Supplementary Information

Cellulose degradation and assimilation by the unicellular phototrophic eukaryote

Chlamydomonas reinhardtii

Olga Blifernez-Klassen, Viktor Klassen, Anja Doebbe, Klaudia Kersting, Philipp Grimm, Lutz Wobbe & Olaf Kruse*

Bielefeld University, Algae Biotechnology & Bioenergy group/ Dep. of Biology - Center for Biotechnology (CeBiTec); Bielefeld/ Germany

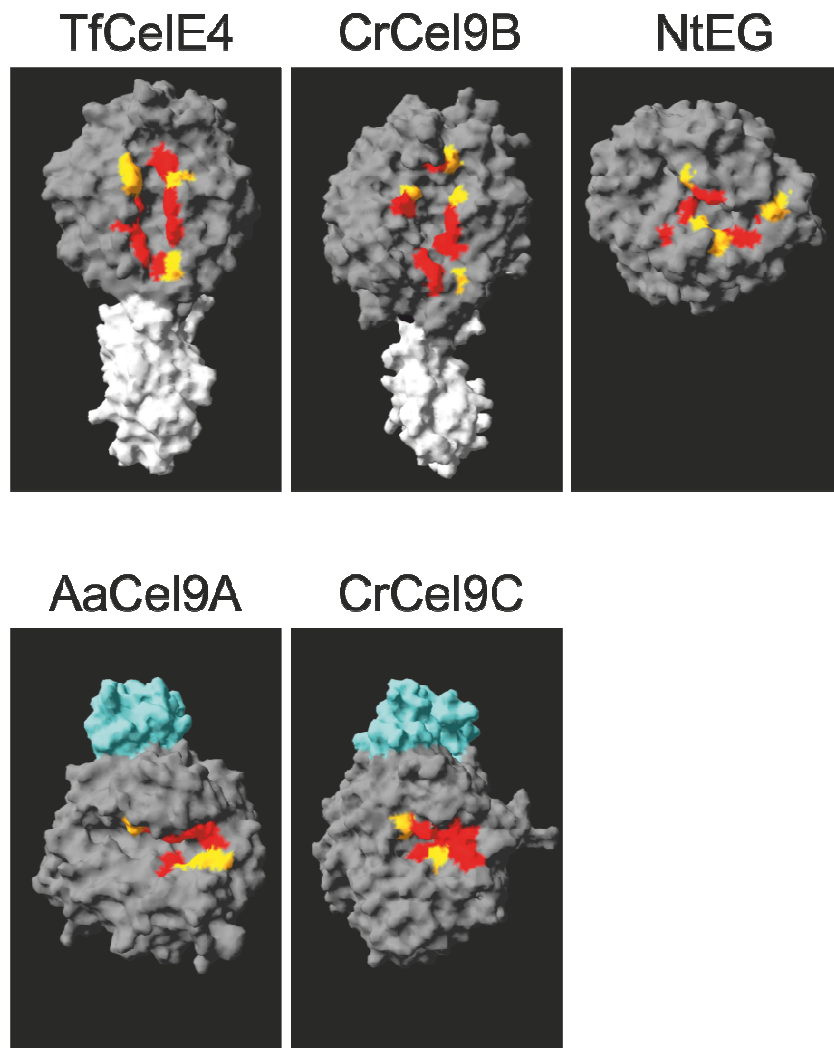*to whom correspondence should be addressed: olaf.kruse@uni-bielefeld.de

This PDF file includes:

> Supplementary Figure S1
>
> Supplementary Tables S1/S2/S3/S4
>
> Supplementary references (60-71)

Supplementary Figure S1

Supplementary Figure S1 - Surface plots and homology models

Surface plots of TfCelE4 from *Thermomonospora fusca* [Protein Data Bank code (PDB): 1JS4][26], NtEG from the termite *Nasutitermes takasagoensis* [PDB 1KS8][60] and AaCel9A from *Alicyclobacillus acidocaldarius* [PDB 3GZK][27]. Homology models of the *Chlamydomonas reinhartii* cellulases CrCel9B (UniProtKB A8JFG8) and CrCel9C (UniProtKB A8JFH1) created with I-TASSER[25]. The catalytic domains are colored in grey and the cellulose binding domain of TfCelE4 is shown in white. A turquoise color indicates the position of the immunglobulin-like domain in AaCel9A. Red and yellow colors within the catalytic domain indicate the conserved aromatic residues (red means fully conserved, yellow non-conserved among the species), which are involved in substrate stacking[26-27,60-61].

