nature portfolio

https://doi.org/10.1038/s41586-024-08553-z

Supplementary information

A metagenomic 'dark matter' enzyme catalyses oxidative cellulose conversion

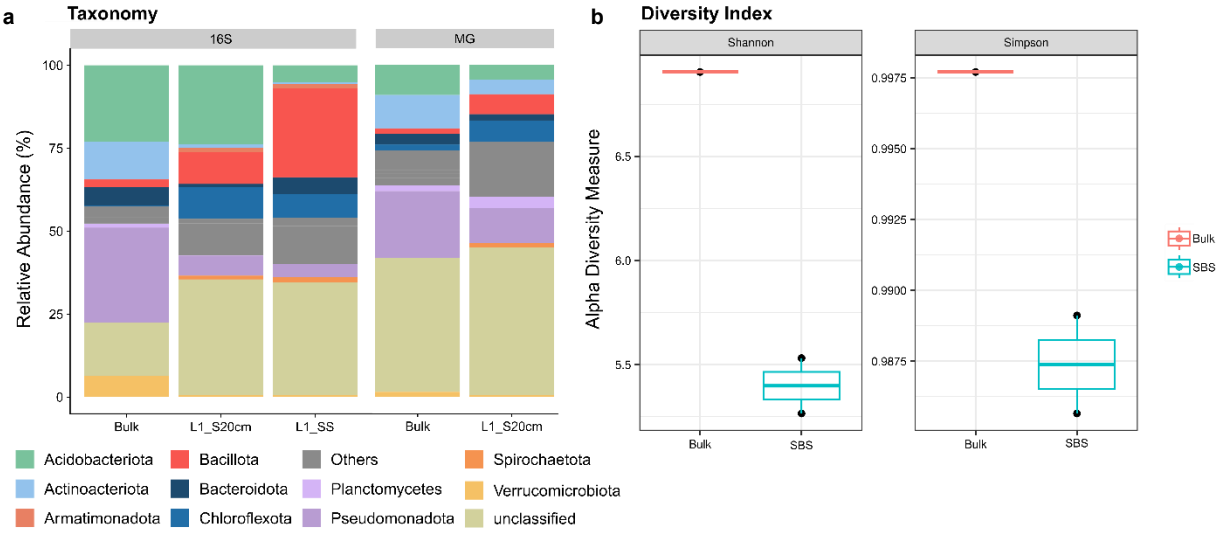
In the format provided by the authors and unedited

Supplementary Information

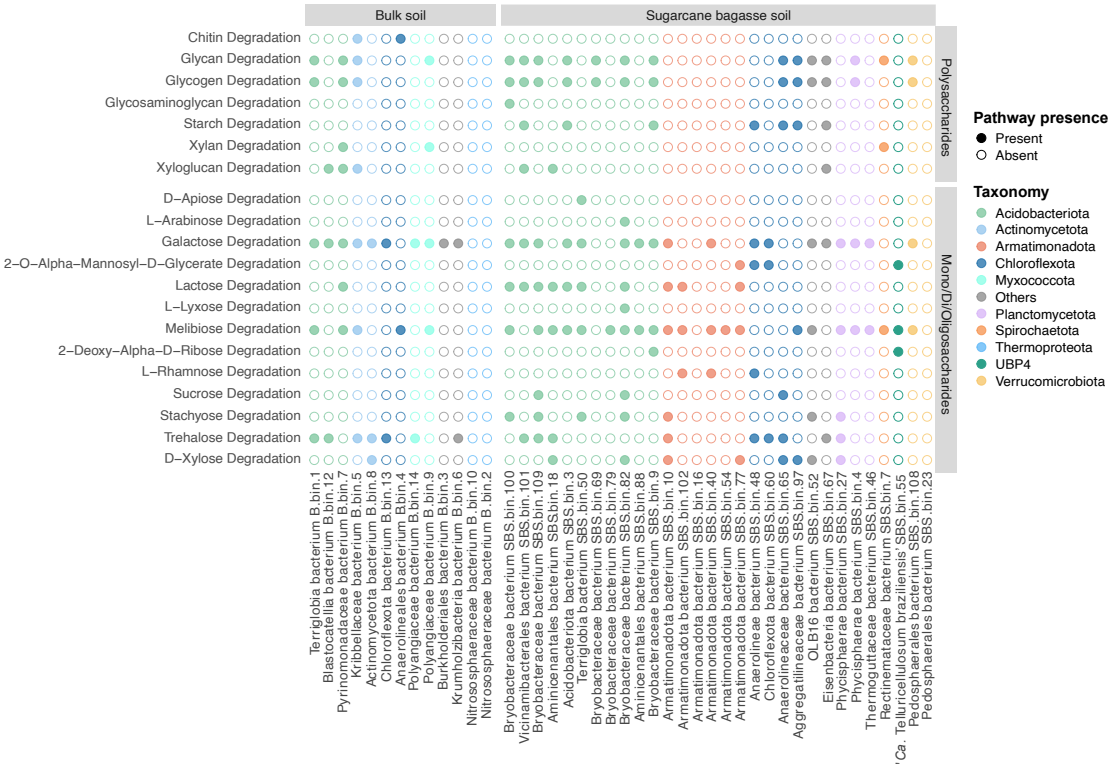
A metagenomic 'dark matter' enzyme catalyzes oxidative cellulose conversion

