

Structure and Enzymatic Characterization of CelD Endoglucanase from the Anaerobic Fungus *Piromyces finnis*

Alexey Dementiev^{2*}, Stephen P. Lillington^{1*}, Shiyan Jin¹, Youngchang Kim², Robert Jedrzejczak², Karolina Michalska², Andrzej Joachimiak^{2,3}, Michelle A. O'Malley^{1,4,5}

¹ Department of Chemical Engineering, University of California, Santa Barbara, CA USA

² Structural Biology Center, X-ray Science Division, Argonne National Laboratory, Argonne, IL 60439, USA,

³ Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637, USA,

⁴ Biological Engineering Program, University of California, Santa Barbara, CA USA

⁵ Joint BioEnergy Institute (JBEI), Emeryville, CA, 94608, USA

*Equal contributions

Corresponding author: Michelle A. O'Malley, momalley@ucsb.edu

Contents:

Table S1 – Amino acid sequences for CelD constructs characterized in this work

Figure S1 – Multiple sequence alignment of homologous GH5 subfamily 4 enzymes to *Piromyces finnis* CelD

Figure S2 – Structural alignment of homologous GH5 subfamily 4 enzymes to *Piromyces finnis* CelD

Figure S3 - Sequence alignment of fungal cellulases from *Piromyces finnis* (CelD), *Thermoascus aurantiacus* (EngI), *Trichoderma reesei* (EgII), and *Piromyces rhizinflata* (EglA)

Figure S4 - Structural comparison of the fungal cellulase catalytic domains.

Figure S5 – Timecourse measurements of reducing sugar concentration for the CelD catalytic domain against several substrates.

Table S1. Amino acid sequences for CelD constructs characterized in this work.

Construct	Amino acid sequence	Predicted Molecular Weight (kDa)
91-452 (catalytic domain only)	AIRDISSLELIKEMRFGWNLGNTLDAECTSWMDYEKNPIGSETCWG NVKTNEDIFKTLMDNQFNVFRIPTTWTGHIGEAPYKINEQWMKRV HEIVDYPYKNGAFVILNIHHETWNHAFETVDEAKVELAQVWKQIA EEFKGYGERLIFEGQNEPRKNGTPVEWNGGDKEGWDVVNAMNAVFL ETVRSSGGNNAKRHLMIPPYAAACNENSFKNFDFPEDDDKVIASVH AYSPYNFALNNGEGAVDKFDASGKNELDWNINLMKKRFVDQGIPMI LGEYGAMNRDNEEERAWEAYMEKITALGVPQVWWDNGVFEGEGE RFGIDRKNLKIVYPSIVAALQKGRGLEVNVLHAITE	41.2
91-452 E154A	AIRDISSLELIKEMRFGWNLGNTLDAECTSWMDYEKNPIGSETCWG NVKTNEDIFKTLMDNQFNVFRIPTTWTGHIGEAPYKINEQWMKRV HEIVDYPYKNGAFVILNIHHETWNHAFETVDEAKVELAQVWKQIA EEFKGYGERLIFEGQNA PRKNGTPVEWNGGDKEGWDVVNAMNAVFL ETVRSSGGNNAKRHLMIPPYAAACNENSFKNFDFPEDDDKVIASVH AYSPYNFALNNGEGAVDKFDASGKNELDWNINLMKKRFVDQGIPMI LGEYGAMNRDNEEERAWEAYMEKITALGVPQVWWDNGVFEGEGE RFGIDRKNLKIVYPSIVAALQKGRGLEVNVLHAITE	41.1
91-536 (catalytic domain + C-terminal double dockerin. Dockerins highlighted in bold)	AIRDISSLELIKEMRFGWNLGNTLDAECTSWMDYEKNPIGSETCWG NVKTNEDIFKTLMDNQFNVFRIPTTWTGHIGEAPYKINEQWMKRV HEIVDYPYKNGAFVILNIHHETWNHAFETVDEAKVELAQVWKQIA EEFKGYGERLIFEGQNEPRKNGTPVEWNGGDKEGWDVVNAMNAVFL ETVRSSGGNNAKRHLMIPPYAAACNENSFKNFDFPEDDDKVIASVH AYSPYNFALNNGEGAVDKFDASGKNELDWNINLMKKRFVDQGIPMI LGEYGAMNRDNEEERAWEAYMEKITALGVPQVWWDNGVFEGEGE RFGIDRKNLKIVYPSIVAALQKGRGLEVNVLHAITE TEPCWSLK YGYECCSPNNTRVVVTDESGKWGVENS DWCGIVDSKDKCWSIPFGY KCCDHCKVLLTDESGKWGELNGEWCGIDTTCK	51.0
1-536 (N-terminal dockerin + catalytic domain + C-terminal double dockerin. Dockerins highlighted in bold.)	MRILNVLSLTGLVIAGTNALDCWANKLGYDCC TSKEVYEDVEGKW GVENGWCGIVEEPTTTVELVEEPTPAVDVEEPTPEVELEPIR DISSLELIKEMRFGWNLGNTLDAECTSWMDYEKNPIGSETCWGNV KTNEDIFKTLMDNQFNVFRIPTTWTGHIGEAPYKINEQWMKRVHEI VDYPYKNGAFVILNIHHETWNHAFETVDEAKVELAQVWKQIAEEF KGYGERLIFEGQNEPRKNGTPVEWNGGDKEGWDVVNAMNAVFL ETVRSSGGNNAKRHLMIPPYAAACNENSFKNFDFPEDDDKVIASVH AYSPYNFALNNGEGAVDKFDASGKNELDWNINLMKKRFVDQGIPMI LGEYGAMNRDNEEERAWEAYMEKITALGVPQVWWDNGVFEGEGE RFGIDRKNLKIVYPSIVAALQKGRGLEVNVLHAITE TEPCWSLK YGYECCSPNNTRVVVTDESGKWGVENS DWCGIVDSKDKCWSIPFGY KCCDHCKVLLTDESGKWGELNGEWCGIDTTCK	60.8

