SUPPLEMENTAL DATA

Sequence, Structure, and Evolution of Cellulases in the Glycoside Hydrolase Family 48

Leonid O. Sukharnikov[a,1,2], Markus Alahuhta [a,1,3], Roman Brunecky [1,3], Amit Upadhyay[1,2], Michael E. Himmel[1,3], Vladimir V. Lunin[1,3,b] and Igor B. Zhulin[1,2,b]

[1]BioEnergy Science Center and [2]Joint Institute for Computational Sciences, University of Tennessee – Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
[3]Biosciences Center, National Renewable Energy Laboratory, Golden, CO 80401, USA

Supplemental Table S1. Comparison of the properties of various GH48 enzymes

Enzyme accession/ organism	Phylum	Thermostability	Substrate degradation					Reference
			CMC*	CC	BMCC	ASC	Non-cellulose	
55818863/Clostridium thermocellum	Firmicutes	yes	yes	yes	yes	yes	β-glucan	(1)
22219179/Clostridium thermocellum	Firmicutes	yes	yes	yes	NA	yes	no (small range (2: xylan, cotton)	(2)
220928179/ Clostridium cellulolyticum (mesophile)	Firmicutes	NA	yes	yes	NA	yes	no (small range (1: xylan)	(3,4)
5705874/ Clostridium cellulovorans	Firmicutes	NA	yes	yes	NA	yes	Xylan (high act.)	(5)
1708082/Clostridium stercorarium	Firmicutes	yes	no	yes	NA	yes	Xylan(possible contamination) No (barley β-glucan)	(6)
37703325/ Ruminococcus albus 8	Firmicutes	NA	NA	NA	yes	NA	NA	(7)
2437819/ Caldicellulosiruptor bescii	Firmicutes	yes	NA[&]	yes[&]	NA[&]	NA[&]	Xylan (?)[&] - low	(8)
1708078/ Caldicellulosiruptor saccharolyticum	Firmicutes	Yes	NA	NA	NA	NA	NA	(9)
Paenibacillus barcinonensis	Firmicutes	NA	no	yes	yes	yes	no	(10)
1708084/ Cellulomonas fimi	Actinobacteria	NA	yes	NA	yes	yes	no (wide range, β-glucan too)	(11)
72162358/ Thermobifida fusca	Actinobacteria	No	yes[$$]	yes[$$]	yes[$$]	yes[$$]	Binds to chitin	(12)
Myxobacter sp. AL-1	Proteobacteria	No	yes	No	NA	yes	NA	(13)
Hahella chejuensis	Proteobacteria	No	?	Yes $$	?	yes$$	?	This work

CMC-caboxymethyl cellulose; CC (Crystalline cellulose) = Avicel, MN300, sigmacel; BMCC –bacterial microcrystalline cellulose; ASC (Acid swollen cellulose) = PASC, SASC etc.

*activity on CMC is usually defined as 'low';

[$]Range of substrates other than cellulose tell how many were checked approximately (number of substrates other than cellulose checked in parenthesis)

[$$] removal of cbm – serious decrease in activity; all activities for T. fusca were very low

[&]The catalytic properties were determined for the entire protein, which includes GH9 and GH48 domains. Only properties of GH9 domain were described separately.

Supplemental Table S2. X-ray data collection and refinement statistics. Statistics for the highest resolution bin are in parenthesis and numbers of atoms includes alternative conformations.

	Hahella chejuensis GH48
Data Collection	
Space group	P21
Unit cell (Å, °)	a= 61.69, b= 78.39, c = 67.08
	$\alpha = \gamma = 90.0$, $\beta = 98.94$
Wavelength (Å)	1.54178
Temperature (K)	100
Resolution (Å)	25.0-1.75 (1.84-1.75)
Unique reflections	63636 (8808)
Observed reflections	411724 (30035)
R_{int} †	0.0942 (0.4910)
Average redundancy	6.47 3.41)
$\langle I \rangle / \langle \sigma(I) \rangle$	12.71 (1.93)
Completeness (%)	99.0 (100)
Refinement	
R/R_{free}	0.154 (0.205)/0.280 (0.318)
Protein atoms	5193
Water molecules	492
Other atoms	152
RMSD from ideal bond length‡ (Å)	0.021

RMSD from ideal bond angles‡ (°)	2.069
Wilson B-factor	13.9
Average B-factor for protein atoms ($Å^2$)	15.0
Average B-factor for water molecules ($Å^2$)	29.3
Ramachandran plot statistics§ (%)	
Allowed	99.7
Favored	97.3
Outliers	2

† $R_{int} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an individual reflection and $\langle I(hkl) \rangle$ is the mean intensity of a group of equivalents; the sums are calculated over all reflections with more than one equivalent measured. ‡ (14). §(15).