Clelton A. Santos[1], Mariana A. B. Morais[1], Fernanda Mandelli[1], Evandro A. Lima[1], Renan Y. Miyamoto[1], Paula M. R. Higasi[1], Evandro A. Araujo[2], Douglas A. A. Paixão[1], Joaquim M. Junior[1], Maria L. Motta[1], Rodrigo S. A. Streit[1], Luana G. Morão[1], Claudio B. C. Silva[1], Lucia D. Wolf[1], Cesar R. F. Terrasan[1], Nathalia R. Bulka[1], Jose A. Diogo[1], Felipe J. Fuzita[1], Felippe M. Colombari[1], Camila R. Santos[1], Priscila T. Rodrigues[1], Daiane B. Silva[3], Sacha Grisel[4,5], Juliana S. Bernardes[3], Nicolas Terrapon[6], Vincent Lombard[6], Antonio J. C. Filho[7], Bernard Henrissat[8], Bastien Bissaro[4], Jean-Guy Berrin[4], Gabriela F. Persinoti[1], Mario T. Murakami[1]*

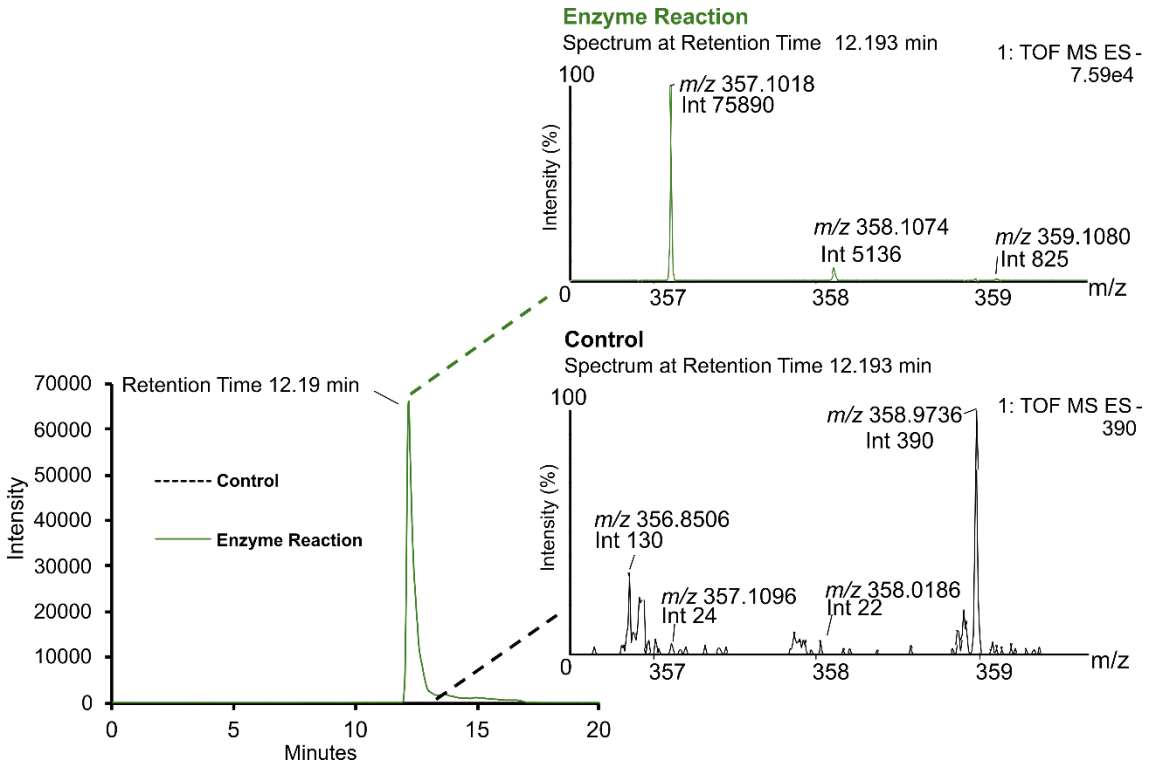
Corresponding author: mario.murakami@lnbr.cnpem.br