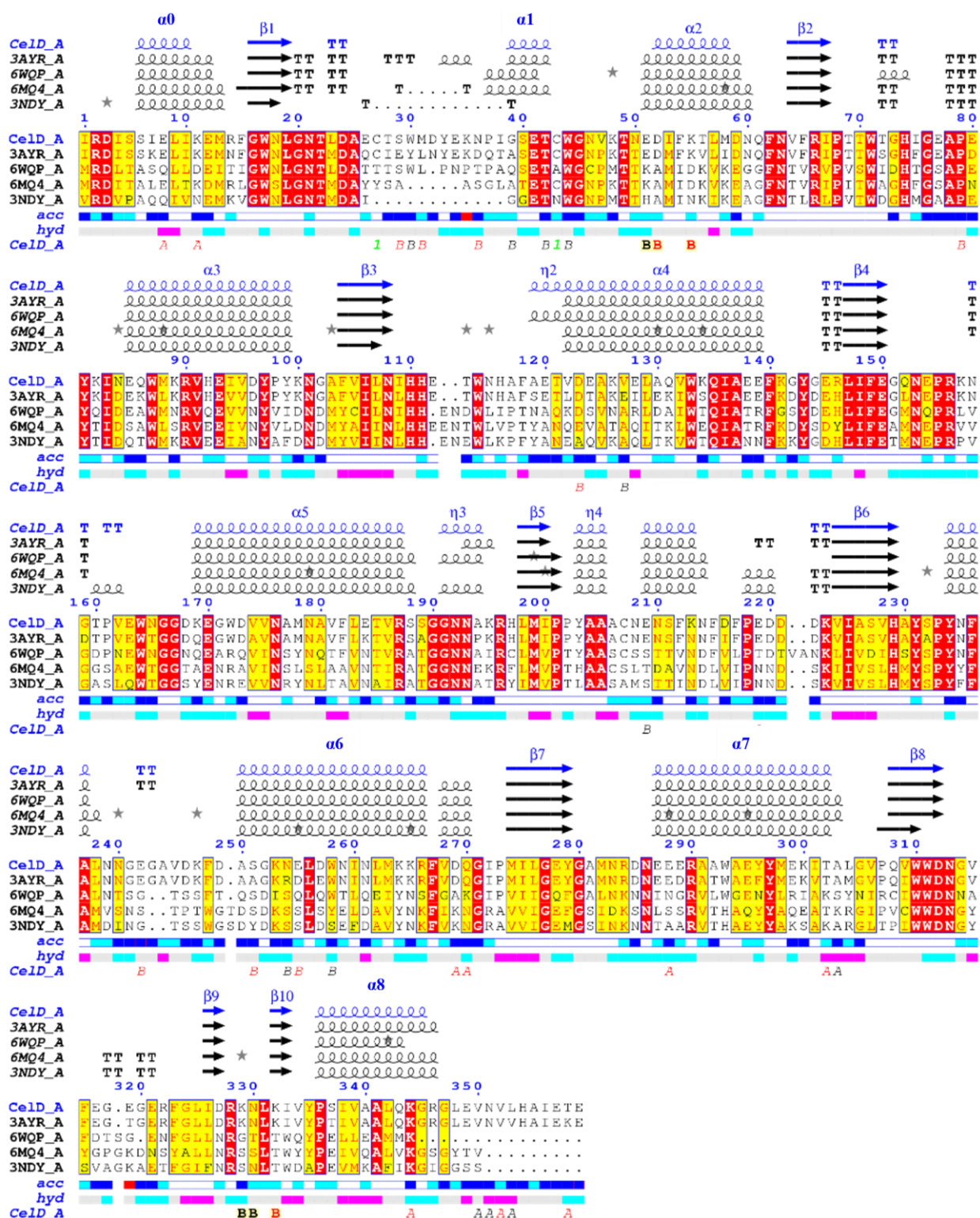


Figure S1. Sequence alignment of homologous GH5_4 enzymes with known structures. Strictly conserved residues are shown in red block and chemically similar residues shown in yellow block. “acc” represents the solvent accessibility and “hyd” the hydropathy of the sequence. 6WQP contains two mutations from the wild-type enzyme sequence (Uniprot