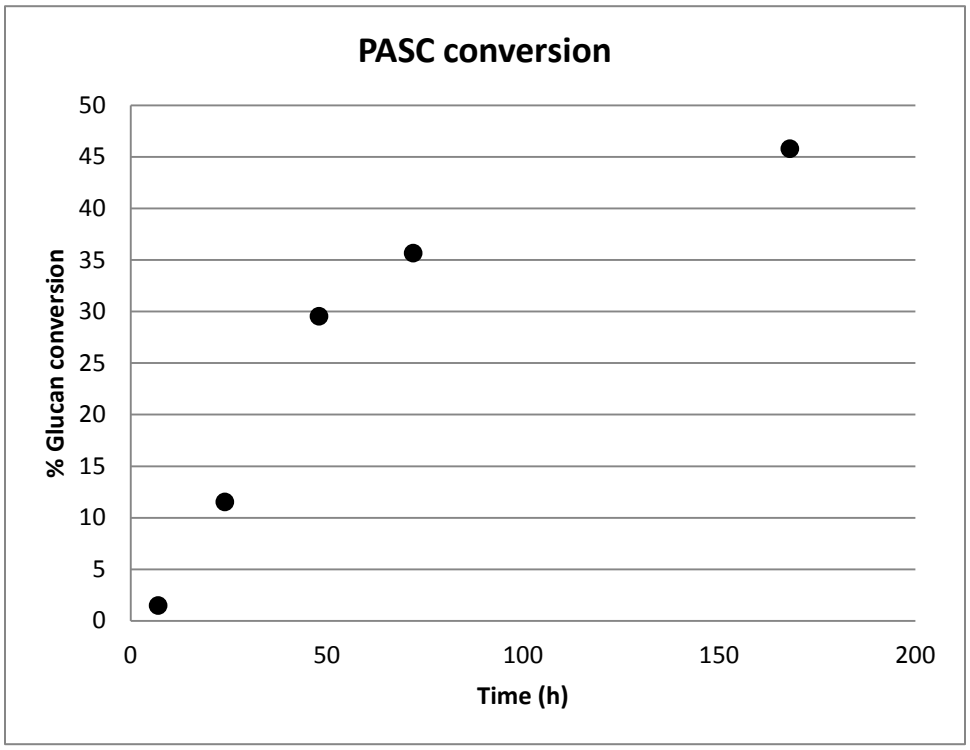
Supplemental Table S3. Conserved in cellulolytic orthologs residues that participate in enzyme-substrate interaction and folding. [*]These residues are located on or adjacent to the loop absent in all insects. [**] Amino acid numbers are as in Cel48F from *C. cellulolyticum* (16)