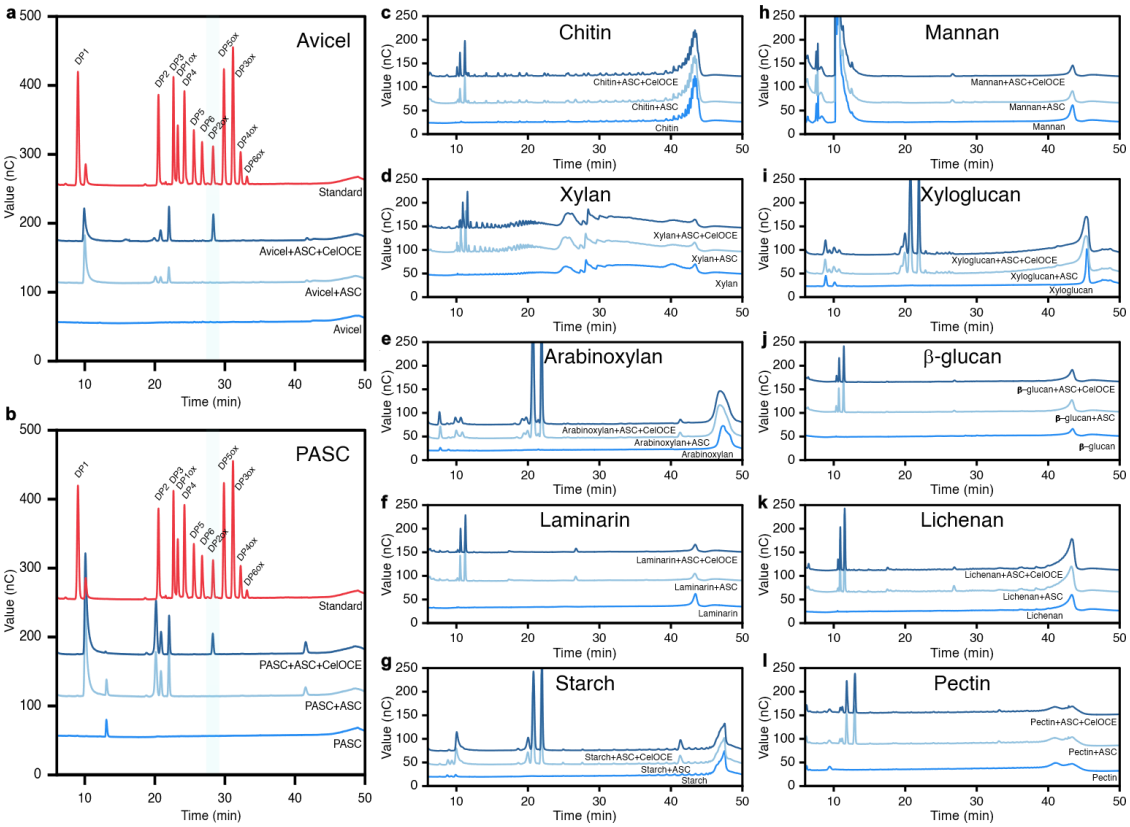
Supplementary Fig. 1. Taxonomic composition and diversity of microbial communities in sugarcane bagasse-covered soil. (**a**) Relative abundance of bacterial phyla based on 16S rRNA gene target sequencing (16S) and whole metagenome (MG) sequencing. Samples include sugarcane bagasse-covered soil collected at 20 cm depth (SBS_20cm) and at the surface (SBS_SS), as well as the bulk control soil. (**b**) Alpha diversity indices (Shannon and Simpson) calculated for each sample using the number of zero-radius operational taxonomic units (zOTUs). This panel highlights the differences in microbial community diversity between the sugarcane bagasse-covered soil and the control soil.



