P. J. M. T. are employees of DuPont Industrial Biosciences, a producer of enzymes for industrial use.

This article contains supplemental Figs. S1–S5.

The atomic coordinates and structure factors (code 5OA5) have been deposited in the Protein Data Bank (http://wwpdb.org/).

[1] To whom correspondence may be addressed. E-mail: frits.goedegebuur@dupont.com.

[2] To whom correspondence may be addressed. E-mail: jerry.stahlberg@slu.se.

[3] To whom correspondence may be addressed. E-mail: christy.payne@uky.edu.

[4] To whom correspondence may be addressed. E-mail: mats.sandgren@slu.se.

[5] The abbreviations used are: EG, endoglucanase; CBH, cellobiohydrolase; CBM, carbohydrate-binding module; GH7, glycoside hydrolase family 7; MD, molecular dynamics; MUL, 4-methylumbelliferyl-β-lactoside; NAG, *N*-acetyl glucosamine; PASC, phosphoric acid swollen cellulose; RMSD, root mean square deviation; RMSF, root mean square fluctuation; PDB, Protein Data Bank; CV, column volumes.

Industrial enzymatic conversion efficiency benefits from increases in temperature by virtue of the fundamental dependence of reaction rate on temperature. However, *H. jecorina* is a mesophilic organism that produces cellulases, including its native Cel7A, that do not have sufficient stability beyond 50 °C (20). Thus, process optimization would greatly benefit from identification of increasingly thermally stable cellulase variants, where improvements to Cel7A stability, the most abundant component of industrial enzyme mixtures (18), offer the largest overall gain.

There are four commonly implemented approaches to identify stabilizing mutations of a given protein. These approaches include (i) comparison of the amino acid sequence of a protein with that of a single more or less thermally stable homologous protein, followed by replacement of selected amino acids by site-directed mutagenesis (21, 22); (ii) site-directed mutagenesis to introduce amino acids derived from the study of the 3D structure of the protein of interest, where proteins with diverse thermal stabilities are compared and amino acid substitutions identified (23–25); (iii) random mutagenesis followed by selection (26–28); and (iv) a structure-guided approach with recombination of stabilizing fragments (SCHEMA) (29, 30). The latter two approaches rely on the ability to both express the enzyme in a microorganism and assay it using high-throughput screening methods. However, heterologous expression of Cel7A from *H. jecorina* in *Escherichia coli*, a traditional high-throughput expression host, has notoriously failed to produce either soluble or active enzyme. In yeast, expression levels are low, and the properties of expressed Cel7s have been unpredictable (31–34). The latter obstacle is probably a result of both the lack of N-terminal glutamine cyclization of Cel7A expressed in yeast and the significant variation in protein glycosylation (35), highlighting the importance of expression host on enzyme activity.

To date, approaches ii and iv have been successfully applied in engineering more thermally stable GH7 cellobiohydrolases. Heinzelman *et al.* (36) used SCHEMA recombination to increase the half-life of a chimeric GH7 expressed in a glycosylation-deficient *Saccharomyces cerevisiae* strain. The variant, which was reported to contain an average of 42 unspecified mutations, retained some hydrolytic activity on soluble cellulose at temperatures up to 70 °C. In a follow-up study, the inclusion of an additional eight mutations improved the T_{50} (temperature at which a 10-min incubation without substrate results in loss of half of the activity) to 72.1 °C (37). In a similar approach, chimeric libraries of Cel7A (not including *H. jecorina* Cel7A) were expressed in *S. cerevisiae* and found to contain stabilized variants (38). Voutilainen *et al.* have also described protein engineering efforts via approach ii, resulting in a 4.5 °C and 9 °C increase in T_m for *Melanocarpus albomyces* Cel7B and *Talaromyces emersonii* Cel7A, respectively (39, 40). Although these studies demonstrate that GH7 enzyme thermal stability can be improved through a variety of protein engineering approaches, we do not yet understand why these mutations, either singular or multiple, result in improved stability. Moreover, there has only been one report of improved thermal stability of *H. jecorina* Cel7A in the scientific literature (41).

As part of a large, broad campaign to find stabilizing mutations in *H. jecorina* Cel7A (42), we developed a host/vector system for heterologous expression of Cel7A in the filamentous fungus *Aspergillus niger* var. *awamori AP4*, using a constitutive promoter to limit the background of host proteins and potential interference from other carbohydrases. We demonstrated the successful expression of GH7 CBHs from various fungi in the *A. niger AP4* system (20). Using approaches i, ii, and iii above, a collection of *H. jecorina* Cel7A variants was generated and found to be stable by thermal inactivation screening (42). In this paper we characterize the most stable of those variants, assessing their thermal stability, residual activity after incubation at elevated temperatures using a model substrate, and activity on amorphous cellulose. We also study combinations of the variants with further improvements in stability. This combinatorial approach has proven effective in other systems (43, 44). To address the fundamental gap in understanding of structural and dynamic contributions to GH7 thermal stability, specifically that of *H. jecorina* Cel7A, we focus on the most stable Cel7A variant, FCA398, with 18 mutations, solving its crystal structure and using high-temperature molecular simulation to evaluate changes in its structural dynamics compared with wild type.

Overall, our study reveals that it is possible to make significant improvements to the thermal stability of *H. jecorina* Cel7A by the accumulation of many mutations, and through a combination of structural and dynamic approaches, we develop rationales for the beneficial effects of most of the mutations made.

Table 1
Protein T_m for selected variants of *H. jecorina* Cel7A

Variant	Mutations	T_m [a]	ΔT_m [a]
		°C	°C
FCA301	None	62.5	0
FCA330	S8P	63.7	+1.2
FCA328	G22D/S278P/T296P	63.6	+1.1
FCA335	T41I	64.2	+1.7
FCA332	N49S	63.7	+1.2
FCA347	A68T	63.7	+1.2
FCA377	N89D	63.5	+1.0
FCA374	S92T	64.4	+1.9
FCA326	S113N	64.0	+1.5
FCA375	S196T/S411F	65.3	+2.8
FCA334	P227L	65.2	+2.7
FCA373	P227A	64.8	+2.3
FCA349	D249K	64.0	+1.5
FCA376	T255P	64.4	+1.9
FCA384	E295K	64.0	+1.5
FCA372	S297L/V403D/T462I	64.5	+2.0
FCA468	N301R	63.0	+0.5
FCA369	T332Y	63.3	+0.8

[a] T_m is unfolding temperature, measured by CD spectroscopy; ΔT_m is T_m value minus the T_m value of FCA301 (wild-type Cel7A).

Results

Expression of thermostabilized H. jecorina Cel7A variants in A. niger var. awamori

To improve the thermal stability of *H. jecorina* Cel7A, mutations have been introduced in the protein molecule using protein engineering. Following high-throughput fungal expression and screening, a large set of point mutations resulting in enhanced residual activity after incubation at elevated temperature was identified (42). In the present study, a group of these were chosen for larger-scale expression in *A. niger* var. *awamori AP4* and purification for more detailed analyses (Table 1), including wild-type Cel7A (FCA301) as a control. Site-specific combination mutants were then generated, expressed, and

Table 2
T_m and $t_{1/2}$ values for selected Cel7A site combination variants

Cel7A variant	Mutations	$t_{½}$ 62 °C	$t_{½}$ 66 °C	$t_{½}$ 69 °C	T_m	ΔT_m
			min		°C	°C
FCA301	None	19.9	9.3	8.6	62.5	0
FCA353	G22D/N49S/A68T/P227L/S278P/T296P	223	ND[a]	8.3	65.9	+3.4
FCA367	S8P/G22D/T41I/N49S/A68T/S113N/P227L /D249K/S278P/T296P/N301R	596	182	53.9	69.9	+7.4
FCA398	S8P/T41I/N49S/A68T/N89D/S92T/S113N /S196T/P227L/D249K/T255P/ S278P/E295K /T296P/T332Y/V403D/S411F/T462I	866	315	75.3	72.9	+10.4

[a] Not determined.

screened, leading to a Cel7A variant (FCA398) with 18 mutations and substantially improved thermal stability. Herein, we present the results for the most thermally stable Cel7A variant, FCA398, and key variants in its development (Table 2).

Determination of protein T_m values

Thermal stability was assessed by monitoring the irreversible thermal denaturation of the purified Cel7A variants with CD spectroscopy for determination of protein T_m. The obtained T_m values and ΔT_m, compared with wild type, are shown in Tables 1 and 2. For comparison, T_m was also determined for Cel7A wild type isolated from a native *H. jecorina* culture; the T_m of natively expressed Cel7A, 62.4 ± 0.3 °C, was the same as that of FCA301 (wild-type Cel7A expressed in *A. niger* var. *awamori*), 62.5 ± 0.3 °C. All the analyzed variants exhibited higher T_m than the wild-type enzyme, as expected from their identification as more thermally stable in the residual activity screening. In general, the stabilizing effect of individual point mutations was small, where ΔT_m was in the range of +1 to +2 °C for a majority, up to +2.7 °C for P227L (FCA334), and +2.8 °C for S196F/S411F (FCA375) (Table 1). When a larger number of sites were combined, a significant enhancement of T_m was achieved, as exemplified by FCA353, FCA367, and FCA398 in which 6, 11, and 18 mutations, respectively, increased the T_m by 3.4, 7.4 and 10.4 °C (Table 2).

The contribution of single-site variants to the thermal stability of the corresponding multi-site combinations was not simply additive. That is, the ΔT_m of combinatorial variants was consistently lower than the direct sum of ΔT_m values for the contributing single-site variants. For instance, one of our best intermediate variants (supplemental Fig. S1), FCA367, with 11 mutations, had a measured ΔT_m of +7.4 °C, whereas the summative ΔT_m of the individual 11 mutations would be +11.4 °C if thermal stability gains were purely additive. To rule out the possibility that some mutations may counteract thermal stability when combined with each other, all 11 sites of FCA367 were recombined using QuikChange, generating a library of Cel7A mutants with different combinations of these sites. Around 2000 clones, with an average of 8 mutations/clone (990 possibilities), were screened for retention of activity at high temperature. The most thermally stable clone contained all 11 sites, suggesting that all of the point mutations of FCA367 contribute to increased thermal stability. FCA367 was further improved by the addition or substitution of stability-associated mutations using fusion PCR techniques to generate site-specific combination mutants with enhanced thermal stability. The most improved variant was FCA398, which exhibits a T_m value of 72.9 °C (*i.e.* 10.4 °C higher than wild-type Cel7A) (Table 2).

Thermal inactivation and half-life, $t_{1/2}$, at elevated temperature

Residual activity measurements were further used to determine inactivation $t_{½}$ for the site-combined variants FCA353, FCA367, and FCA398 compared with wild-type FCA301. The activity of the purified enzymes on 4-methylumbelliferyl-β-lactoside (MUL) was measured at 50 °C, before and after incubation at elevated temperatures, 62, 66, and 69 °C. Samples were taken at regular time points up to 20 h for activity measurements on MUL. This experimental design measured irreversible thermal inactivation. The half-life is defined as the time taken for a decrease of activity to 50% of the initial activity. The results show that the Cel7A variants were more stable than wild type, in agreement with measured T_m values. Even a seemingly small increase in T_m substantially improves the lifetime of the enzyme at high temperature (in the case of FCA353), and for FCA398, the most stable variant, we see large improvements (Table 2); the half-life is 44-, 34-, and 9-fold enhanced compared with wild type at 62, 66, and 69 °C, respectively.

Degradation of phosphoric acid swollen cellulose (PASC) by Cel7A variant FCA398 at elevated temperature

Wild-type FCA301 and the most stable Cel7A variant, FCA398, were incubated at 53, 65, and 75 °C with 1% PASC as substrate, and the cellulose degradation was monitored over the course of 3 days (72 h) by quantification of released soluble sugar by HPLC (Fig. 1). FCA398 is clearly more active and gave higher sugar yields at all three temperatures than did FCA301 at any temperature. For FCA301, the activity was highest at 53 °C and slightly lower at 65 °C, whereas the enzyme was practically inactive at 75 °C. FCA398 retained relatively high activity even at 75 °C, leveling off after ~48 h, and its highest activity was observed at 65 °C.

3D structure of the catalytic domain of the Cel7A FCA398 variant

To understand the structural basis for improved thermal stability over wild type, the structure of the deglycosylated catalytic domain of FCA398 was determined by X-ray crystallography (Fig. 2). The protein crystallized in space group $P2_12_12$ with two protein chains in the asymmetric unit, in a new crystal form with unit cell parameters and crystal packing that are different from that of any previous structure of *H. jecorina* Cel7A. The structure of the FCA398 catalytic domain was solved by molecular replacement and was refined at 2.1 Å resolution to final R and R_{free} values of 19.3 and 23.9%, respectively. Statistics on data collection, refinement, and the final structure model are given in Table 3. Ramachandran plots are provided in

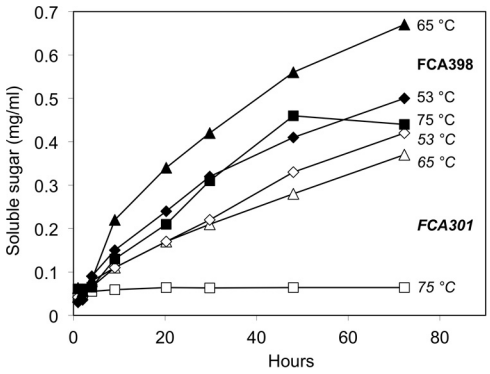

Figure 1. Hydrolysis of PASC by the most stable Cel7A variant FCA398 (*filled symbols*) compared with wild type (FCA301; *open symbols*) at 53 °C (*diamonds*), 65 °C (*triangles*), and 75 °C (*squares*), using 1% PASC and 0.5 mg enzyme/g cellulose at pH 5.5. The amount of soluble sugar was determined by HPLC.

supplemental Fig. S2). The structure model has been deposited with the Protein Data Bank (PDB accession code 5OA5).