Accession no. D4LAX7), E154Q and E278Q in the multiple sequence alignment. Figure generated with Endscript (Robert and Gouet 2014).

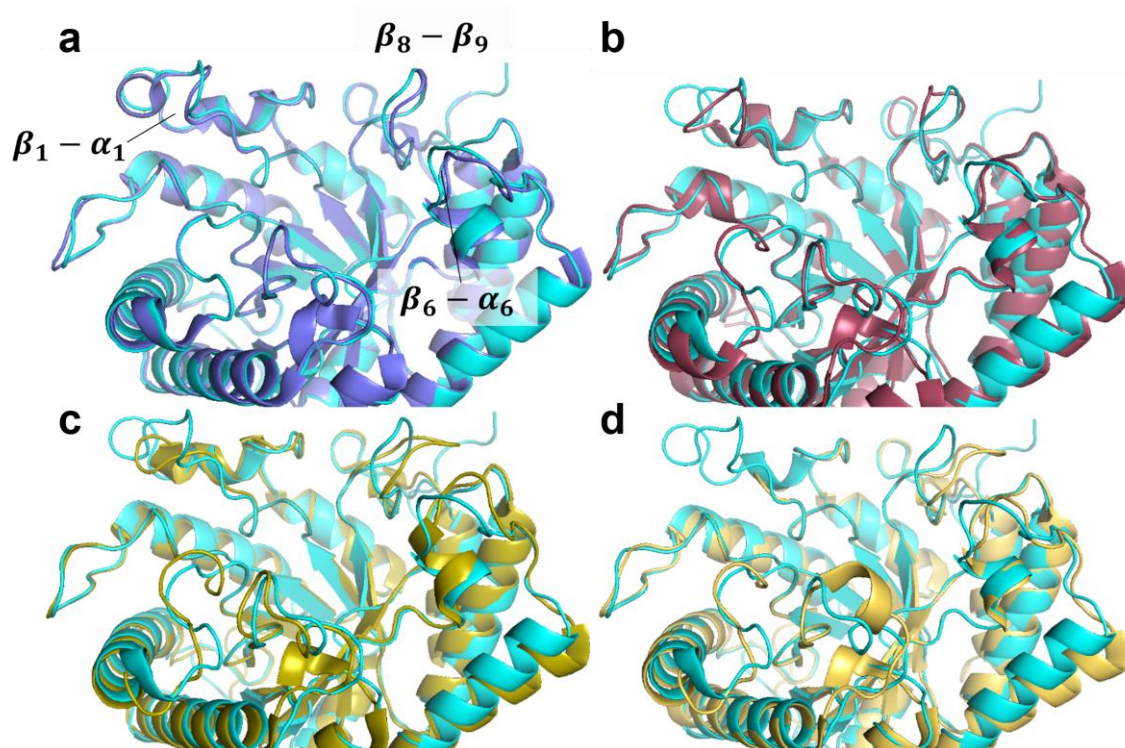


Figure S2. Structural alignment of CelD (cyan) to 3AYR (a), 6WQP (b), 6MQ4 (c), and 3NDZ (d). The loops exhibiting the greatest structural diversity are annotated and CelD is shown in the same orientation in each panel.