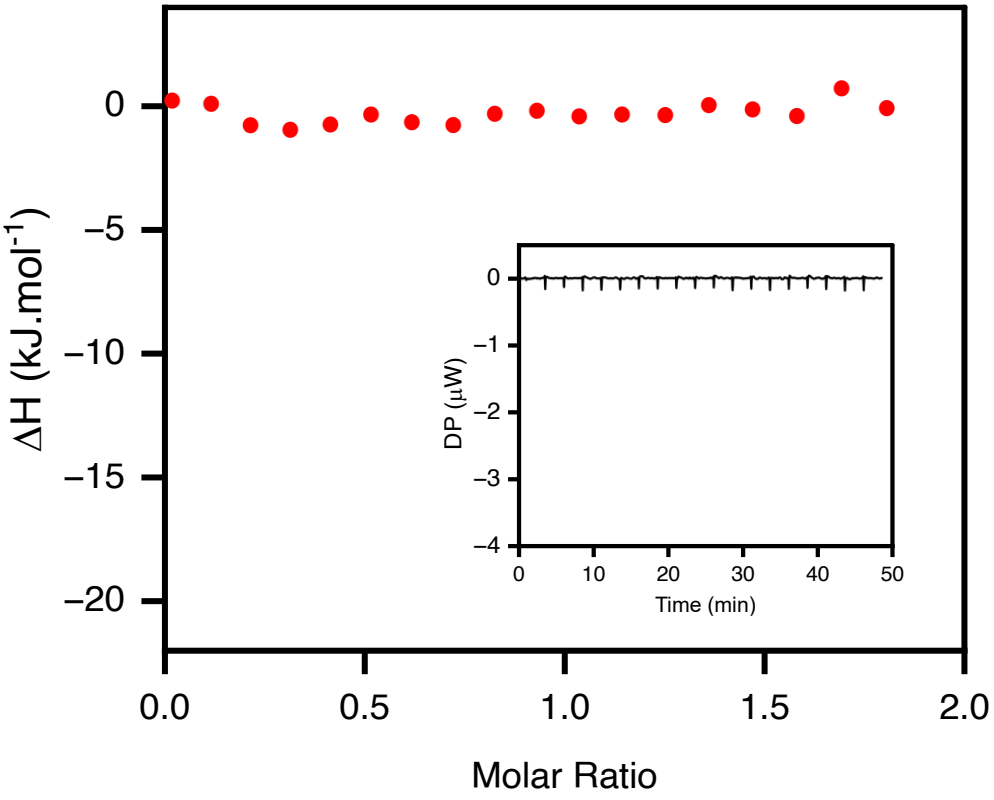
Supplementary Fig. 2. Reconstruction of carbohydrate-related metabolic pathways. Prediction of metabolic pathways related to carbohydrate breakdown for the top 30 MAGs exhibiting the highest number of predicted GHs. A pathway is predicted to be present when any MetaCyc reaction that is hierarchically under the parental pathway term displayed above was predicted by gapseq v1.1 software. SBS represents Sugarcane Bagasse-covered Soil collected 20 cm below ground, while B stands for Bulk control soil collected from the same area but not covered with sugarcane bagasse.



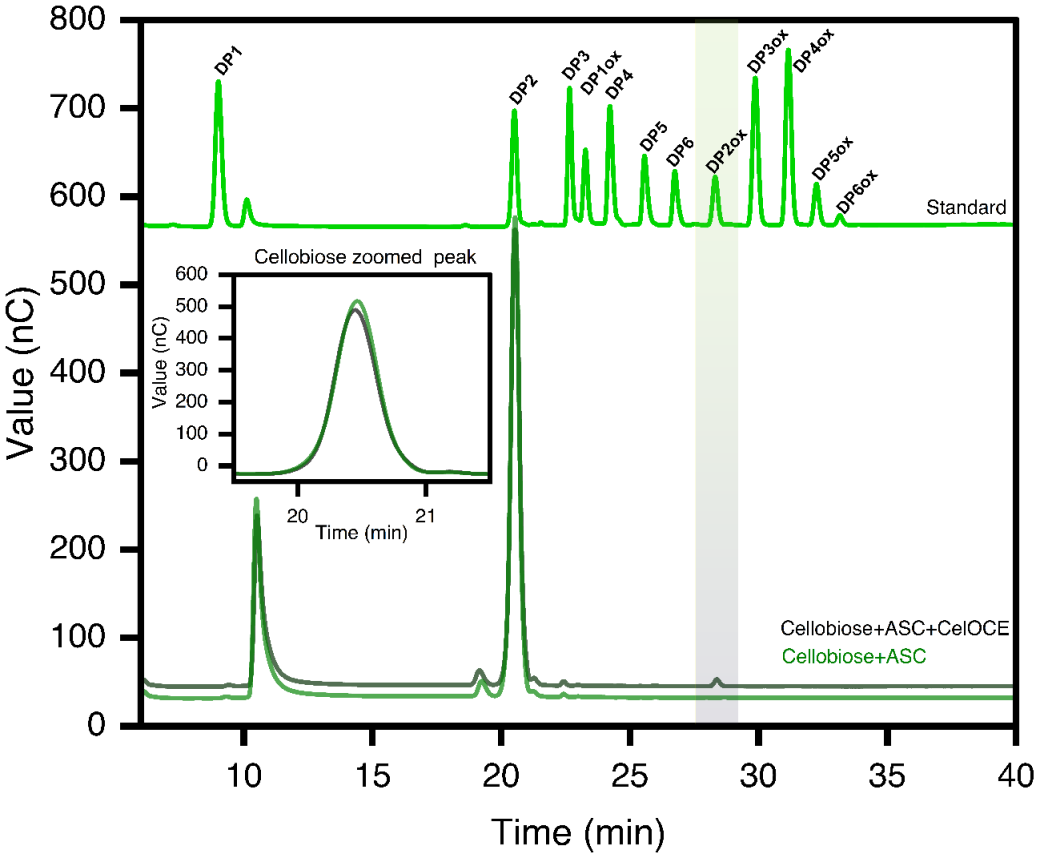
Supplementary Fig. 3. Cellobionic acid identified as a CelOCE product by mass spectrometry. Left panel, extracted ion chromatograms within 10 ppm error from the exact monoisotopic mass $[M-H]^-$ = 357.1033 Da of cellobionic acid are shown. Upper-right and lower-right panels, zoomed-in views of the *m/z* range 357-359 from the spectra corresponding to the retention time of 12.19 min for the enzyme and control reactions, respectively. The observed *m/z* of 357.1018 in the enzyme reaction corresponds to the expected isotopic distribution for cellobionic acid.