Supplementary Table S1.

| Protein | UniProt | Phytozome | GF | Secretion | | | *C-score* | Structural homolog |
				TargetP1.1 **(RC/score)**	**SignalP 4.0** **(DS)**	**SecretomeP 2.0** **(SecPS)**	EC No. (EC score) *function*	**(PDB)** *TM score*
CrBGl1	A8I718		1	other location (RC2/0.868)	non-secreted (0.135)	non-secreted (0.423)	*-2.29* 3.2.1.21 (0.9082) *β-glucosidase*	Taglu1b (2DGA) *0.6569*
		Cre07.g320850	1	other location (RC2/0.868	non-secreted (0.135)	non-secreted (0.367)		
CrBGl2	A8IEM7		1	other location (RC2/0.817)	non-secreted (0.110)	secreted (0.657)	*-2.60* 3.2.1.21 (0.6136) *β-glucosidase*	*Taglu1b* (2DGA) *0.6136*
		Cre03.g171050		other location (RC2/0.817)	non-secreted (0.110)	non-secreted (0.319)		
CrCel9B	A8JFG8		9	secreted (RC2/0.88)	secreted (0.741)	secreted (0.757)	*-0.28* 3.2.1.4 (1.5814) *endo-1,4-β-D-glucanase*	*endo/exo-cellulase E4* (1JS4) *0.8213*
		Cre17.g730600		mitochondria (RC4/0.614)	non-secreted (0.145)	non-secreted (0.364)		
CrCel9C	A8JFH1		9	chloroplast (RC5/0.228)	non-secreted (0.321)	secreted (0.506)	*-0.09* 3.2.1.4 (0.8799) *endo-1,4-β-D-glucanase*	*AaCel9A* (3GZK) *0.7840*
		Cre17.g730550		secreted (RC1/0.915)	secreted (0.699)	non-secreted (0.402)		
CrCel9D	A8HW25		9	other location (RC2/0.911)	non-secreted (0.111)	secreted (0.629)	*-2.55* 3.2.1.4 (0.9158) *endo-1,4-β-D-glucanase*	*Cel9G* (1GA2) *0.6731*
		Cre06.g270500	9	other location (RC2/0.895)	non-secreted (0.111)	secreted (0.521)		

Supplementary Table S1 - Identification of cellulase genes in the nuclear genome of *C. reinhardtii*

A BLAST survey of the *C. reinhardtii* nuclear genome revealed the existence of endoglucanases and β-glucosidases. UniProtKB[62] data base accession numbers as well as the Phytozome 7.0 locus names[63] are given for available protein/gene annotations. Glycoside hydrolases family (GF) classification was performed using the Carbohydrate-Active Enzymes (CAZy) server[64-66]. Furthermore, we used the prediction server SignalP 4.0[67] in combination[68] with TargetP 1.1[69] to determine the likelihood of secretion for both available protein models (UniProt and Phytozome). The TargetP 1.1 localization prediction (settings: plant; no cutoffs)[69] is given with the reliability class in brackets. For the SignalP 4.0 results D-score values[67] (DS) are indicated in brackets as a measure for the prediction quality. Secretome P2.0[70] predictions (setting: mammalian) together with SecP-scores (SecPS) were included to consider non-classical (leaderless) secretion. The potential function of the proteins was inferred from *in silico* models created with I-TASSER[25] and confidence scores (C-score) as well as Enzyme Commission (EC) numbers with EC scores as a confidence measure are given. For each glycoside hydrolase the PDB entry showing the closest structural similarity is indicated together with respective TM scores as a measure for the similarity. A correct fold can be expected for models having C-score values >-1.5 and determination of structure class / protein family requires TM scores > 0.5[25].

Supplementary Table S2.