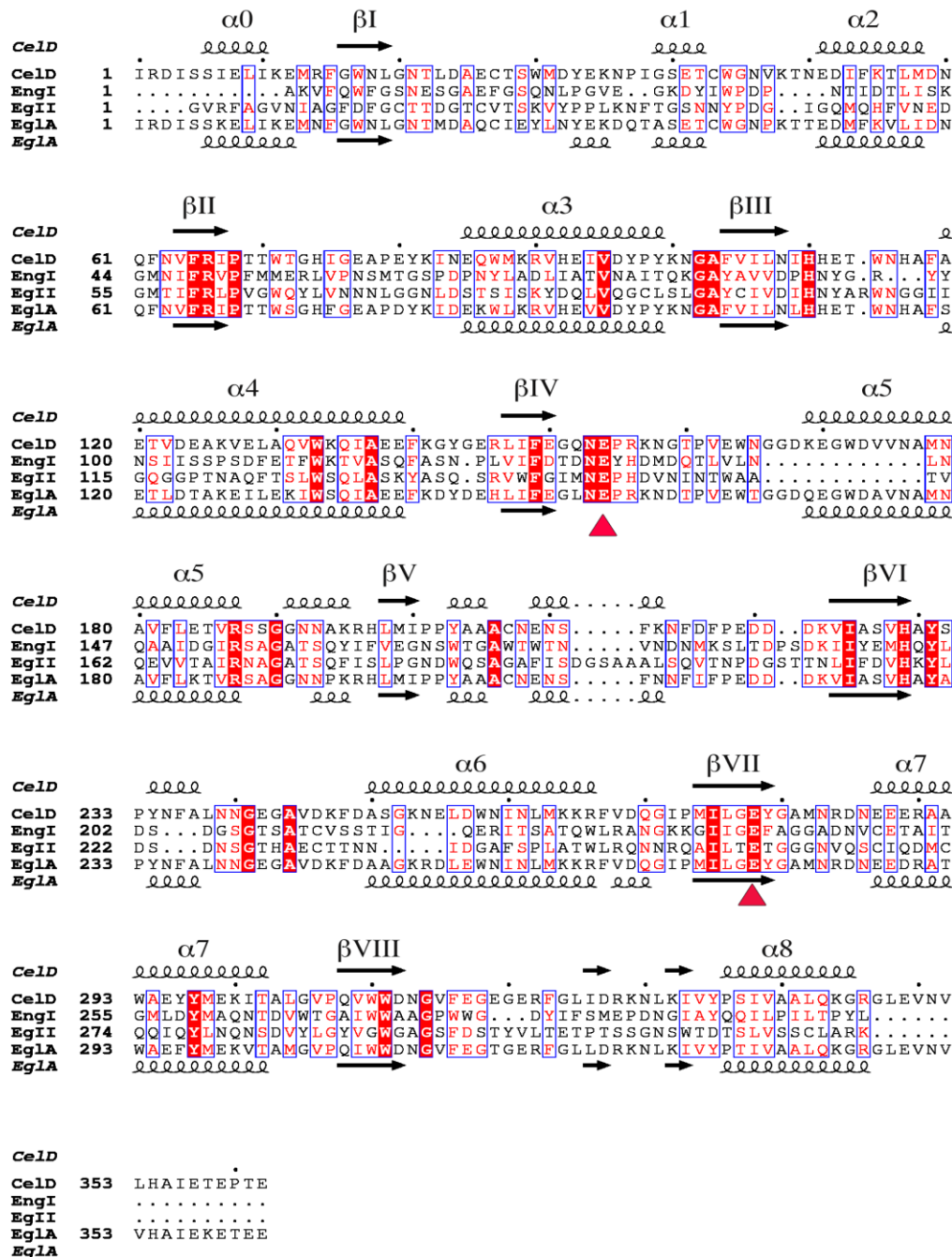
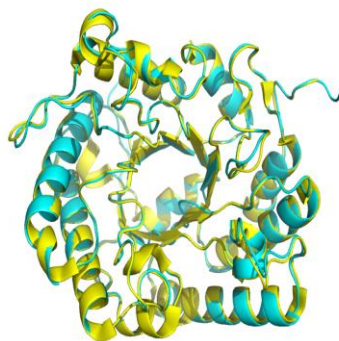