Both protein chains, A and B, of the asymmetric unit contain all amino acid residues (positions 1–434) of the FCA398 catalytic domain and three *N*-acetyl glucosamine residues (NAG), which are covalently linked to Asn^{64}, Asn^{113}, and Asn^{270}, respectively. Glycosylation at Asn^{64}, as observed in this *Aspergillus*-expressed protein, has not been seen previously in *H. jecorina* Cel7A structures or glycosylation analyses (45), despite the presence of an *N*-glycosylation sequence motif at this site also in the wild-type enzyme. At Asn^{113}, a new *N*-glycosylation motif is introduced by the S113N mutation in FCA398.

The structures of the two chains are practically identical with a root mean square deviation (RMSD) of 0.13 Å. The FCA398 structure is very similar to previously deposited *H. jecorina* Cel7A catalytic domain structures; superposition with the Cel7A-cellononaose complex (PDB code 4C4C (46)) gives RMSD values of 0.35 and 0.34 Å relative to FCA398 chain A and B, respectively. There is only one loop region, 244–254, where the main chain of FCA398 deviates significantly. This loop bends toward and partially encloses the catalytic center of the enzyme. It has previously been referred to as the "exo-loop" (47) and, more recently, loop B3 (48), and flexibility in the loop has been reported previously (47, 49, 50). The largest backbone shift between the FCA398 and 4C4C structures is seen at residue 249 (4.6 Å between Cα atoms), which is the site of the D249K mutation (Fig. 2, *A* and *C*).

Molecular dynamics (MD) simulations of Cel7A FCA301 and the FCA398 variant

Molecular dynamics simulations of wild-type Cel7A and FCA398 at 300, 475, and 525 K suggest that the 17 mutations in the FCA398 catalytic domain improve thermal stability by retaining tighter packing of the core domain at elevated temperatures. This is illustrated by two separate analyses of the triplicate simulation trajectories: root mean square fluctuation (RMSF) of the protein backbone and variation in native contacts maintained over time. First, the RMSF of the protein backbone was determined for each of the 18 simulation trajectories. The RMSF of the triplicate simulations were averaged together to obtain an average RMSF of a given enzyme at a given temperature (Fig. 3). The RMSF measures deviation of a residue from its time-averaged position over the course of a simulation. Higher RMSF in the protein backbone typically indicates a more flexible, potentially less thermally stable, region of the protein. At 300 K, the average RMSF of wild type and FCA398 are practically indistinguishable. However, with increasing temperature (475 and 525 K), the FCA398 variant fluctuates less in comparison with wild type, implying that the 17 selected mutations in the catalytic domain increased the stability of the protein. At 475 K, the stabilizing effect of the set of mutations had the most significant impact in the following residue ranges: 243–256, 310–358, and 378–396. At 525 K, we observe the same stabilization effects as in the set of simulations at 475 K, but also the region between 260 and 285 benefits from the stabilizing effects of the FCA398 mutations.

Quantitative determination of the total number of native contacts from MD simulations illustrates that FCA398 retained important molecular interactions within the core of the protein more effectively than wild-type Cel7A when faced with thermal stress. First, the number of native contacts formed by each residue was identified. Here, a native contact was defined as any amino acid whose side chain center of geometry was within 6.5 Å of the reference amino residue's Cα. The total number of native contacts (Fig. 4) is the sum of the native contacts formed by all residues in the protein. The total numbers of native contacts formed by both wild-type Cel7A and FCA398 were identical at the 300 K reference temperature over the entirety of the 50-ns MD simulations, although only 15 ns is shown in Fig. 4 for comparison with high temperature behavior. As expected at high temperatures, the protein began to partially unfold at ~7 ns, destroying native contacts present at 300 K. Regions having low initial numbers of native contacts began to unfold first during high-temperature simulations (supplemental Fig. S3), corresponding well with the increasing regional fluctuation described above.

The relative difference in number of native contacts retained over the 15-ns high-temperature simulations is easiest to see by examining the difference in the number of native contacts formed by each residue over the entire simulation. For both 475 and 525 K, the native contact heat maps of wild-type Cel7A and FCA398 have been subtracted from each other to reveal regions in which FCA398 maintains a greater number of native contacts at high temperature (supplemental Fig. S4). The region between residues 320 and 340 appears to benefit the most from the stabilizing effect of FCA398 mutations, forming and maintaining 2–6 new native contacts in many locations through the region. The FCA398 regions exhibiting the highest number of maintained native contacts relative to wild type largely correspond to the regions of decreased flexibility identified from RMSF. Notably, several of the FCA398 point mutations (P227L, D249K, T255P, and T332Y) were made in these regions. Interestingly, small regions around residues 160 (525 K) and 250 (475 K) exhibit fewer native contacts in FCA398. As described above, the increase in melting temperature due to an added mutation was never simply additive, and this localized destabilization of the FCA398 variant suggests a mechanism for this

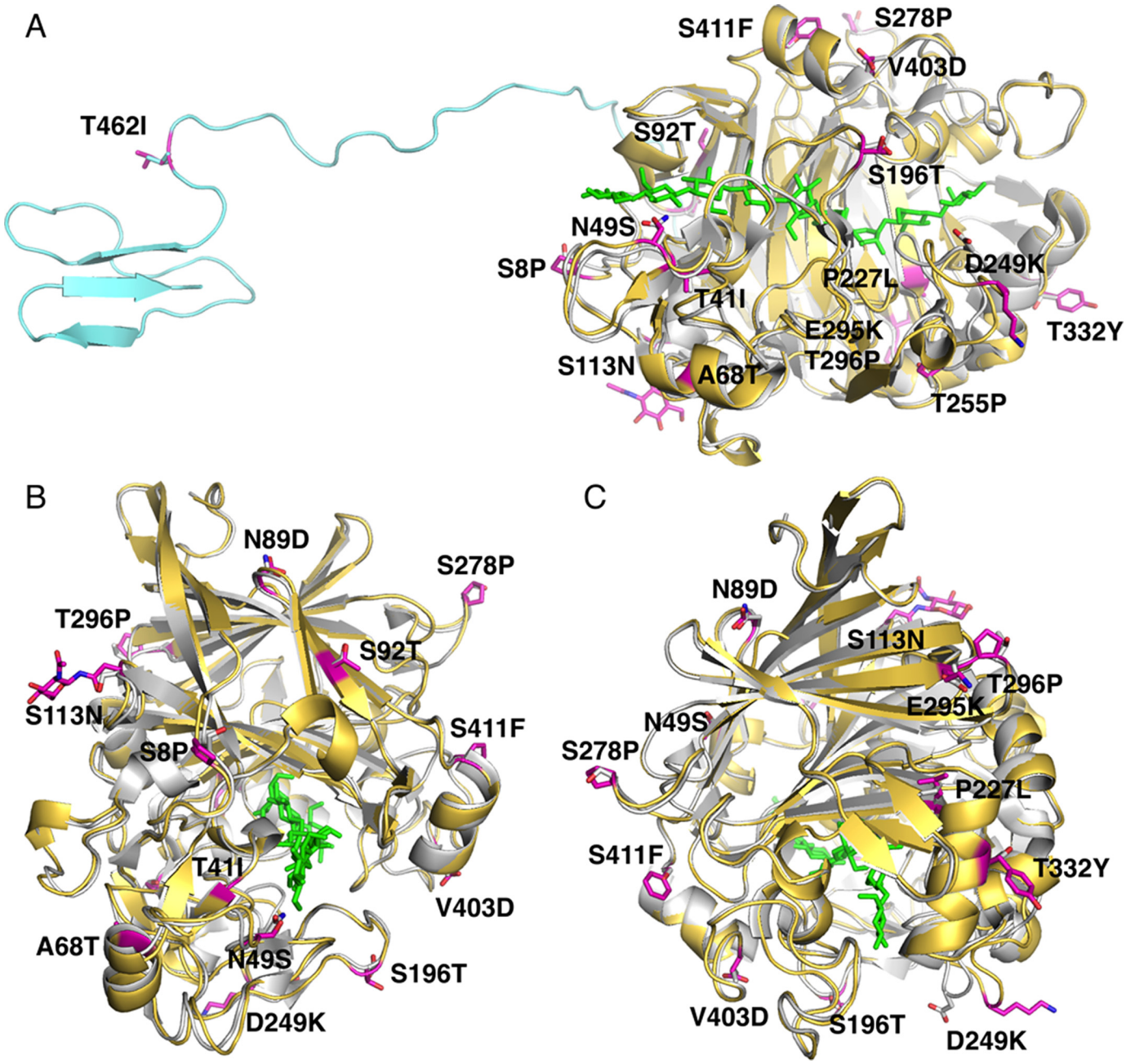

Figure 2. Overview of the 18 mutation sites (*magenta*) in the *H. jecorina* Cel7A FCA398 variant and comparison of the crystal structure of the catalytic domain of FCA398 (*yellow*) with that of Cel7A (*light gray*) in complex with cellononaose (*green*; PDB code 4C4C) (83). *A*, the linker–CBM region (*light blue*) from a previous model of the full-length Cel7A enzyme from MD simulation (102) shows the location of the T462I mutation at the linker–CBM junction. *B*, view showing the cellulose-binding tunnel entrance side of the catalytic domain. *C*, view showing the product binding side of the catalytic domain.

observation, namely that overall stabilization comes at the cost of localized regions of destabilization.

Discussion

In this work, we examine stabilizing mutations of *H. jecorina* Cel7A and show that the accumulation of such mutations is an effective strategy in the creation of more stable variants. Although the increase in stability, as measured by thermal melting T_m, by adding stable variants together was not as large as a purely arithmetic addition, the approach was effective in allowing us to progress from single variants with an increase in T_m of up to 2.7 °C (relative to wild type) to combinatorial variants with T_m increases of 3.4 °C, of 7.4 °C and, by accumulating 18 mutations (in FCA398), of 10.4 °C above the starting point. This stability increase led to greatly reduced rates of thermal inactivation and increased hydrolysis of cellulose (PASC) at elevated temperatures.

Most of the 18 point mutations introduced in the FCA398 variant are located on the surface of the protein and distributed over the whole molecule (Fig. 2). The side chains at six of these sites (N49S, N89D, S92T, S196T, D249K, and T332Y) point out into the surrounding solution, and there are no obvious structural explanations for why these mutations might stabilize the molecule.

Table 3

X-ray data collection and refinement statistics for the structure of *H. jecorina* Cel7A variant FCA398

Values were calculated with O (98), MOLEMAN (99), LSQMAN (100), and Refmac version 5.0 (76). Values for the outer shell are given in parentheses.

Parameter	Value
PDB code	5OA5
Diffraction source	ESRF ID23-2
Wavelength (Å)	0.8726
Temperature (K)	100
Detector	CCD MAR225
Rotation range per image (degrees)	0.5
Space group	$P2_12_12$
a, *b*, *c* (Å)	102.94, 92.10, 102.20
α, β, γ (degrees)	90, 90, 90
Resolution range (Å)	41–2.1 (2.21–2.10)
Total no. of reflections	211,322
No. of unique reflections	56,990
Completeness (%)	99.6 (99.6)
Redundancy	3.71 (3.73)
$\langle I/\sigma(I)\rangle$	7.2 (3.0)
R_{meas} [a]	0.15 (0.47)
Resolution range in refinement (Å)	40–2.1 (2.15–2.10)
σ cut-off	2.0
No. of reflections, working set	53,983 (3975)
No. of reflections, test set	2882 (183)
Final R_{cryst}	0.197 (0.256)
Final R_{free}	0.238 (0.306)
No. of non-hydrogen atoms	
Protein	6480
Non-protein	90
Water	546
RMSD	
Bonds (Å)	0.013
Angles (degrees)	1.65
Average *B* factors (Å²)	
Overall	17.7
Protein	16.9
Water	25.0
Stringent Ramachandran outliers (%) [b]	1.7

[a] $R_{meas} = \Sigma \sqrt{(n/(n-1))}|IhI - \langle Ih\rangle|/\Sigma|\langle Ih\rangle|$.
[b] According to the stringent boundary definition of Kleywegt and Jones (101).