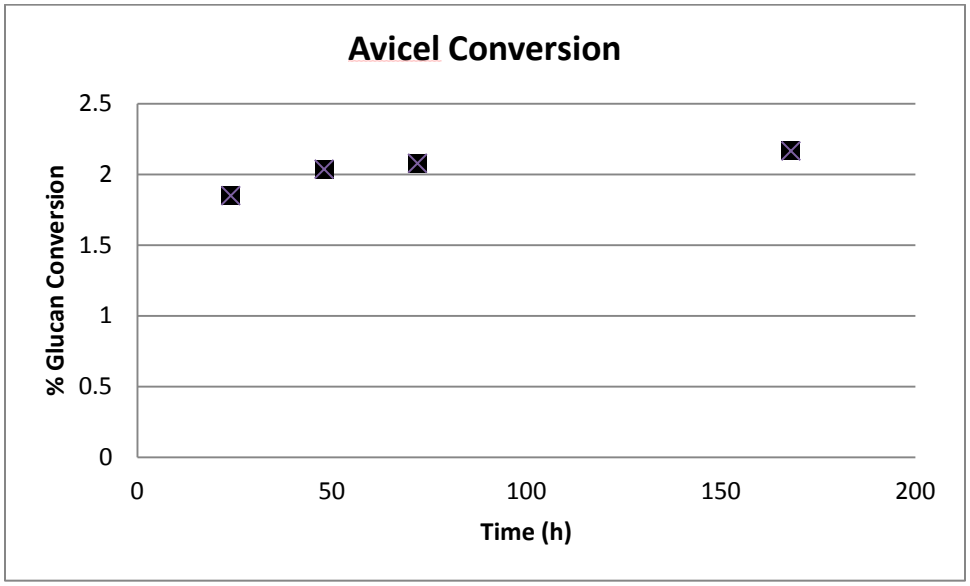
Residue[**]	Conservation	Missense mutations		Function
		Conservative mutations	Non-conservative mutations	
W154	100.00%	-	-	Hydrophobic Stacking Interactions
Y299	100.00%	-	-	
W310	100.00%	-	-	
W312	100.00%	-	-	
W411	100.00%	-	-	
I314	98.00%	-	E-1%	
Y403	92.00%	W-7%	-	
F180	80.00%	Y- 20%	-	
N178	100.00%	-	-	Hydrogen Bonding
Q181	100.00%	-	-	
G183	100.00%	-	-	
E186	100.00%	-	-	
K274	100.00%	-	-	
Y275	100.00%	-	-	
W298	100.00%	-	-	
P406	100.00%	-	-	
D494	100.00%	-	-	
W611	100.00%	-	-	
Q222	98.00%	-	K-1%	
T226	98.00%	S-1%	-	
E542	95.00%			
V402	100.00%	-	-	Calcium coordination
D405	98.00%	N-1%	-	
E190	96.00%	N-1%	K-1%	
R549	96.00%	K-1%	Q-1%	
Q185	83.00%	E-7%; N-1%	S-3%; M,A,R-1%;	
N481	100.00%	-	-	Cellulolytic properties[*]
W472	98.00%	F-1%	-	
L484	96.00%	Y,F-1%	-	

Supplemental Table S4. Percetage of missense mutations of a corresponding amino acid residue in proteobacteria, fungi and insects compared to cellulolytic orthologs. [*]These residues are located on or adjacent to the loop absent in all insects. [**] Amino acid numbers are as in Cel48F from *C. cellulolyticum* (16)

Cellulolytic orthologs	Proteobacteria	Fungi	Insects	Function
W154	0	0	0	Hydrophobic Stacking Interactions
Y299	0	0	21%	
W310	0	0	21%	
W312	0	0	0	
W411	0	0	0	
I314	0	0	21%	
Y403	0	0	5%	
F180	0	0	47%	
N178	0	0	0	Hydrogen Bonding
Q181	0	0	26%	
G183	0	0	16%	
E186	0	0	16%	
K274	0	0	11%	
Y275	0	0	16%	
W298	0	0	0	
P406	0	0	11%	
D494	0	0	32%	
W611	0	0	0	
Q222	0	0	21%	
T226	0	0	47%	
E542	0	0	11%	
V402	0	0	21%	Calcium coordination
D405	0	0	16%	
E190	0	100%	95%	
R549	0	100%	47%	
Q185	0	100%	26%	
N481	0	0	100%	Cellulolytic properties[*]
W472	0	0	100%	
L484	0	0	100%	

Supplemental Figure S1. Cellulolytic activity of the GH48 enzyme from *H. chejuensis*. Conversion of PASC using the GH48 catalytic domain at a 80mg/g enzyme loading. Conversion of Avicel using the GH48 catalytic domain at a 15mg/g enzyme loading.