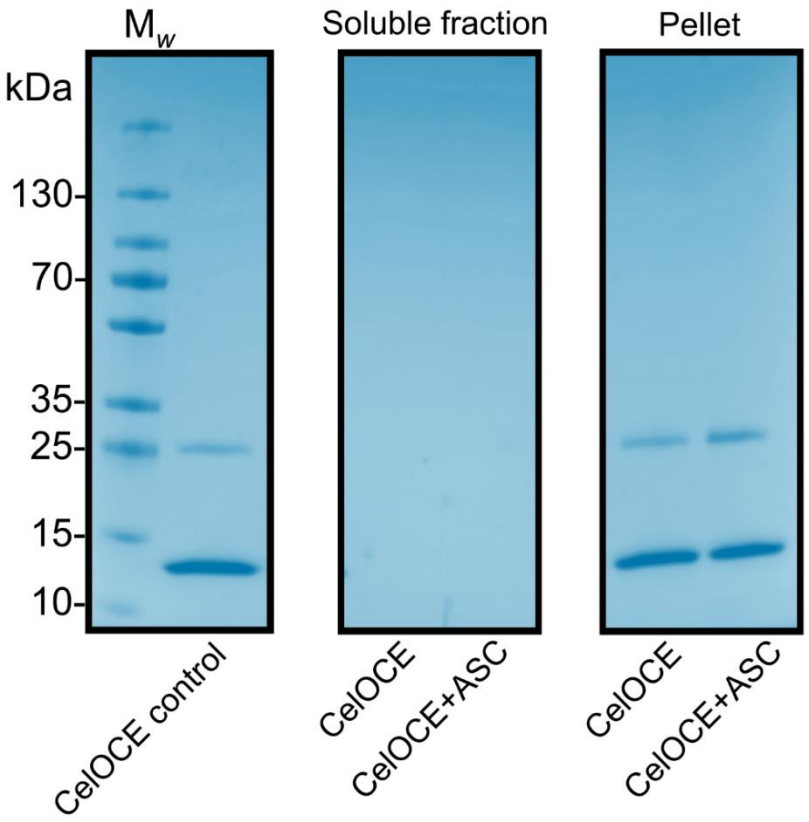
Supplementary Fig. 4. HPAEC-PAD profiles of CelOCE acting on different polysaccharides. CelOCE was incubated with Avicel (**a**), PASC (**b**), chitin (**c**), xylan (**d**), arabinoxylan (**e**), laminarin (**f**), starch (**g**), mannan (**h**), xyloglucan (**i**), mixed-linked β-glucan (**j**), lichenan (**k**) and pectin (**l**). The reaction started with the addition of 1 mM ascorbic acid. Control reactions using only the polysaccharide and ascorbic acid without enzyme were also tested. Cellobionic acid production was only detected from lignocellulose (Fig. 2b), Avicel and PASC. The chromatograms present the profile of one of the three independent experiments ($n = 3$) with each polysaccharide tested, alone, and in the presence of ascorbic acid (ASC) and ASC plus CelOCE.