All mutations are found in the catalytic module except one, T462I, which is located at the junction between the linker peptide and the CBM (Fig. 2*A*). Thr^{462} corresponds to the first residue, Thr^{1}, in the NMR structure of the CBM from *H. jecorina* Cel7A (PDB code 1CBH (51)). Replacement with isoleucine at this site increases hydrophobic interaction with surrounding residues ($Pro^{16/477}$, $Val^{18/479}$, and $Ala^{20/481}$; 1CBH/FCA398 numbering) and could, thereby, strengthen the linker–CBM connection. However, T462I appeared neither as a single point variant nor in FCA367. Accordingly, the results suggest that T462I is an unlikely contributor to an overall increase in thermal stability for the FCA398 full-length protein.

None of the mutations in the FCA398 catalytic domain are at sites that are completely conserved in the phylogenetically diverse GH7 CBH sequences, examined in the recent paper by Hobdey *et al.* (52). Only one residue, Asn^{49}, is conserved among the known GH7 CBH structures. Five of the specific amino acid substitutions in FCA398 (Leu^{227}, Pro^{255}, Lys^{295}, Pro^{296}, and Tyr^{332}) are not found in the sequences examined by Hobdey *et al.* (52). Of the remainder, eight (Pro^{8}, Thr^{68}, Asp^{89}, Thr^{92}, Asn^{113}, Thr^{196}, Asp^{403}, and Phe^{411}) are found among the GH7 CBHs with known structures, and four more (Ile^{41}, Ser^{49}, Lys^{249}, and Pro^{278}) are found in the larger set of sequences. Only one of the mutated sites in FCA398 is common with mutations reported in *H. jecorina* Cel7A literature; the T255P site in FCA398 corresponds to T260V in the study by Smith *et al.* (41).

The catalytic domain of Cel7A is built around two β-sheets packed against each other to form a β-sandwich that curves around the active site. Long loops extend from the β-sheets and fold into irregular loop regions, a few α-helices, and short β-strands at the surface of the protein; this is where the MD simulations show elevated backbone fluctuations and stabilization of FCA398 relative to wild type (Fig. 3). The largest differences between FCA398 and wild type at high temperature are seen in surface regions clustered around the side of the protein that harbor the product-binding site (to the *right* in Fig. 2*A*).

Mutation P227L

Near the product site, we find the P227L mutation, which gave the highest ΔT_m of the single point mutations analyzed (+2.7 °C; Table 1). It is the only mutation site in FCA398 where the side chain is completely buried within the protein, sitting on a β-strand anti-parallel to the strand that carries the catalytic residues. The flanking residues, Thr^{226} and His^{228}, point into the product subsite +1, whereas the 227 side chain points in the opposite direction, into the hydrophobic core (Fig. 5). The mutation does not appear to cause significant conformational changes, although the β-strand is slightly shifted at Leu^{227} in FCA398 compared with Pro^{227} in the 4C4C structure (0.8 Å between Cα atoms). The larger leucine side chain seems to fill the void better than proline, with more van der Waals contacts and tighter hydrophobic packing with surrounding side chains (Met^{213}, Ile^{300}, Leu^{326}, Leu^{349}, and Phe^{352}; Fig. 5). The P227L mutation probably contributes to the observed stabilization of the 310–358 segment by strengthening the hydrophobic interaction with several hydrophobic residues in that segment. Furthermore, the peptide nitrogen at Leu^{227} is accessible for a new hydrogen bond (3.0 Å) to the backbone oxygen of Cys^{261} in the short 261–263 β-strand at the edge of the inner β-sheet. Cys^{261}, in turn, is bound by a disulfide bridge to Cys^{331} (next to the T332Y mutation) in the 328–338 α-helix at the surface (Fig. 5).

The *V403D* and *S411F* sites are at either end of a surface helix (at the top of Fig. 2*A*), which is among the regions that unfold first and where stabilization of FCA398 is indicated by the high-temperature MD simulations. The V403D mutation introduces a new hydrogen bond in FCA398 between the Asp^{403} side-chain and the backbone amide of Gln^{406} in the helix. This type of hydrogen bonding, referred to as N-terminal capping, stabilizes the α-helix (53, 54), which in turn should stabilize the whole protein (55). At the other end of the helix, the S411F mutation introduces an aromatic residue partially buried underneath the side chain of Gln^{410}. The Phe^{411} in FCA398 provides a larger hydrophobic interface with underlying amino acids (Pro^{137}, Leu^{140}, and Val^{407}), which should anchor the helix more firmly at the surface of the catalytic domain. Interestingly, an aromatic residue at this position is also present in the documented thermostable cellobiohydrolases *M. albomyces* Cel7B and *Thermoascus aurantiacus* Cel7A (39, 56).

Tunnel entrance region

Near the entrance to the cellulose binding tunnel, at subsite −7, the extended B1 loop (residues 41–63) and the following short α-helix (residues 64–70) cover the protein surface. Three of the FCA398 mutations, T41I, N49S, and A68T, are found in

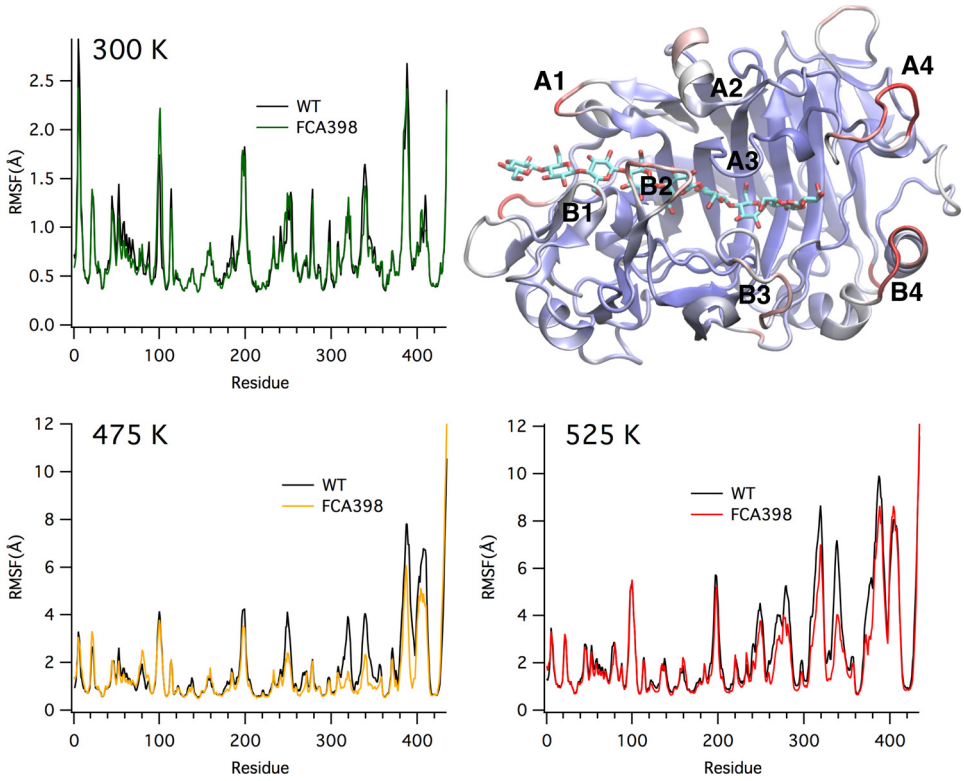

Figure 3. RMSF of the wild-type Cel7A and FCA398 variant backbone from molecular dynamics simulations at 300, 475, and 525 K. In all cases, the RMSF shown was determined by averaging values from three independent simulation trajectories. At 300 K, both wild type and the variant exhibit nearly identical dynamic flexibility, but at high temperatures, wild-type Cel7A is prone to higher backbone fluctuation relative to the FCA398 variant. In the *top right panel*, the wild-type Cel7A structure is shown *colored* by its RMSF at 300 K. The *red regions* represent backbone segments with the greatest fluctuation, whereas *blue* indicates the lowest fluctuation. The RMSF values shown on the structure have been scaled to a maximum of 2.5. Loops are *labeled* in *black type* on the structure.

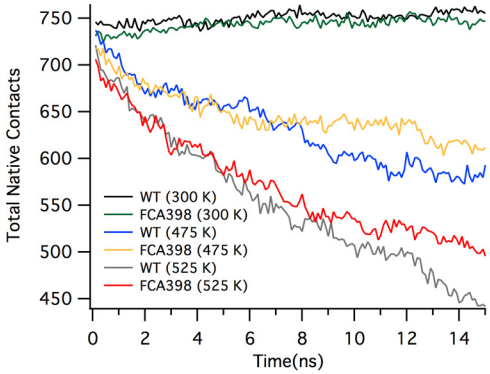

Figure 4. Total number of native contacts retained by wild-type Cel7A and FCA398 mutant over the course of MD simulations at 300, 475, and 525 K. The *lines* represent the average of three independent MD simulations for each enzyme at a given temperature.

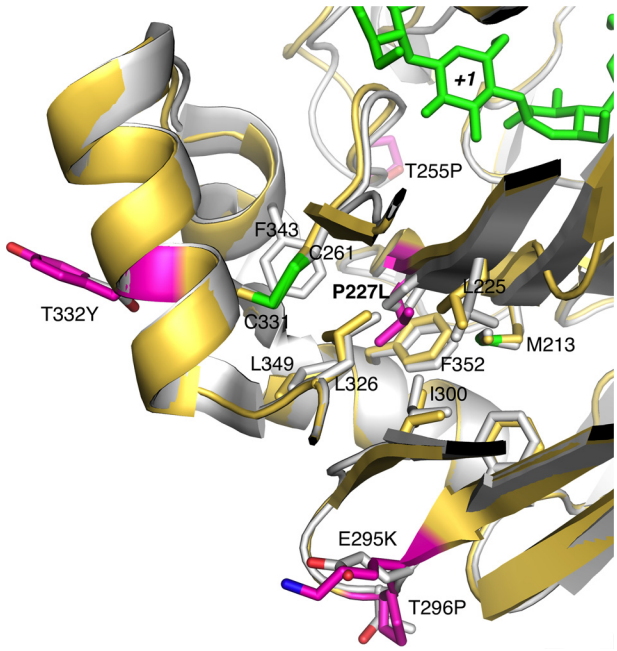

Figure 5. Comparison of the Cel7A FCA398 (*yellow*) and 4C4C (*white*) structures around the P227L mutation site. Selected side chains are shown in *stick representations*, with mutated residues in FCA398 in *magenta*.

this region. Residue Thr[41] is pointing into and is partially buried in a hydrophobic pocket in the structure, surrounded by Tyr[51], Trp[56], Leu[73], Lys[181], Phe[182], and the Cys[50]–Cys[71] disulfide bridge. The replacement of Thr[41] with an isoleucine in FCA398 displaces two water molecules and introduces several new hydrophobic contacts with the surrounding hydrophobic amino acids. These new hydrophobic contacts should contribute to the higher stability of the variant protein, as indicated by the 1.7 °C increase of T_m for the single T41I mutation (Table 1). Our MD simulations also demonstrated that, at high temperatures (supplemental Fig. S3), residue 41 in FCA398 retained more native contacts than in the wild type, indicating more favorable interactions with the neighboring amino acid residues.

The A68T site is located in the 64–70 α-helix, partially buried at the interface between the helix, the N-terminal pyroglu-

tamate residue, one loop turn formed by residues 183–185, and Thr^{160} in another loop turn at the base of loop B2. The FCA398 Thr^{68} side chain makes additional van der Waals contacts with surrounding residues and forms two new hydrogen bonds (with the backbone carbonyl oxygen at position 64 within the helix and with the side chain of Gln^{182}, which in turn bonds to Thr^{160}). The A68T mutation is thus likely to anchor the helix more firmly at the surface and may also contribute to stabilization of the neighboring loop regions. The MD simulations revealed that the A68T mutation in FCA398 variant significantly increased hydrogen bond interactions with Gln^{186} and Ala^{187} at native temperature, as well as at the two high temperatures, in comparison with the wild type. These newly formed interactions are assumed to provide additional stability to the disulfide bridge between Cys^{61} and Cys^{67}. Together, these contributions result in an increase of T_m by about 1.2 °C in the single A68T mutation (Table 1).

Surface loop turns

13 of the FCA398 mutations are at the surface of the catalytic module, and 10 of these sites are in, or close to, loop turns at the surface (S8P, N49S, N89D, S113N, S196T, D249K, T255P, S278P, E295K, and T296P). All four of the new proline residues introduced into FCA398 are found in such surface turns. The importance of proline residues in thermal stability of enzymes has been described by Matthews *et al.* (57). As a prime example, Muslin *et al.* (58) showed the effect of proline insertions on the thermostability of a barley α-glucosidase. Introduction of proline residues in FCA398 at these four positions, S8P, T255P, S278P, and T296P, probably stabilizes the Cel7A molecule by restricting main-chain flexibility and, thus, the mobility at these turns. MD simulations suggest that the T255P and S78P mutations have the largest impact on flexibility (Fig. 3). The S8P, T255P, and T296P mutations also provide a larger area of hydrophobic contact with underlying protein atoms, which may contribute further to stabilization of the respective surface turns. The T296P site is in a hairpin turn (residues 295–298), which also carries the E295K mutation. The Lys^{295} side chain introduced in FCA398 forms a new salt bridge with Glu^{325}, which may add to stabilization of the 295–298 turn and also contribute to the stabilization effect on the 310–358 region seen in the high-temperature MD simulations (Fig. 3 and supplemental Fig. S4).