Protein	UniProtKB	Phytozome	GHF	possible function (NCBI) / EC-number
CrGHF2A	A8IZ85	Cre08.g379450	2	β-galactosidase (EC 3.2.1.23)
CrGHF3A	A8JBP6	Cre19.g752200	3	xylan 1,4-β-xylosidase (EC 3.2.1.37) β-glucosidase (EC 3.2.1.21) α-L-arabinofuranosidase (EC 3.2.1.55)
CrGHF5A	A8HN72	Cre01.g048350	5	β-mannosidase (EC 3.2.1.25) mannan endo-β-1,4-mannosidase (EC 3.2.1.78) β-1,3-mannanase (EC 3.2.1.-)
CrGHF5B	A8I258	Cre02.g098000	5	
CrGHF5C	A8ITT2	Cre07.g343950	5	
CrGHF5D	A8JDR6	Cre12.g492800	5	
CrGHF5E	A8JCF5	Cre07.g337750	5	
CrGHF16A	A8IEP7	Cre03.g170700	16	endo-1,3-β-glucanase (laminarinase, EC 3.2.1.39) endo-1,3(4)-β-glucanase (EC 3.2.1.6)
CrGHF16B	A8IED3	Cre03.g173100	16	
CrGHF16C	A8I2Q2	Cre02.g115950	16	
CrGHF27A (CrAGA1)	A8HQ67	Cre01.g026250	27	α-galactosidase (EC 3.2.1.22)
CrGHF28A	A8IYS7	Cre12.g556350	28	polygalacturonase (pectinase, EC 3.2.1.15) exo-polygalacturonase (EC 3.2.1.67)
CrGHF31A (CrAGL1)	A8IWM1	Cre03.g194700	31	α-glucosidase (EC 3.2.1.20) α-1,3-glucosidase (EC 3.2.1.84) α-xylosidase (EC 3.2.1.177)
CrGHF31B	A8IRW5	Cre03.g190500	31	
CrGHF32A	A8ILB2	Cre12.g488000	32	ß-fructofuranosidase (invertase, EC 3.2.1.26) endo-inulinase (EC 3.2.1.7) sucrose:sucrose 1-fructosyltransferase (EC 2.4.1.99) fructan:fructan 6G-fructosyltransferase (EC 2.4.1.243) fructan β-(2,1)-fructosidase/1-exohydrolase (EC 3.2.1.153) fructan β-(2,6)-fructosidase/6-exohydrolase (EC 3.2.1.154)
CrGHF32B	A8ILB4	Cre12.g488050	32	
CrGHF32C	-	Cre12.g501900	32	
CrGHF32D	-	Cre12.g501850	32	
CrGHF32E	-	Cre12.g507029	32	
CrGHF32F	-	Cre13.g582300	32	
CrGHF38A	A8IG58	Cre10.g437950	38	α-mannosidase (EC 3.2.1.24)
CrGHF42A	A8J0P1	Cre02.g141600	42	β-galactosidase (EC 3.2.1.23) mannan endo-β-1,4-mannosidase (EC 3.2.1.78) β-1,3-mannanase (EC 3.2.1.-)
CrGHF47A (CrAMD1)	A8JCH8	Cre07.g336600	47	α-mannosidase (EC 3.2.1.113)
CrGHF47B	-	Cre06.g301600	47	
CrGHF63A	A8HTR3	Cre13.g579750	63	processing α-glucosidase (EC 3.2.1.106) α-1,3-glucosidase (EC 3.2.1.84) α-glucosidase (EC 3.2.1.20)
CrGHF81A	A8JHA2	Cre12.g513400	81	endo-β-1,3-glucanase (EC 3.2.1.39)
Cellulase-like (PF12876) proteins				
CrCelL1	A8I673	Cre07.g314850	PF12876	mannan endo-β-1,4-mannosidase (EC 3.2.1.78) β-1,3-mannanase (EC 3.2.1.-)

Supplementary Table S2 - Identification of hemicellulase genes in the genome of *C. reinhardtii*

C. reinhardtii glycoside hydrolase family (GHF) members[62-66] potentially implicated in the degradation of complex polysaccharides other than cellulose. UniProtKB[62] data base accession numbers are given in addition to the Phytozome 7.0 locus names[63] of identified genes. Functional predictions were inferred from NCBI BLAST[71] results.