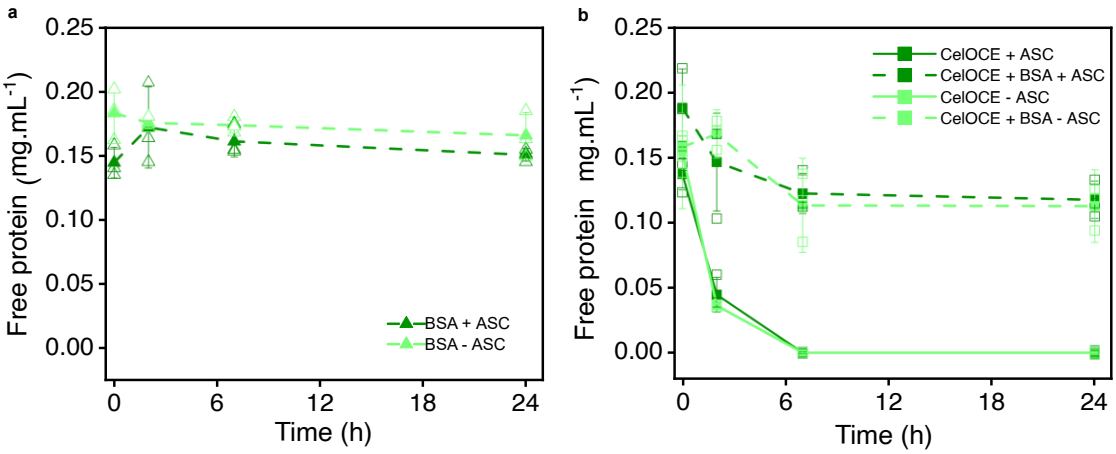
Supplementary Fig. 5. Isothermal titration calorimetry analysis of cellobiose binding to CelOCE. The ITC thermogram demonstrates a lack of detectable heat change upon cellobiose titration into CelOCE, indicating no significant binding affinity under the tested conditions**.**



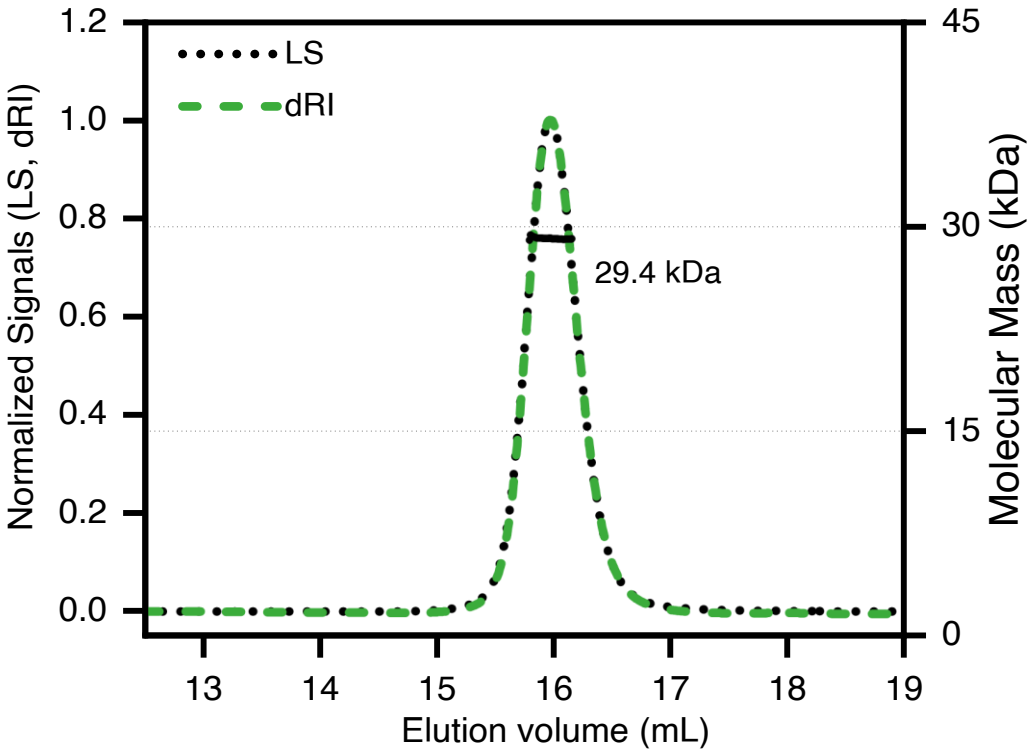
Supplementary Fig. 6. HPAEC-PAD chromatograms of control and experimental reactions with CelOCE and 50 µM cellobiose as a substrate. Standards of non-oxidized and oxidized cellooligosaccharides are shown for reference. These data highlight no relevant cellobiose consumption after 16 h reaction time. Only residual presence of cellobionic acid was observed under the tested conditions.