New N-glycosylation at S113N site

One extra NAG residue in the FCA398 structure compared with the wild-type Cel7A structure was found covalently bound to Asn^{113} (Fig. 6). The S113N mutation, also at a surface hairpin turn, introduces an additional consensus glycosylation motif (Asn-Asp-Thr) in the amino acid sequence, which has been recognized by the *N*-glycosylation machinery in the endoplasmic reticulum of the expression host *A. niger* to produce a new glycosylation. This single mutation resulted in an increase in T_m of 1.5 °C (Table 1). Mutating this residue to an aspartic acid did not result in a significant increase in T_m compared with wild type (data not shown), suggesting that the stability effect of the S113N mutation may be the result of an altered glycosylation pattern of the protein. Interestingly, an *N*-glycosylation motif is present at the corresponding position in the GH7 CBHs from the close relative *Trichoderma harzianum*, as well as the distantly related social amoeba *Dictyostelium purpureum* and *D. discoideum*. An attached NAG residue is indeed visible at this site in the structures of the two former enzymes (52, 59).

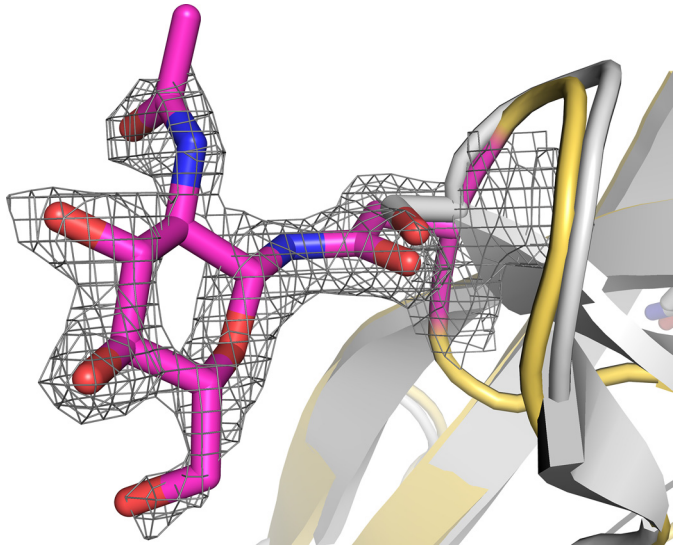

Figure 6. The S113N mutation in the Cel7A FCA398 variant introduces a new *N*-glycosylation site at a surface hairpin turn. An *N*-acetyl glucosamine residue attached to Asn^{113} is visible in the electron density map of the FCA398 crystal structure. The $2F_oF_c - F_c$ map is contoured at a σ level of 1.8. The Cel7A 4C4C structure is in *light gray*, and FCA398 is shown in *yellow* with mutated sites in *magenta*.

Role of cystine residues in stability

A striking observation is that several mutations are in the vicinity of a cystine residue and interact, either directly or indirectly, with the cystine residues. N49S and A68T are near Cys^{50}, Cys^{61}, Cys^{67}, Cys^{71}, and Cys^{72}, within the 41–74 loop. Four of these cystine residues make short disulfide bridges (Cys^{50}–Cys^{71} and Cys^{61}–Cys^{67}) within the loop, and the fifth, Cys^{72}, makes a bridge with Cys^{4} at the N terminus of the enzyme. Another example is at the apex of a loop extending from residue Thr^{231} to Asp^{259}, where a threonine has been replaced with a proline at position 255. The rigidity of this loop, probably increased by this mutation, in turn, stabilizes the position of Cys^{256}, in a disulfide bond with Cys^{230}. MD simulations support the proposal that the T255P mutation stabilized the protein and the disulfide bond between Cys^{230} and Cys^{256} (Fig. 3). The single mutant T255P improves T_m by around 2 °C (Table 1).

Interactions with cystines could be of importance for the thermal stability of Cel7A; some of the previously determined structures of *H. jecorina* Cel7A, solved by our group, have shown that the thiol group of some of these cystines can have multiple conformations. The N49S and A68T mutations, for example, may contribute to FCA398 stability by stabilizing low-energy conformations of the disulfides in the 41–74 loop. The *H. jecorina* Cel7A molecule contains 10 disulfide bridges in the catalytic domain and 2 in the CBM (SwissProt P62694). Disulfide bridges between cystine residues can also be important factors in the thermal stability of enzymes. The cystines involved in disulfide bridges represent about 5% of the residues of the Cel7A catalytic domain, leading one to speculate that

disulfides are important for the thermal stability of Cel7A; this idea is supported by examples from the literature. For example, Voutilainen *et al.* (39) showed that introducing a disulfide bridge found in wild-type *H. jecorina* Cel7A into the thermostable *M. albomyces* Cel7B increased the unfolding temperature of this enzyme by 4 °C. Although no disulfides were formed or broken during our experiments or simulations, presumably these covalent bridges limit the possible motions in their vicinity and could indicate regions that would otherwise be sites of local unfolding on the path to thermal denaturation. Thus, mutations near disulfides could further stabilize vulnerable regions and have an indirect effect by stabilizing the disulfide conformation.

In conclusion, we have developed a robust pathway of mutagenesis of the *H. jecorina* Cel7A molecule by combining mutated sites and adding their individual influences to improve the thermal stability of Cel7A. The best Cel7A variant obtained, FCA398, inactivated much more slowly than the wild type at elevated temperatures, having a half-life 44 times that of wild type at 62 °C, and had an increase of 10.4 °C in its thermal denaturation melting point. We have shown that this improvement in the thermostability of the Cel7A molecule enables degradation of amorphous cellulose at elevated temperatures, up to 75 °C. The variant FCA398 contained 18 mutated sites compared with the wild-type enzyme, and we developed rationales for the stabilizing effects of most of the mutations on the Cel7A molecule by applying structural and molecular dynamics analyses.

Experimental procedures

H. jecorina Cel7A combinatorial variant construction

Sites with increased thermal stability were identified and selected by screening libraries of *H. jecorina* Cel7A mutants heterologously expressed in *A. niger* var. *awamori* AP4 (42). Residual activity of small-scale (100 μl) culture supernatants was assessed by measuring cellobiohydrolase activity against MUL (Sigma, M2405) at 50 °C, as described previously (20), before and after heat treatment at elevated temperature (62 °C, for initial rounds, or 69 °C, as more stable variants were introduced as library starting points) for 4 h.

Specific mutagenized primers were developed, and mutated sites were combined using fusion PCR techniques as described (60). The PCR fragments containing the mutagenized Cel7A gene *cbh1* were cloned in *E. coli* strain MAX Efficiency DH5α (Invitrogen), using plasmid pDONR^{TM}201 (Invitrogen), and then transferred to the *E. coli*/*A. niger* shuttle expression vector plasmid pRAXdes, which was subsequently transformed into *A. niger* var. *awamori* AP4 (61), as described previously (20, 62). General recombinant DNA procedures were adapted from the literature (63). The mutated genes were sequenced by BaseClear Holding BV (Leiden, The Netherlands), and sequence data were analyzed using the software package VectorNTI®. The nucleotide sequences were translated into protein sequence and compared with the wild-type sequence to identify the mutated sites. The selected Cel7A variants that contributed to the design of the combinatorial variant FCA398 are listed in Table 1. Table 2 lists combinatorial variants. Variant FCA301 is wild-type *H. jecorina* Cel7A expressed in *A. niger* var. *awamori* as a control.

Recombination of mutated sites of variant FCA367

All 11 mutated sites of FCA367 were separately introduced into 5′-phosphorylated mutagenized primers. These 11 primers were then used to mutagenize Cel7A to produce a library of mutants with several different combinations, using the QuikChange® multisite-directed mutagenesis kit (Stratagene, La Jolla, CA, catalog no. 200518). A total of around 2000 clones, containing an average of 8 mutations/clone (990 possibilities), were screened for thermal stability, as measured by retention of activity after incubation at high temperature (69 °C).

Expression and purification of selected Cel7A variants

The *A. niger* var. *awamori* transformants were grown in 500-ml shake flasks for 3 days at 37 °C (62). Expressed Cel7A variants were then purified from culture supernatants by hydrophobic interaction chromatography on phenyl-Sepharose resin (GE Healthcare catalog no. 17-0973-05) equilibrated with 5 column volumes (CV) of 0.020 M sodium phosphate, 0.5 M ammonium sulfate at pH 6.8. Ammonium sulfate was added to the supernatants to a final concentration of 0.5 M, and the pH was adjusted to 6.8. After filtration, the supernatant was loaded onto the column. The column was washed with 10 CV of equilibration buffer and then eluted with a 10-CV gradient from 0.5 to 0 M ammonium sulfate in 0.020 M sodium phosphate, pH 6.8. Cel7A eluted approximately mid-gradient. Fractions were collected and pooled on the basis of SDS-PAGE analysis.

Determination of T_m

Protein T_m values were determined for the purified enzymes by CD spectroscopy, as described previously (20). The thermal unfolding of the Cel7A variants was not reversible; in our measurements of Cel7A melting points, the native, starting, CD spectrum (and, specifically, the 230 nm value) was not recovered upon cooling.

Thermal inactivation half-life measurements

For determination of irreversible inactivation $t_{1/2}$, the purified variants, FCA301, FCA353, FCA367, and FCA398, were incubated at 62, 66, and 69 °C in a microtiter plate incubator (iEMS^{TM} microplate incubator/shaker HT, Thermo Fisher Scientific). Samples were taken at regular time intervals up to 20 h, cooled on ice for 10 min, and assayed for residual activity on MUL (measured at 50 °C, as described (20)). The half-life was derived from the slope (k) of the natural logarithm of residual activity *versus* time using the formula $t_{1/2} = \ln 2/k$.

Cellulose conversion assay using PASC as substrate

PASC (1% (w/v)) (64, 65) prepared from Sigmacell type 20 (Sigma catalog no. 9004-34-6) was incubated with FCA301 (wild type) or variant FCA398 (0.5 mg enzyme/g cellulose), in 200 μl of 50 mM sodium acetate buffer, pH 5.5, in sealed microtiter plates at 53, 65, and 75 °C in an Eppendorf Thermomixer shaking at 700 rpm. Samples were taken at regular intervals, and soluble sugars were quantified by HPLC as described (66).

Preparation and crystallization of Cel7A variant FCA398 catalytic domain

Before crystallization, the *H. jecorina* Cel7A FCA398 variant enzyme (expressed in *A. niger* var. *awamori* AP4 and purified as described above) was deglycosylated and treated with papain to remove the C-terminal peptide linker and carbohydrate-binding module from the full-length protein, using a procedure similar to that described previously (67). The FCA398 protein, in 100 mM sodium acetate, pH 5.0, and 20 mM zinc acetate, was mixed with α-mannosidase (Sigma-Aldrich) and endo-*N*-acetyl glucose aminidase H to a final concentration of 40:2:1 (Cel7A/α-mannosidase/endo-*N*-acetyl glucose aminidase H, w/w/w) and incubated at room temperature for 24 h. Activated papain (Sigma-Aldrich) was then added to a final concentration of 50:1 (Cel7A/papain, w/w). After overnight incubation at room temperature, the buffer was changed to 20 mM sodium acetate, pH 5.0, and the protein was loaded onto a Source 30Q anion exchange column (GE Healthcare) and eluted by a linear 0.1–1.0 M sodium chloride gradient. Fractions containing FCA398 catalytic domain were collected, and after buffer exchange to 25 mM sodium morpholine ethane sulfonic acid, pH 6.0, the protein was concentrated to 12 mg/ml. The purity of the protein was >95%, as determined by SDS-PAGE. Note that one of the mutations in FCA398, T462I, is in the linker and, hence, is removed by this process.

Initial screening for crystallization conditions for the FCA398 variant was performed with the JCSG+ screen (Qiagen). Well-diffracting crystals were obtained in drops with 30% JCSG+ solution D9 (25.5% PEG 4000, 0.17 M ammonium sulfate, and 15% glycerol) and 70% protein solution (~12 mg/ml), equilibrated against D9 solution at 20 °C (68). Large, single, square-shaped crystals grew to a size of 10 × 200 × 200 μm in sitting drops within 1 week. The crystals belong to the space group $P2_12_12$ with the cell dimensions $a = 102.9$ Å, $b = 92.1$ Å, $c = 102.2$ Å and have a calculated V_m of 2.48 (69, 70), a solvent content of 50.5%, and two Cel7A molecules in the asymmetric unit.

X-ray data collection, structure solution, and refinement

Single crystals were picked with cryoloops and flash-frozen in liquid nitrogen. X-ray diffraction data were collected at 100 K on beam line ID23-2 at the European Synchrotron Radiation Facility (Grenoble, France). Data collection and processing statistics for the final structure are given in Table 3. The data were integrated with XDS and scaled with SCALA in the CCP4 package (71–73). 5% of the reflections were set aside for calculation of R_{free} (74). The structure of the *H. jecorina* Cel7A FCA398 catalytic domain was solved by molecular replacement using PHASER with a wild-type *H. jecorina* Cel7A structure as the search model (PDB code 2V3I (75)). The structure was refined with alternating cycles of model building using COOT and maximum likelihood refinement in Refmac version 5.0 (76). Most water molecules in the structure model were located automatically using water picking protocols in the refinement program. These water molecules were then manually selected for inclusion or discarded by visual inspection. A summary of the data collection and refinement statistics is given in Table 3. All structural comparisons were made with COOT, and figures were prepared with PyMOL (77). Coordinates and structure-factor amplitudes have been deposited with the Protein Data Bank (78) under PDB accession code 5OA5.