Supplementary Table S3.

Observed	Mr (expt)	Mr (calc)	ppm	Miss	Score	Expect	Rank	Peptide
971.5182	970.5109	970.5097	1.30	0	33	0.0079	1	K.APEKPHHR.S
1086.5726	1085.5653	1085.5618	3.22	0	40	0.0016	1	R.AAGQWDIAVR.N
1349.7686	1348.7613	1348.7575	2.80	0	49	0.0001	1	R.HTANAALVALAAAR.G
1366.6650	1365.6577	1365.6525	3.87	0	45	0.00031	1	R.GDAGAGLEVAYSTR.V
1774.7815	1773.7742	1773.7781	-2.18	0	15	0.21	1	R.FYEAQMSGNVPSWSR.A + Oxidation (M)

Supplementary Table S3 - Mass spectrometric identification of CMCases

MASCOT (version 2.2.04) MS/MS Ion search results of mass spectrometric analyses performed with the band extracted from the zymogram shown in Fig. 1C. Both proteins CrCel9B (UniProtKB A8JFG8) and CrCel9C (UniProtKB A8JFH1) were identified with identical scores (MASCOT search score 98 / significance threshold 23). Observed, expected and calculated nominal masses (Mr) of peptides are given along with deviations (ppm). The number of missed cleavages (Miss), individual ion scores (Score), expect values (Expect), rank positions (Rank) and the amino acid sequence (Peptide) for each identified peptide are given. Identified peptides are all derived from the catalytic domains (amino acid 42-544 in A8JFG8 and 105-587 in A8JFH1) of the cellulases. The identity between both proteins in this region is 83.3% and the overall identity of the proteins is 64%.

Supplementary Table S4.

Name	Sequence 5´ → 3´
A8JFG8 fwd	ACTGACTGGGGCTGTAACCA
A8JFG8 rev	GCGGTCCTTGTAAGCTTGTT
A8JFH1 fwd	CGTCTATGGGGACGTGAACT
A8JFH1 rev	AAGTCTGGGTGGAGGGAATC
A8HW25 fwd	GCATGGACTGGAACCTCAAT
A8HW25 rev	TTGGCTGTGGTGTACTTGGA
A8I718 fwd	GTGTGGCAGAACAGCTTTGA
A8I718 rev	CAGCACTGGATGTCTTCGTG
A8IEM7 fwd	ACTGGTGGGGCATCAACTAC
A8IEM7 rev	GTACATGGGAATGCCGAACT
actin fwd	ACGACACCACCTTCAACTCC
actin rev	ACATTTGCTGGAAGGTGGAC

Supplementary Table S4 - Oligonucleotides used for RT-Q-PCR analyses

Supplementary references

60.   Khademi, S., Guarino, L.A., Watanabe, H., Tokuda, G. & Meyer, E.F. Structure of an endoglucanase from termite, *Nasutitermes takasagoensis*. *Acta Crystallogr.* **58D**, 653-659 (2002).

61.   Mandelman, D. *et al.* X-Ray crystal structure of the multidomain endoglucanase Cel9G from *Clostridium cellulolyticum* complexed with natural and synthetic cello-oligosaccharides. *J. Bacteriol.* **185**, 4127-4135 (2003).

62.   http://www.uniprot.org/

63.   http://www.phytozome.net

64.   http://www.cazy.org/

65.   Cantarel, B.L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37**, D233-D238 (2009).

66.   Henrissat, B. A classification of glycosyl hydrolases based on amino-acid-sequence similarities. *Biochem. J.* **280**, 309-316 (1991).

67.   Petersen, T.N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Meth.* **8**, 785-786 (2011).

68.   Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2**, 953-971 (2007).

69.   Emanuelsson, O., Nielsen, H., Brunak, S.R. & von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005-1016 (2000).

70.   Bendtsen, J. D., L. J. Jensen, N. Blom, G. Von Heijne & S. Brunak. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng. Des. Sel.* **17,** 349-356 (2004).

71.   http://blast.ncbi.nlm.nih.gov/Blast.cgi