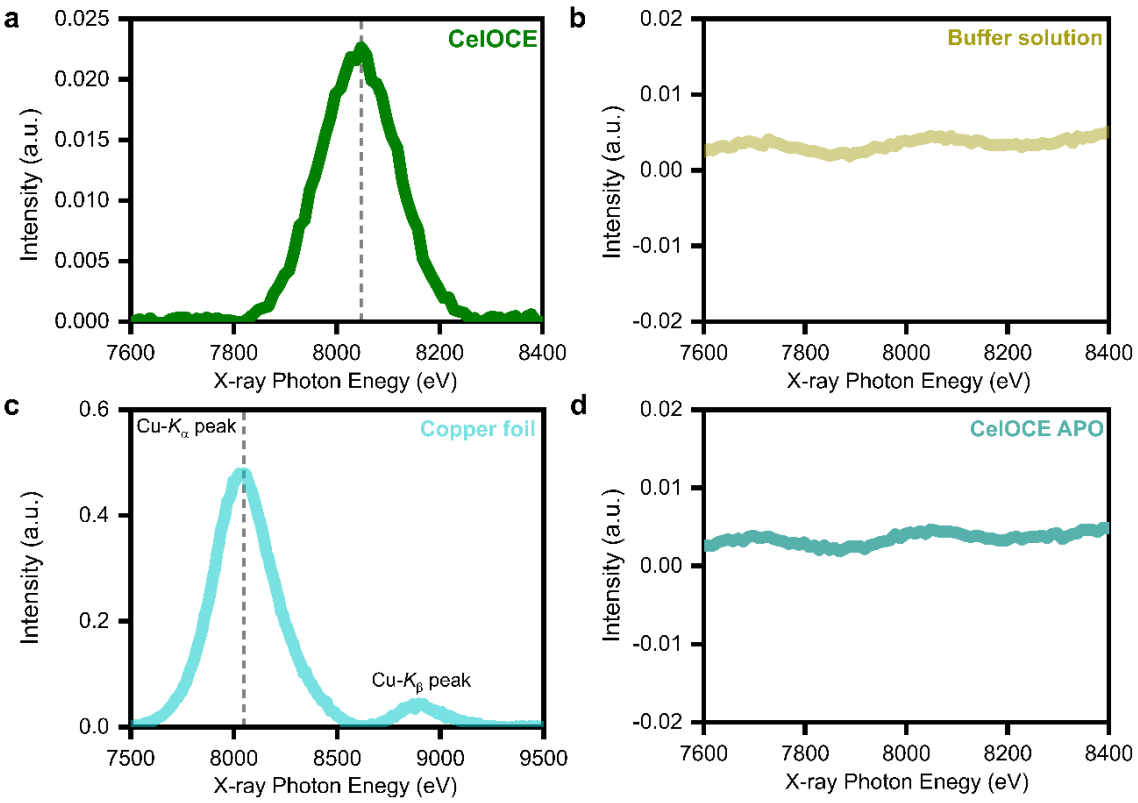
Supplementary Fig. 7. Qualitative CelOCE binding assay by affinity gel electrophoresis. Binding of CelOCE to microcrystalline cellulose (Avicel) in the presence and absence of ascorbic acid. The presence of a minor second band in the gel likely represents the dimeric form of the enzyme, which hydrodynamic studies indicate is the predominant conformation in solution. Uncropped gel image is presented in Supplementary Fig. 30a.