Molecular dynamics simulation

To evaluate the role of mutations in enhancing thermal stability, classical MD simulations of wild-type Cel7A and the FCA398 variant were performed at three temperatures: 300, 475, and 525 K. The simulation at 300 K was used to identify baseline dynamics of the two enzymes under standard conditions, providing a point of comparison for behavior at elevated temperatures. High-temperature simulations, at 475 and 525 K, were performed to observe the structural changes that occur as part of the thermal unfolding mechanisms. The selected temperatures are necessarily higher than those the proteins would be exposed to *in vivo*, so that unfolding was observable within a feasible timeframe for molecular simulation. Although the temperatures are quite high, we do not anticipate adverse impact on predicted thermal unfolding mechanisms; prior studies, conducted at simulation temperatures up to 600 K, have established that unfolding mechanisms are relatively independent of simulation temperature (79–81). Comparisons between an averaged set of trajectories, thus defining the unfolding mechanism, were made on the basis of native contacts, where a native contact is any residue within 6.5 Å of a residue of interest (82).

Wild-type Cel7A and the FCA398 variant were constructed from deposited PDB structures, 4C4C (83) and 5OA5, respectively. The E217Q mutation in the 4C4C structure was reversed from Gln to Glu to obtain native Cel7A, and the cellononaose oligomer was removed from the catalytic tunnel. The FCA398 model required no manipulation of the active site. In this study, we modeled only the catalytic domains of the enzymes, maintaining consistency with the crystal structures. Accordingly, the FCA398 model contained only 17 of the 18 mutations (Table 2), as the T462I appears in the linker domain of the full-length protein. The catalytic domains were modeled in a deglycosylated state; based on previous observations, glycosylation effects on protein dynamics are marginal over the MD simulation timescales (84, 85). Protonation states were determined using H++ at pH 5 with an internal and external dielectric constant of 10 and 80, respectively (86–88). Selected protonation states and disulfide bonds are provided in supplemental Table S1. The enzymes were then explicitly solvated in water in an 80 × 80 × 80 Å cubic box, and sodium ions were added to ensure charge neutrality of the systems. The resultant systems contained roughly 52,000 atoms.

After system construction, the models were minimized, heated, and equilibrated before data collection. First, the water molecules and ions were minimized for 1000 steps of steepest descent in CHARMM, keeping the protein fixed (89). This was followed by 1000 steps of steepest descent in which the entire system was allowed to move freely. After minimization, the systems were heated from 100 to 300 K in the *NVE* ensemble over 20 ps in 50 K increments. At this point, three independent 300 K simulations were started from different random number seeds for both wild-type Cel7A and FCA398 (supplemental Fig. S4). These six simulations were then density-equilibrated for 100 ps in the *NPT* ensemble in CHARMM using a Nosè–Hoover thermostat and barostat (90, 91). The CHARMM36

all-atom force field with CMAP corrections was used to model the proteins (89, 92, 93), and the modified TIP3P force field was used to model water (94, 95). Long-range electrostatic interactions were modeled using the particle mesh Ewald approach. The nonbonded cut-off distance was 10 Å, with a switching distance of 9 Å and a nonbonded pair list distance of 12 Å.

The equilibrated systems were simulated for 50 ns in NAMD at 300 K in the *NVT* ensemble using a 2-fs time step (96). High-temperature simulations were started from a 10-ns snapshot of the 300 K simulations and run for 15 ns in the *NVT* ensemble (supplemental Fig. S5). In all, 18 simulations were performed: three replicates at 300, 475, and 525 K for both wild-type Cel7A and FCA398 (supplemental Fig. S4). VMD was used to visualize the trajectories (97).

Author contributions—C. M. conceived and coordinated the study. F. G., J. S., C. M. P., C. M., and M. S. compiled all results and wrote the paper. F. G., L. D., P. K., and P. J. M. T. conducted the *Aspergillus* expression, screening of thermostabilized variants, and larger-scale expression and provided protein for further analysis. V. H. and P. G. purified proteins and performed analyses to generate T_m data. B. R. K. and E. A. L. developed and performed the biochemical analyses. S. K., H. H., and M. S. prepared the FCA398 protein for crystallization, crystallized it, and performed the X-ray crystallography part of the study. S. J. and C. M. P. performed, analyzed, and interpreted the MD simulations and prepared the related figures (Figs. 3 and 4 and supplemental Figs. S3–S5). J. S., C. M. P., and C. M. combined the structural and MD results and developed rationales for stabilizing effect at mutated sites, and J. S. prepared the structural figures (Figs. 2, 5, and 6). All authors reviewed the results and approved the final version of the manuscript.

Acknowledgments—Computing resources were provided by the University of Kentucky and National Science Foundation (NSF) Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by NSF Grant ACI-1053575 (under allocation MCB090159). We thank Dr. Nils Mikkelsen (Swedish University of Agricultural Sciences) for help with X-ray data collection.

References

1. Chundawat, S. P. S., Beckham, G. T., Himmel, M. E., and Dale, B. E. (2011) Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annu. Rev. Chem. Biomol. Eng.* **2,** 121–145
2. Payne, C. M., Knott, B. C., Mayes, H. B., Hansson, H., Himmel, M. E., Sandgren, M., Ståhlberg, J., and Beckham, G. T. (2015) Fungal cellulases. *Chem. Rev.* **115,** 1308–1448
3. Kubicek, C. P. (2013) Systems biological approaches towards understanding cellulase production by *Trichoderma reesei*. *J. Biotechnol.* **163,** 133–142
4. Okada, H., Tada, K., Sekiya, T., Yokoyama, K., Takahashi, A., Tohda, H., Kumagai, H., and Morikawa, Y. (1998) Molecular characterization and heterologous expression of the gene encoding a low-molecular-mass endoglucanase from *Trichoderma reesei* QM9414. *Appl. Environ. Microbiol.* **64,** 555–563
5. Saloheimo, M., Lehtovaara, P., Penttilä, M., Teeri, T. T., Ståhlberg, J., Johansson, G., Pettersson, G., Claeyssens, M., Tomme, P., and Knowles, J. K. (1988) EGIII, a new endoglucanase from *Trichoderma reesei*: the characterization of both gene and enzyme. *Gene* **63,** 11–22
6. Saloheimo, A., Henrissat, B., Hoffrén, A. M., Teleman, O., and Penttilä, M. (1994) A novel, small endoglucanase gene, egl5, from *Trichoderma* reesei isolated by expression in yeast. *Mol. Microbiol.* **13,** 219–228
7. Saloheimo, M., Nakari-Setälä, T., Tenkanen, M., and Penttilä, M. (1997) cDNA cloning of a *Trichoderma reesei* cellulase and demonstration of endoglucanase activity by expression in yeast. *Eur. J. Biochem.* **249,** 584–591
8. Teeri, T. T., Lehtovaara, P., Kauppinen, S., Salovuori, I., and Knowles, J. (1987) Homologous domains in *Trichoderma reesei* cellulolytic enzymes: gene sequence and expression of cellobiohydrolase II. *Gene* **51,** 43–52
9. Ward, M., Wu, S., Dauberman, J., Weiss, G., Larenas, E., Bower, B., Rey, M., Clarkson, K., and Bott, R. (1993) Cloning, sequence and preliminary structural analysis of a small, high pI endoglucanase (EGIII) from *Trichoderma reesei.* In *The Tricel 93 symposium* (Suominen, P., and Reinikainen, T. eds), Vol. 8, pp. 153–158, Foundation for Biotechnical and Industrial Fermentation Research, Espoo, Finland
10. Foreman, P. K., Brown, D., Dankmeyer, L., Dean, R., Diener, S., Dunn-Coleman, N. S., Goedegebuur, F., Houfek, T. D., England, G. J., Kelley, A. S., Meerman, H. J., Mitchell, T., Mitchinson, C., Olivares, H. A., Teunissen, P. J., *et al.* (2003) Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. *J. Biol. Chem.* **278,** 31988–31997
11. Barnett, C. C., Berka, R. M., and Fowler, T. (1991) Cloning and amplification of the gene encoding an extracellular β-glucosidase from *Trichoderma reesei*: evidence for improved rates of saccharification of cellulosic substrates. *Biotechnology* **9,** 562–567
12. Takashima, S., Iikura, H., Nakamura, A., Hidaka, M., Masaki, H., and Uozumi, T. (1999) Comparison of gene structures and enzymatic properties between two endoglucanases from *Humicola grisea*. *J. Biotechnol.* **67,** 85–97
13. Li, X. L., Spániková, S., de Vries, R. P., and Biely, P. (2007) Identification of genes encoding microbial glucuronoyl esterases. *FEBS Lett.* **581,** 4029–4035
14. Jacobson, F., Karkehabadi, S., Hansson, H., Goedegebuur, F., Wallace, L., Mitchinson, C., Piens, K., Stals, I., and Sandgren, M. (2013) The crystal structure of the core domain of a cellulose induced protein (Cip1) from *Hypocrea jecorina*, at 1.5 Å resolution. *PLoS One* **8,** e70562
15. Cherry, J. R., and Fidantsef, A. L. (2003) Directed evolution of industrial enzymes: an update. *Curr. Opin. Biotechnol.* **14,** 438–443
16. Suominen, P. L., Mäntylä, A. L., Karhunen, T., Hakola, S., and Nevalainen, H. (1993) High-frequency one-step gene replacement in *Trichoderma reesei.* 2. Effects of deletions of individual cellulase genes. *Mol. Gen. Genet.* **241,** 523–530
17. Ilmén, M., Saloheimo, A., Onnela, M. L., and Penttilä, M. E. (1997) Regulation of cellulase gene expression in the filamentous fungus *Trichoderma reesei*. *Appl. Environ. Microbiol.* **63,** 1298–1306
18. Chundawat, S. P. S., Lipton, M. S., Purvine, S. O., Uppugundla, N., Gao, D., Balan, V., and Dale, B. E. (2011) Proteomics-based compositional analysis of complex cellulase-hemicellulase mixtures. *J. Proteome Res.* **10,** 4365–4372
19. Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., and Foust, T. D. (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* **315,** 804–807
20. Momeni, M. H., Goedegebuur, F., Hansson, H., Karkehabadi, S., Askarieh, G., Mitchinson, C., Larenas, E. A., Ståhlberg, J., and Sandgren, M. (2014) Expression, crystal structure and cellulase activity of the thermostable cellobiohydrolase Cel7A from the fungus *Humicola grisea* var. *thermoidea*. *Acta Crystallogr. D Biol. Crystallogr.* **70,** 2356–2366
21. Lehmann, M., Loch, C., Middendorf, A., Studer, D., Lassen, S. F., Pasamontes, L., van Loon, A. P., and Wyss, M. (2002) The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng.* **15,** 403–411
22. van den Burg, B., Dijkstra, B. W., van der Vinne, B., Stulp, B. K., Eijsink, V. G., and Venema, G. (1993) Introduction of disulfide bonds into *Bacillus subtilis* neutral protease. *Protein Eng.* **6,** 521–527
23. Sandgren, M., Gualfetti, P. J., Shaw, A., Gross, L. S., Saldajeno, M., Day, A. G., Jones, T. A., and Mitchinson, C. (2003) Comparison of family 12 glycoside hydrolases and recruited substitutions important for thermal stability. *Protein Sci.* **12,** 848–860
24. Goedegebuur, F., Fowler, T., Phillips, J., van der Kley, P., van Solingen, P., Dankmeyer, L., and Power, S. D. (2002) Cloning and relational analysis of 15 novel fungal endoglucanases from family 12 glycosyl hydrolase. *Curr. Genet.* **41,** 89–98