REFERENCES

1. Berger, E., Zhang, D., Zverlov, V. V., and Schwarz, W. H. (2007) Two noncellulosomal cellulases of Clostridium thermocellum, Cel9I and Cel48Y, hydrolyse crystalline cellulose synergistically. *FEMS Microbiol. Lett.* **268**, 194-201
2. Kruus, K., Wang, W. K., Ching, J., and Wu, J. H. (1995) Exoglucanase activities of the recombinant Clostridium thermocellum CelS, a major cellulosome component. *J. Bacteriol.* **177**, 1641-1644
3. Reverbel-Leroy, C., Belaich, A., Bernadac, A., Gaudin, C., Belaich, J.-P., and Tardif, C. (1996) Molecular study and overexpression of the *Clostridium cellulolyticum* celF cellulase gene in *Escherichia coli*. *Microbiology* **142**, 1013-1023
4. Parsiegla, G., Reverbel-Leroy, C., Tardif, C., Belaich, J. P., Driguez, H., and Haser, R. (2000) Crystal structures of the cellulase Cel48F in complex with inhibitors and substrates give insights into its processive action. *Biochemistry* **39**, 11238-11246
5. Liu, C.-C., and Doi, R. H. (1998) Properties of exgS, a gene for a major subunit of the *Clostridium cellulovorans* cellulosome. *Gene* **211**, 39-47
6. Bronnenmeier, K., Rücknagel, K. P., and Staudenbauer, W. L. (1991) Purification and properties of a novel type of exo-1,4-beta-glucanase (avicelase II) from the cellulolytic thermophile *Clostridium stercorarium. Eur. J. Biochem.* **200**, 379-385
7. Devillard, E., Goodheart, D. B., Karnati, S. K., Bayer, E. A., Lamed, R., Miron, J., Nelson, K. E., and Morrison, M. (2004) *Ruminococcus albus* 8 mutants defective in cellulose degradation are deficient in two processive endocellulases, Cel48A and Cel9B, both of which possess a novel modular architecture. *J. Bacteriol.* **186**, 136-145
8. Zverlov, V., Mahr, S., Riedel, K., and Bronnenmeier, K. (1998) Properties and gene structure of a bifunctional cellulolytic enzyme (CelA) from the extreme thermophile 'Anaerocellum thermophilum' with separate glycosyl hydrolase family 9 and 48 catalytic domains. *Microbiology* **180**, 3091-3099
9. Te'o, V. S., Saul, D. J., and Bergquist, P. L. (1995) CelA, another gene coding for a multidomain cellulase from the extreme thermophile *Caldocellum saccharolyticum*. *Appl. Microbiol. Biotechnol.* **43**, 291-296
10. Sanchez, M. M., Pastor, F. I., and Diaz, P. (2003) Exo-mode of action of cellobiohydrolase Cel48C from *Paenibacillus sp.* BP-23. *Eur. J. Biochem.* **270**, 2913-2919
11. Shen, H., Gilkes, N. R., Kilburn, D. G., Miller, R. C., Jr, and Warren, R. A. (1995) Cellobiohydrolase B, a second exo-cellobiohydrolase from the cellulolytic bacterium *Cellulomonas fimi*. *Biochem. J.* **311**, 67-74
12. Irwin, D. C., Zhang, S., and Wilson, D. B. (2000) Cloning, expression and characterization of a family 48 exocellulase, Cel48A, from *Thermobifida fusca*. *Eur. J. Biochem.* **267**, 4988-4997
13. Ramírez-Ramírez, N., Romero-García, E. R., Calderón, V. C., Avitia, C. I., Téllez-Valencia, A., and Pedraza-Reyes, M. (2008) Expression, characterization and synergistic interactions of *Myxobacter sp*. AL-1 Cel9 and Cel48 glycosyl hydrolases. *Int. J. Mol. Sci.* **9**, 247-257
14. Engh, R. A., and Huber, R. (1991) Accurate Bond and Angle Parameters for X-Ray Protein-Structure Refinement. *Acta Crystallogr. A* **47**, 392-400
15. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12-21
16. Parsiegla, G., Juy, M., Reverbel-Leroy, C., Tardif, C., Belaïch, J. P., Driguez, H., and Haser, R. (1998) The crystal structure of the processive endocellulase CelF of *Clostridium cellulolyticum* in complex with a thiooligosaccharide inhibitor at 2.0 A resolution. *EMBO J.* **17**, 5551-5562