Figure S3. Sequence alignment of fungal cellulases from *Piromyces finnis* (CelD), *Thermoascus aurantiacus* (EngI), *Trichoderma reesei* (EgII), and *Piromyces rhizinflata* (EgIA). Strictly conserved residues are shown in red block, and chemically similar residues in red text. The residue numbering is shown on the left for each catalytic domain. Dashed lines indicate deletions. The acid/base and nucleophile residues are indicated below by red triangles. The secondary structure elements of CelD and EgIA are shown above and below the alignment, respectively. The sequence alignment revealed close homology between CelD and EgIA cellulases (82% amino acid residue identity). The figure was generated with ESPrpt (<http://esprpt.ibcp.fr>) (Robert and Gouet 2014).

a



b



c



Figure S4. Structural comparison of the fungal cellulase catalytic domains. (A, B, C) Superposition of the CelD structure reported in this study (cyan ribbon) and the catalytic domain of EglA (PDB code 3AYR, yellow ribbon), EgII (PDB code 3QR3, red ribbon), and EngI (PDB code 1GZJ, magenta ribbon).

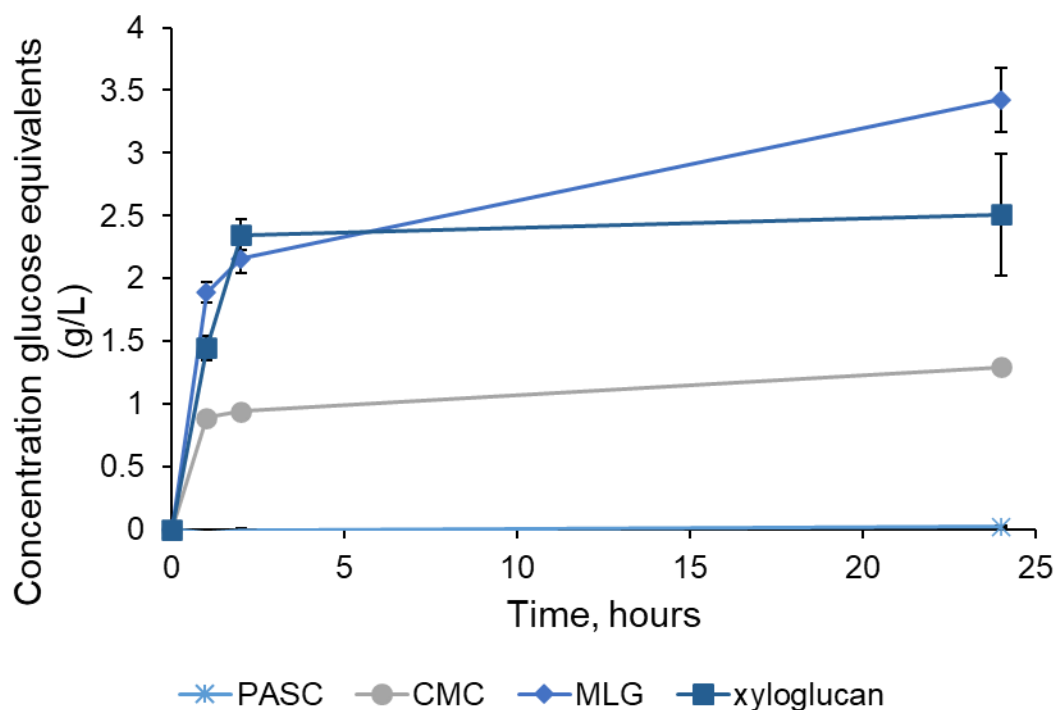


Figure S5. Timecourse measurement of reducing sugar concentration for CelD catalytic domain against (hemi)cellulosic substrates. Data are presented as the mean $\pm$ standard deviation of technical triplicates. PASC: phosphoric acid swollen cellulose; CMC: carboxymethylcellulose; MLG: mixed linkage glucan (β -D-glucan).

Supplementary References

Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42:W320–W324.
<https://doi.org/10.1093/nar/gku316>