25. Sammond, D. W., Kastelowitz, N., Himmel, M. E., Yin, H., Crowley, M. F., and Bomble, Y. J. (2016) Comparing residue clusters from thermophilic and mesophilic enzymes reveals adaptive mechanisms. *PLoS One* **11,** e0145848
26. Lehmann, M., Pasamontes, L., Lassen, S. F., and Wyss, M. (2000) The consensus concept for thermostability engineering of proteins. *Biochim. Biophys. Acta* **1543,** 408–415
27. Kim, P., Yoon, S. H., Seo, M. J., Oh, D. K., and Choi, J. H. (2001) Improvement of tagatose conversion rate by genetic evolution of thermostable galactose isomerase. *Biotechnol. Appl. Biochem.* **34,** 99–102
28. Sakaue, R., and Kajiyama, N. (2003) Thermostabilization of bacterial fructosyl-amino acid oxidase by directed evolution. *Appl. Environ. Microbiol.* **69,** 139–145
29. Meyer, M. M., Hochrein, L., and Arnold, F. H. (2006) Structure-guided SCHEMA recombination of distantly related β-lactamases. *Protein Eng. Des. Sel.* **19,** 563–570
30. Heinzelman, P., Snow, C. D., Wu, I., Nguyen, C., Villalobos, A., Govindarajan, S., Minshull, J., and Arnold, F. H. (2009) A family of thermostable fungal cellulases created by structure-guided recombination. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 5610–5615
31. Godbole, S., Decker, S. R., Nieves, R. A., Adney, W. S., Vinzant, T. B., Baker, J. O., Thomas, S. R., and Himmel, M. E. (1999) Cloning and expression of *Trichoderma reesei* cellobiohydrolase I in *Pichia pastoris*. *Biotechnol. Prog.* **15,** 828–833
32. Laymon, R. A., Adney, W. S., Mohagheghi, A., Himmel, M. E., and Thomas, S. R. (1996) Cloning and expression of full-length *Trichoderma reesei* cellobiohydrolase I cDNAs in *Escherichia coli*. *Appl. Biochem. Biotechnol.* **57**, 389–397
33. Reinikainen, T., Ruohonen, L., Nevanen, T., Laaksonen, L., Kraulis, P., Jones, T. A., Knowles, J. K., and Teeri, T. T. (1992) Investigation of the function of mutated cellulose-binding domains of *Trichoderma reesei* cellobiohydrolase I. *Proteins* **14,** 475–482
34. Penttilä, M. E., André, L., Lehtovaara, P., Bailey, M., Teeri, T. T., and Knowles, J. K. (1988) Efficient secretion of two fungal cellobiohydrolases by *Saccharomyces cerevisiae*. *Gene* **63,** 103–112
35. Dana, C. M., Dotson-Fagerstrom, A., Roche, C. M., Kal, S. M., Chokhawala, H. A., Blanch, H. W., and Clark, D. S. (2014) The importance of pyroglutamate in cellulase Cel7A. *Biotechnol. Bioeng.* **111,** 842–847
36. Heinzelman, P., Komor, R., Kanaan, A., Romero, P., Yu, X., Mohler, S., Snow, C., and Arnold, F. (2010) Efficient screening of fungal cellobiohydrolase class I enzymes for thermostabilizing sequence blocks by SCHEMA structure-guided recombination. *Protein Eng. Des. Sel.* **23,** 871–880
37. Komor, R. S., Romero, P. A., Xie, C. B., and Arnold, F. H. (2012) Highly thermostable fungal cellobiohydrolase I (Cel7A) engineered using predictive methods. *Protein Eng. Des. Sel.* **25,** 827–833
38. Dana, C. M., Saija, P., Kal, S. M., Bryan, M. B., Blanch, H. W., and Clark, D. S. (2012) Biased clique shuffling reveals stabilizing mutations in cellulase Cel7A. *Biotechnol. Bioeng.* **109,** 2710–2719
39. Voutilainen, S. P., Boer, H., Alapuranen, M., Jänis, J., Vehmaanperä, J., and Koivula, A. (2009) Improving the thermostability and activity of *Melanocarpus albomyces* cellobiohydrolase Cel7B. *Appl. Microbiol. Biotechnol.* **83,** 261–272
40. Voutilainen, S. P., Murray, P. G., Tuohy, M. G., and Koivula, A. (2010) Expression of *Talaromyces emersonii* cellobiohydrolase Cel7A in *Saccharomyces cerevisiae* and rational mutagenesis to improve its thermostability and activity. *Protein Eng. Des. Sel.* **23,** 69–79
41. Smith, M. A., Bedbrook, C. N., Wu, T., and Arnold, F. H. (2013) *Hypocrea jecorina* cellobiohydrolase I stabilizing mutations identified using noncontiguous recombination. *ACS Synth. Biol.* **2,** 690–696
42. Day, A., Goedegebuur, F., Gualfetti, P., Mitchinson, C., Neefe, P., Sandgren, M., Shaw, A., and Stahlberg, J. (January 28, 2014) Variant *Hypocrea jecorina* CBH1 cellulases. United States Patent 8,637,294
43. Kanaya, S. (2000) Stabilization of an enzyme with protein engineering technology. In *Biotechnology for Sustainable Utilization of Biological Resources in the Tropics*, Vol. 14, pp. 186–192, International Center for Biotechnology, Osaka University, Osaka, Japan
44. Gray, K. A., Richardson, T. H., Kretz, K., Short, J. M., Bartnek, F., Knowles, R., Kan, L., Swanson, P. E., and Robertson, D. E. (2001) Rapid evolution of reversible denaturation and elevated melting temperature in a microbial haloalkane. *Adv. Synth. Catal.* **343,** 607–617
45. Stals, I., Sandra, K., Devreese, B., Van Beeumen, J., and Claeyssens, M. (2004) Factors influencing glycosylation of *Trichoderma reesei* cellulases. II: *N*-glycosylation of Cel7A core protein isolated from different strains. *Glycobiology* **14,** 725–737
46. Knott, B. C., Crowley, M. F., Himmel, M. E., Ståhlberg, J., and Beckham, G. T. (2014) Carbohydrate–protein interactions that drive processive polysaccharide translocation in enzymes revealed from a computational study of cellobiohydrolase processivity. *J. Am. Chem. Soc.* **136,** 8810–8819
47. von Ossowski, I., Ståhlberg, J., Koivula, A., Piens, K., Becker, D., Boer, H., Harle, R., Harris, M., Divne, C., Mahdi, S., Zhao, Y., Driguez, H., Claeyssens, M., Sinnott, M. L., and Teeri, T. T. (2003) Engineering the exo-loop of *Trichoderma reesei* cellobiohydrolase, Cel7A: a comparison with *Phanerochaete chrysosporium* Cel7D. *J. Mol. Biol.* **333,** 817–829
48. Momeni, M. H., Payne, C. M., Hansson, H., Mikkelsen, N. E., Svedberg, J., Engström Å., Sandgren, M., Beckham, G. T., and Ståhlberg, J. (2013) Structural, biochemical, and computational characterization of the glycoside hydrolase family 7 cellobiohydrolase of the tree-killing fungus *Heterobasidion irregulare*. *J. Biol. Chem.* **288,** 5861–5872
49. Divne, C., Ståhlberg, J., Teeri, T. T., and Jones, T. A. (1998) High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from *Trichoderma reesei*. *J. Mol. Biol.* **275,** 309–325
50. Muńoz, I. G., Mowbray, S. L., and Stahlberg, J. (2003) The catalytic module of Cel7D from *Phanerochaete chrysosporium* as a chiral selector: structural studies of its complex with the β blocker (*R*)-propranolol. *Acta Crystallogr. D Biol. Crystallogr.* **59,** 637–643
51. Kraulis, J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J., and Gronenborn, A. M. (1989) Determination of the 3-dimensional solution structure of the C-terminal domain of cellobiohydrolase-I from *Trichoderma reesei*: a study using nuclear magnetic-resonance and hybrid distance geometry dynamical simulated annealing. *Biochemistry* **28,** 7241–7257
52. Hobdey, S. E., Knott, B. C., Haddad Momeni, M., Taylor, L. E., 2nd, Borisova, A. S., Podkaminer, K. K., VanderWall, T. A., Himmel, M. E., Decker, S. R., Beckham, G. T., and Ståhlberg, J. (2016) Biochemical and structural characterizations of two dictyostelium cellobiohydrolases from the amoebozoa kingdom reveal a high level of conservation between distant phylogenetic trees of life. *Appl. Environ. Microbiol.* **82,** 3395–3409
53. Serrano, L., and Fersht, A. R. (1989) Capping and α-helix stability. *Nature* **342,** 296–299
54. Bell, J. A., Becktel, W. J., Sauer, U., Baase, W. A., and Matthews, B. W. (1992) Dissection of helix capping in T4 lysozyme by structural and thermodynamic analysis of six amino acid substitutions at Thr 59. *Biochemistry* **31,** 3590–3596
55. Mitchinson, C., and Baldwin, R. L. (1986) The design and production of semisynthetic ribonucleases with increased thermostability by incorporation of S-peptide analogues with enhanced helical stability. *Proteins* **1,** 23–33
56. Benko, Z., Drahos, E., Szengyel, Z., Puranen, T., Vehmaanperä, J., and Réczey, K. (2007) *Thermoascus aurantiacus* CBHI/Cel7A production in *Trichoderma reesei* on alternative carbon sources. *Appl. Biochem. Biotechnol.* **137**, 195–204
57. Matthews, B. W., Nicholson, H., and Becktel, W. J. (1987) Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc. Natl. Acad. Sci. U.S.A.* **84,** 6663–6667
58. Muslin, E. H., Clark, S. E., and Henson, C. A. (2002) The effect of proline insertions on the thermostability of a barley α-glucosidase. *Protein Eng.* **15,** 29–33
59. Textor, L. C., Colussi, F., Silveira, R. L., Serpa, V., de Mello, B. L., Muniz, J. R. C., Squina, F. M., Pereira, N., Jr., Skaf, M. S., and Polikarpov, I. (2013) Joint X-ray crystallographic and molecular dynamics study of cellobiohydrolase I from *Trichoderma harzianum*: deciphering the structural features of cellobiohydrolase catalytic activity. *FEBS J.* **280,** 56–69
60. Turunen, O., Etuaho, K., Fenel, F., Vehmaanperä, J., Wu, X., Rouvinen, J., and Leisola, M. (2001) A combination of weakly stabilizing mutations with a disulfide bridge in the α-helix region of *Trichoderma reesei* endo-

1,4-β-xylanase II increases the thermal stability through synergism. *J. Biotechnol.* **88,** 37–46
61. Berka, R. M., and Barnett, C. C. (1989) The development of gene expression systems for filamentous fungi. *Biotechnol. Adv.* **7,** 127–154
62. Cao, Q. N., Stubbs, M., Ngo, K. Q., Ward, M., Cunningham, A., Pai, E. F., Tu, G. C., and Hofmann, T. (2000) Penicillopepsin-JT2, a recombinant enzyme from *Penicillium janthinellum* and the contribution of a hydrogen bond in subsite S3 to k_{cat}. *Protein Sci.* **9,** 991–1001
63. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
64. Walseth, C. S. (1952) The influence of the fine structure of cellulose on the action of cellulases. *TAPPI* **35,** 233–238
65. Wood, T. M. (1971) The cellulase of *Fusarium solani*: purification and specificity of the β-(1→4)-glucanase and the β-D-glucosidase components. *Biochem. J.* **121,** 353–362
66. Rignall, T. R., Baker, J. O., McCarter, S. L., Adney, W. S., Vinzant, T. B., Decker, S. R., and Himmel, M. E. (2002) Effect of single active-site cleft mutation on product specificity in a thermostable bacterial cellulase. *Appl. Biochem. Biotechnol.* **98**, 383–394
67. Ståhlberg, J., Divne, C., Koivula, A., Piens, K., Claeyssens, M., Teeri, T. T., and Jones, T. A. (1996) Activity studies and crystal structures of catalytically deficient mutants of cellobiohydrolase I from *Trichoderma reesei*. *J. Mol. Biol.* **264,** 337–349
68. McPherson, A. (1982) *Preparation and Analysis of Protein Crystals*, Wiley, New York
69. Matthews, B. W. (1968) Solvent content of protein crystals. *J. Mol. Biol.* **33,** 491–497
70. Kantardjieff, K. A., and Rupp, B. (2003) Matthews coefficient probabilities: improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Sci.* **12,** 1865–1871
71. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 72–82
72. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 125–132
73. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 235–242
74. Brünger, A. T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355,** 472–475
75. Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K., Teeri, T. T., and Jones, T. A. (1994) The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* **265,** 524–528
76. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255
77. Harris, M., and Jones, T. A. (2001) Molray: a web interface between O and the POV-Ray ray tracer. *Acta Crystallogr. D Biol. Crystallogr.* **57,** 1201–1203
78. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112,** 535–542
79. Day, R., Bennion, B. J., Ham, S., and Daggett, V. (2002) Increasing temperature accelerates protein unfolding without changing the pathway of unfolding. *J. Mol. Biol.* **322,** 189–203
80. Huang, X., and Zhou, H. X. (2006) Similarity and difference in the unfolding of thermophilic and mesophilic cold shock proteins studied by molecular dynamics. *Biophys. J.* **91,** 2451–2463
81. Sham, Y. Y., Ma, B., Tsai, C. J., and Nussinov, R. (2002) Thermal unfolding molecular dynamics simulation of *Escherichia coli* dihydrofolate reductase: thermal stability of protein domains and unfolding pathway. *Proteins Struct. Funct. Genet.* **46,** 308–320
82. Sheinerman, F. B., and Brooks, C. L., 3rd (1998) Calculations on folding of segment B1 of streptococcal protein G. *J. Mol. Biol.* **278,** 439–456
83. Knott, B. C., Haddad Momeni, M., Crowley, M. F., Mackenzie, L. F., Götz, A. W., Sandgren, M., Withers, S. G., Ståhlberg, J., and Beckham, G. T. (2014) The mechanism of cellulose hydrolysis by a two-step, retaining cellobiohydrolase elucidated by structural and transition path sampling studies. *J. Am. Chem. Soc.* **136,** 321–329
84. Taylor, C. B., Payne, C. M., Himmel, M. E., Crowley, M. F., McCabe, C., and Beckham, G. T. (2013) Binding site dynamics and aromatic-carbohydrate interactions in processive and non-processive family 7 glycoside hydrolases. *J. Phys. Chem. B* **117,** 4924–4933
85. Payne, C. M., Bomble, Y. J., Taylor, C. B., McCabe, C., Himmel, M. E., Crowley, M. F., and Beckham, G. T. (2011) Multiple functions of aromatic-carbohydrate interactions in a processive cellulase examined with molecular simulation. *J. Biol. Chem.* **286,** 41028–41035
86. Anandakrishnan, R., Aguilar, B., and Onufriev, A. V. (2012) H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **40,** W537–W541
87. Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., and Onufriev, A. (2005) H++: a server for estimating pK_as and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **33,** W368–W371
88. Myers, J., Grothaus, G., Narayanan, S., and Onufriev, A. (2006) A simple clustering algorithm can be accurate enough for use in calculations of p*K*s in macromolecules. *Proteins* **63,** 928–938
89. Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., *et al.* (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30,** 1545–1614
90. Hoover, W. G. (1985) Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **31,** 1695–1697
91. Nose, S., and Klein, M. L. (1983) Constant pressure molecular-dynamics for molecular-systems. *Mol. Phys.* **50,** 1055–1076
92. Mackerell, A. D., Jr., Feig, M., Brooks, C. L., 3rd (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25,** 1400–1415
93. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T., Mattos, C., *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102,** 3586–3616
94. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–935
95. Durell, S. R., Brooks, B. R., and Bennaim, A. (1994) Solvent-induced forces between 2 hydrophilic groups. *J. Phys. Chem.* **98,** 2198–2202
96. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–1802
97. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph. Model.* **14,** 33–38
98. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119
99. Kleywegt, G. J., and Jones, T. A. (1996) xdlMAPMAN and xdlDATAMAN: programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Crystallogr. D Biol. Crystallogr.* **52,** 826–828
100. Kleywegt, G. J., and Jones, T. A. (1997) Detecting folding motifs and similarities in protein structures. *Methods Enzymol.* **277,** 525–545
101. Kleywegt, G. J., and Jones, T. A. (1996) Phi/Psi-cology: Ramachandran revisited. *Structure* **4,** 1395–1400
102. Payne, C. M., Jiang, W., Shirts, M. R., Himmel, M. E., Crowley, M. F., and Beckham, G. T. (2013) Glycoside hydrolase processivity is directly related to oligosaccharide binding free energy. *J. Am. Chem. Soc.* **135,** 18831–18839

Improving the thermal stability of cellobiohydrolase Cel7A from *Hypocrea jecorina* by directed evolution

Frits Goedegebuur, Lydia Dankmeyer, Peter Gualfetti, Saeid Karkehabadi, Henrik Hansson, Suvamay Jana, Vicky Huynh, Bradley R. Kelemen, Paulien Kruithof, Edmund A. Larenas, Pauline J. M. Teunissen, Jerry Ståhlberg, Christina M. Payne, Colin Mitchinson and Mats Sandgren

J. Biol. Chem. 2017, 292:17418-17430.
doi: 10.1074/jbc.M117.803270 originally published online August 31, 2017

Improving the thermal stability of cellobiohydrolase Cel7A from *Hypocrea jecorina* by directed evolution

Frits Goedegebuur[a], Lydia Dankmeyer[a], Peter Gualfetti[b], Saeid Karkehabadi[c], Henrik Hansson[c], Suvamay Jana[d], Vicky Huynh[b], Bradley R. Kelemen[b], Paulien Kruithof[a], Edmund A. Larenas[b], Pauline J.M. Teunissen[b], Jerry Ståhlberg[c], Christina M. Payne[d], Colin Mitchinson[b], and Mats Sandgren[c]

[a] DuPont Industrial Biosciences, Archimedesweg 30, Leiden, 2333CN, The Netherlands;
[b] DuPont Industrial Biosciences, 925 Page Mill Road, Palo Alto, CA 94304, USA;
[c] Department of Chemistry and Biotechnology, Swedish University of Agricultural Sciences, PO box 7015, Uppsala, SE-75007, Sweden;
[d] Department of Chemical and Materials Engineering, University of Kentucky, 177 F. Paul Anderson Tower, Lexington, KY 40506, USA

Running title: *H. jecorina Cel7A thermal stabilization*

To whom correspondence should be addressed: Frits Goedegebuur, frits.goedegebuur@dupont.com; Christina M Payne, christy.payne@uky.edu; Jerry Ståhlberg, jerry.stahlberg@slu.se; Mats Sandgren, mats.sandgren@slu.se

SUPPLEMENTAL DATA

The Supplemental Data contains:

- Residual activity data on methylumbelliferyl-β-lactoside (Figure S1),
- Ramachandran plots for the final structure model of Cel7A FCA398 catalytic model, PDB ID 5OA5 (Figure S2),
- Additional information on the setup of the molecular simulations (Table S1),
- Thermal unfolding trajectories from molecular dynamics simulations (average of three) of wild-type Cel7A and FCA398 at 475 K (Figure S3),
- Difference in number of native contacts per residue between wildtype Cel7A and FCA398 as a function of simulation time at 475 K and 525 K (Figure S4), and
- Visual illustration of the molecular dynamics simulations protocol (Figure S5)

Residual Activity Measurements on Methylumbelliferyl-β-lactoside

Residual activity was assessed by measuring cellobiohydrolase activity against 4-methylumbelliferyl-β-lactoside (MUL; Sigma Chemicals, M2405) at 50 °C, as previously described ((1,2)), before and after heat treatment at elevated temperature. The *A. niger* var. *awamori* expressed and purified Cel7A variants FCA301 (wild type), FCA334 (single-site P227L mutant), FCA353 (6 mutations) and FCA367 (11 mutations) were heat treated for 4 h at 58, 60, 62, 64, 66 and 68 °C in an iEMS microtiter plate incubator, followed by cooling on ice for 10 minutes, before the activity assay on MUL at 50 °C. Residual activity is presented as retained activity (measured at 50°C) after heat treatment, in percentage of the activity without prior heat treatment (Figure S1). The results show that the Cel7A variants are more stable than wild type, in agreement with measured T_m values. The results also demonstrate that even a seemingly small increase in T_m substantially improves the life-time of the enzyme at high temperature.

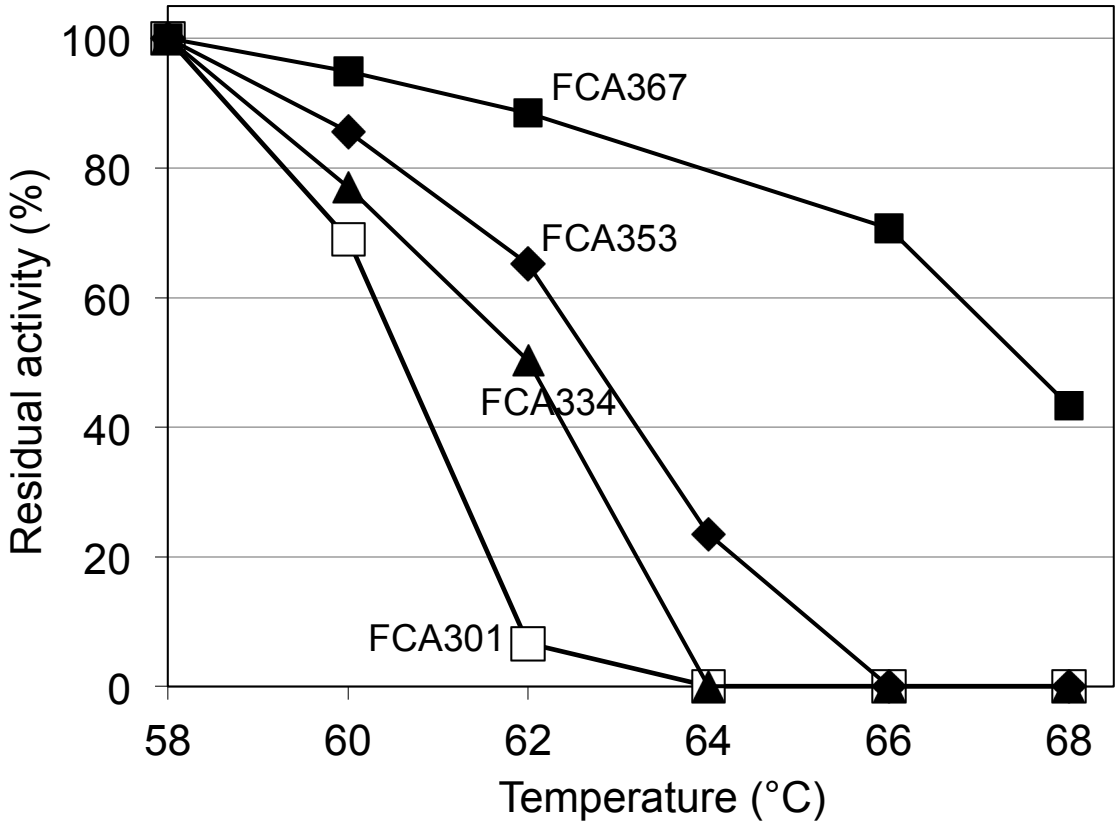

FIGURE S1. Residual activity on MUL of Cel7A wild type (FCA301; open square), single-site P227L mutant FCA344 (triangle) and the combined variants FCA353 (diamond) and FCA367 (filled square), after incubation for 4 h at 58-68 °C. The percentage of residual activity was calculated by dividing the remaining activity by the initial activity and multiplying with 100.

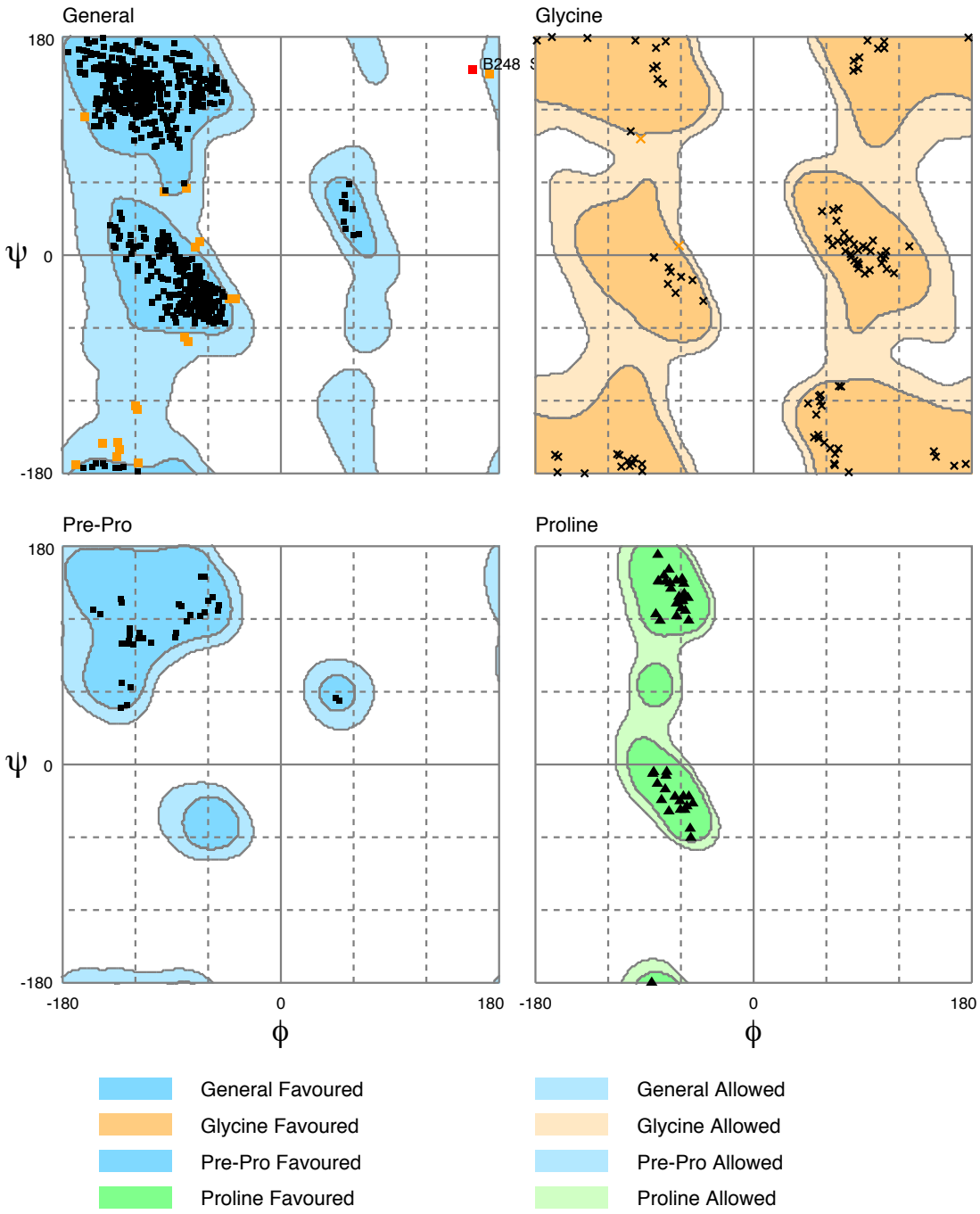

FIGURE S2. Ramachandran plots for the protein mainchain of the final Cel7A FCA398 structure model (PDB ID 5OA5), generated by RAMPAGE (http://www-cryst.bioc.cam.ac.uk/rampage/)(3).

TABLE S1. Molecular dynamics simulation parameters.

	Cel7A wild type	FCA398
Catalytic Domain PDB Structure	4C4C	5OA5
Periodic Boundary Conditions	80 Å × 80 Å × 80 Å	80 Å × 80 Å × 80 Å
Atoms	51,776	51,785
Protonated Residues	Glu65; Glu119; Glu217; Glu223; Glu325; Glu334; Asp214; Asp249; Asp262; Asp368; Asp378	Glu65; Glu119; Glu217; Glu223; Glu325; Glu334; Asp214; Asp368; Asp378
Disulfide Bonds	Cys4 - Cys72; Cys19 - Cys25; Cys50 -71; Cys61 - Cys67; Cys138 - Cys397; Cys172 - Cys210; Cys176 - Cys209; Cys230 - Cys256; Cys238 - Cys243; Cys261 - Cys331	Cys4 - Cys72; Cys19 - Cys25; Cys50 -71; Cys61 - Cys67; Cys138 - Cys397; Cys172 - Cys210; Cys176 - Cys209; Cys230 - Cys256; Cys238 - Cys243; Cys261 - Cys331
Sodium Ions	14	13

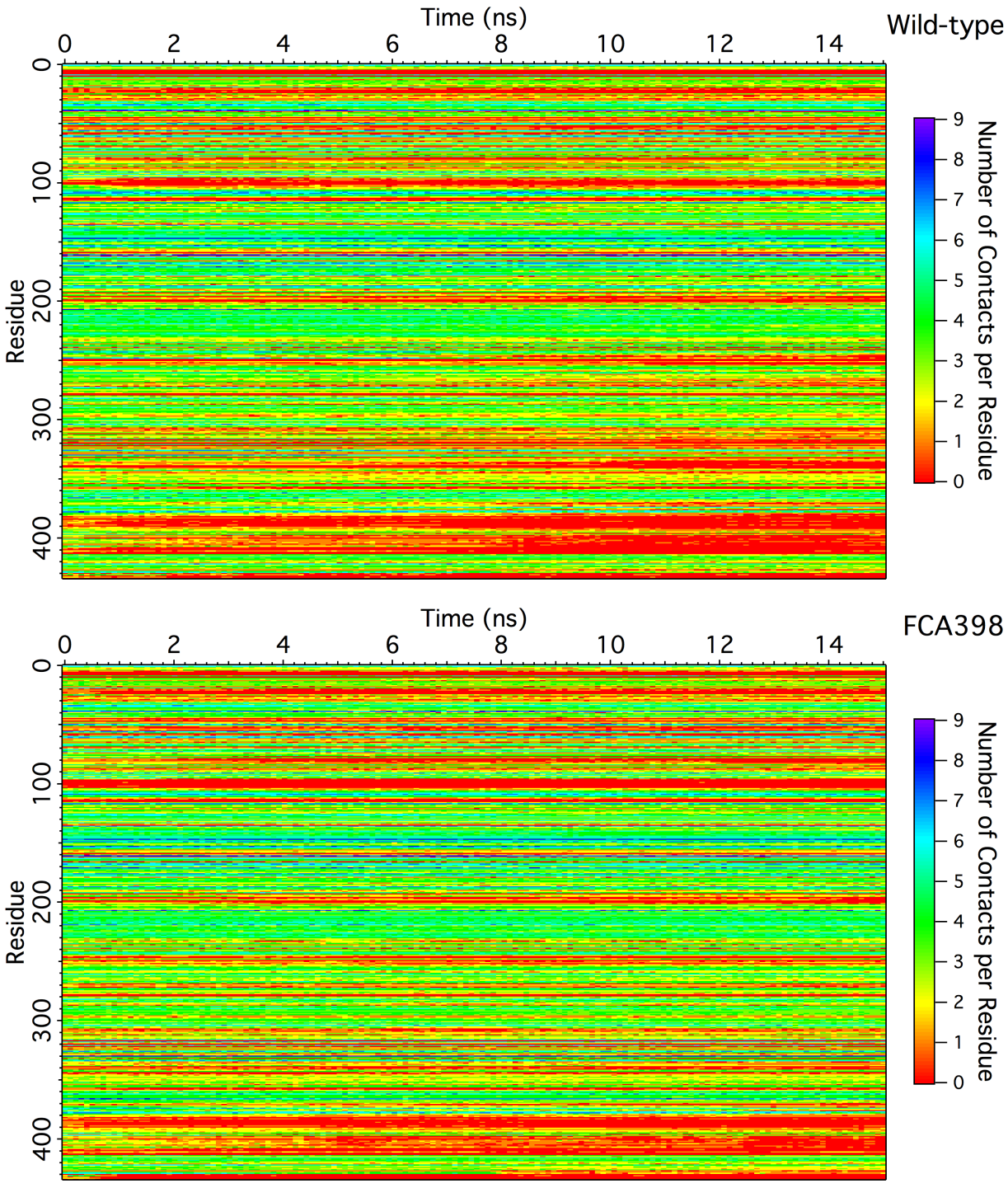

FIGURE S3. Plots of native contacts maintained across the length of the protein as a function of simulation time at 475 K. Wild-type Cel7A is shown in the top panel, and the FCA398 variant is shown in the bottom panel. The number of native contacts for a given residue is shown as a color, where red indicates no contacts exist and indigo indicates the maximum observed number of contacts (9) exist. For example, in the more thermally unstable wild-type enzyme (top), residues 310 to 350 region lose contacts more rapidly over the course of the high-temperature simulation than does FCA398 (bottom), beginning at approximately 5 ns. The number of contacts is the average from three independent simulations.

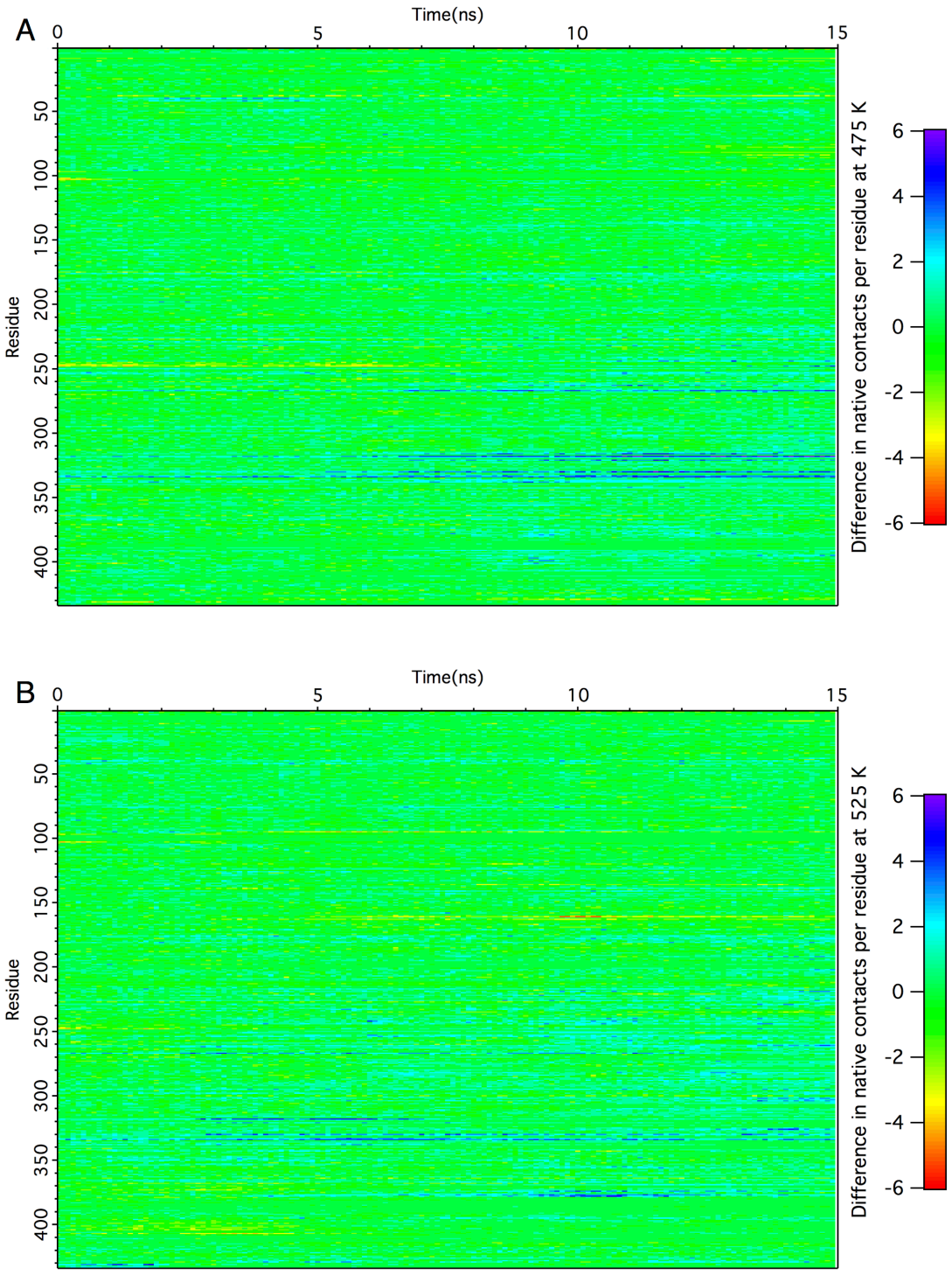

FIGURE S4. Mathematical difference between the number of contacts formed by a given residue in FCA398 and wild type Cel7A at (A) 475 K and (B) 525 K as a function of simulation time. Positive regions (violet/blue/cyan shading) illustrate regions where more protein-protein contact was retained in the FCA398 variant. Negative regions (red/orange/yellow shading) reveal regions where more protein-protein contact was retained by the wild type. Green shading corresponds to regions of the protein in which protein-protein contact and, thus, thermal tolerance was unaffected by mutation.

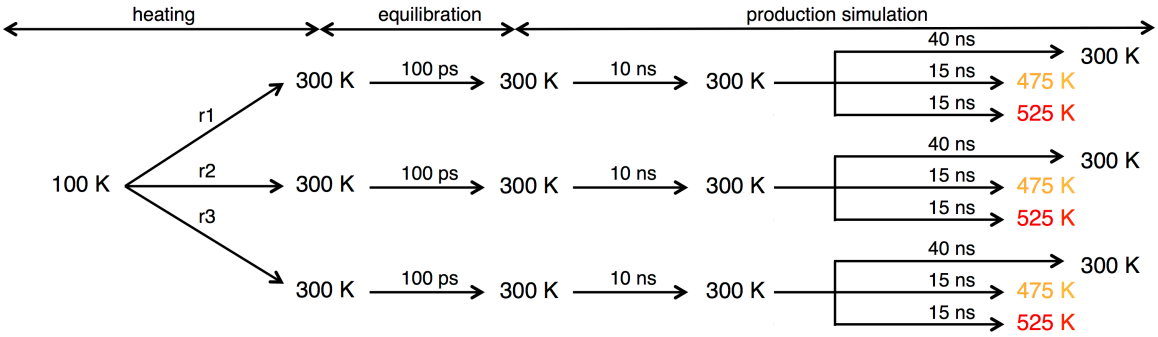

FIGURE S5. Schematic illustration of MD simulations performed in this study. All simulations began with a model constructed from crystal structure, either the wild-type Cel7A or the FCA398 structure, at 100 K. With three different random number seeds (r1, r2, and r3), the structures were heated to 300 K, resulting in three independent simulations. These were then equilibrated for 100 ps in CHARMM. Using NAMD, the three simulations were simulated for an additional 10 ns at 300 K in NAMD. From the final 10 ns snapshot of each of the three simulations, an additional three simulations were generated at 300 K, 475 K, and 525 K. The 300 K simulations were run for an additional 40 ns, totaling 50 ns of simulation data at 300 K. The high-temperature simulations, 475 K and 525 K, were run for 15 ns each. In total, 18 simulations, 9 for each enzyme, were conducted for this study.

REFERENCES

1. Day, A., Goedegebuur, F., Gualfetti, P., Mitchinson, C., Neefe, P., Sandgren, M., Shaw, A., and Stahlberg, J. (2007) Novel variant *Hypocrea jecorina* CBH1 cellulases. Patent US 11/728,219. US 20070173431 A1.
2. Momeni, M. H., Goedegebuur, F., Hansson, H., Karkehabadi, S., Askarieh, G., Mitchinson, C., Larenas, E. A., Stahlberg, J., and Sandgren, M. (2014) Expression, crystal structure and cellulase activity of the thermostable cellobiohydrolase Cel7A from the fungus *Humicola grisea var. thermoidea*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 2356-2366.
3. Lovell, S. C., Davis, I. W., Arendall III, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S. and Richardson, D. C. (2002) Structure validation by Cα geometry: φ/ψ and Cβ deviation. Proteins: Structure, Function & Genetics. **50**, 437-450