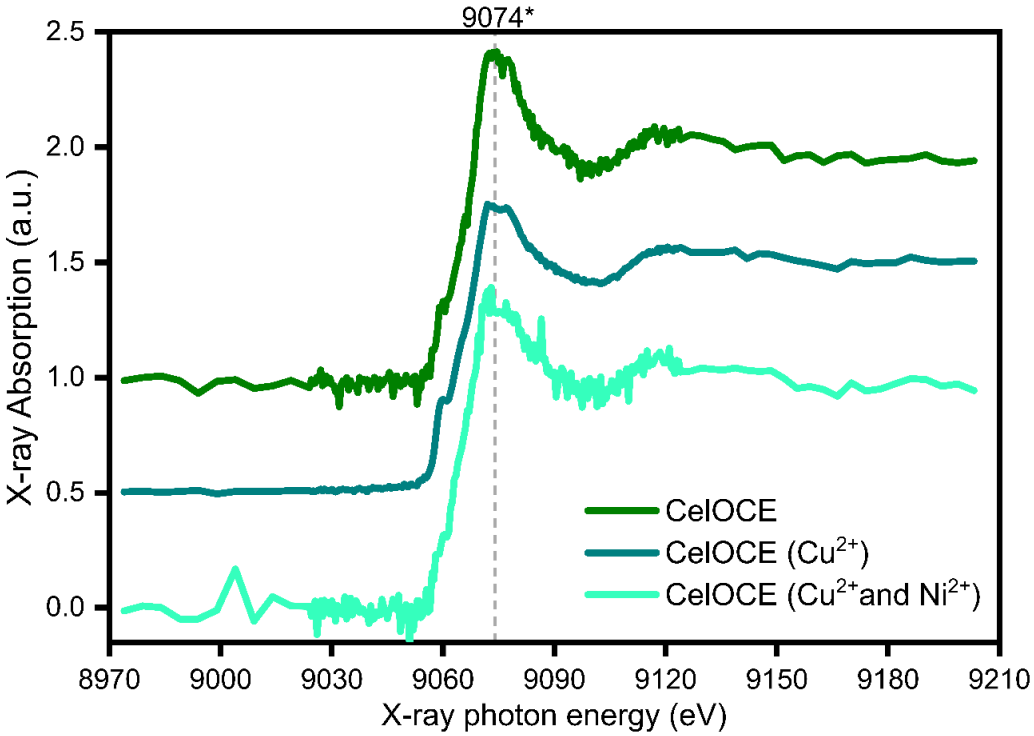
Supplementary Fig. 8. CelOCE competitive binding assay. Binding capacity of CelOCE to 5% (*w/v*) Avicel in the presence of bovine serum albumin (BSA) as blocking agent. (**a**) Control experiments were conducted to assess the individual binding of BSA to Avicel. These experiments involved incubating BSA alone with Avicel, ensuring that BSA did not bind to Avicel. **(b)** Comparison of the binding capacity of oxidized and reduced CelOCE to Avicel, both in the presence and absence of BSA. All reactions were performed at the same mass concentration of CelOCE (0.17 mg/mL). To visualize the specific binding of CelOCE in the presence of BSA, the BSA-only curves were subtracted from the corresponding "CelOCE + BSA" curves. Results are expressed as mean ± standard deviation from three independent experiments ($n = 3$).



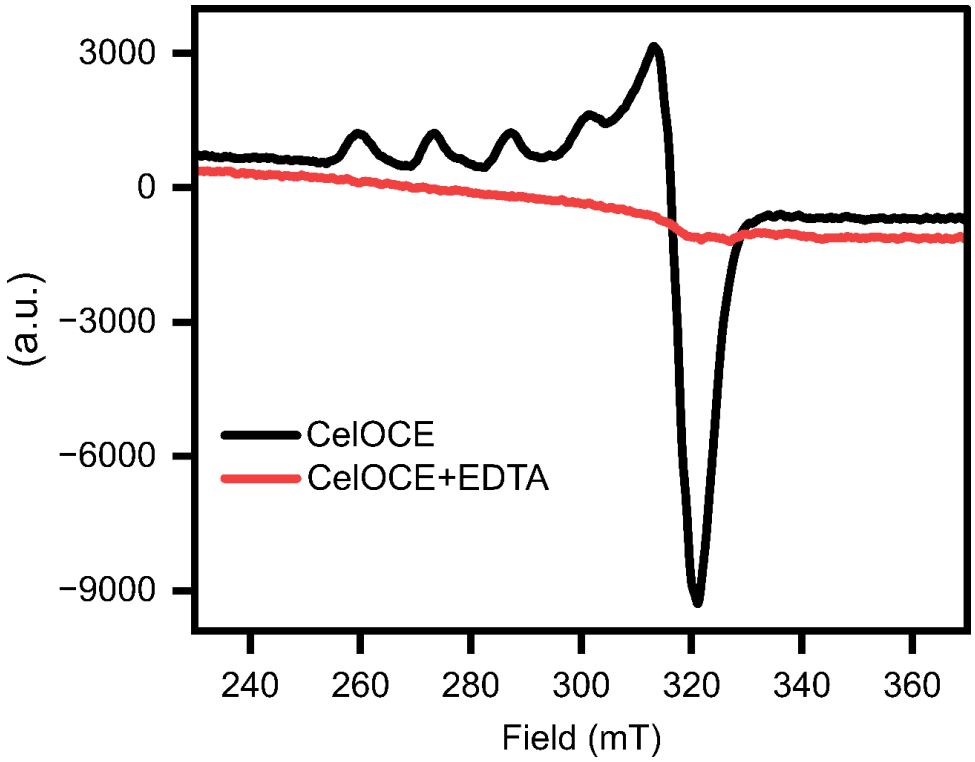
Supplementary Fig. 9. CelOCE exists as a dimer in solution. Analytical size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) analysis of CelOCE using a Superdex 200 column. The chromatograms show the light scattering signal at a 90° angle (green), the refractive index (black), and the calculated molar mass for each peak. Data of three independent experiments ($n = 3$) confirm that CelOCE predominantly exists as a dimer in solution.



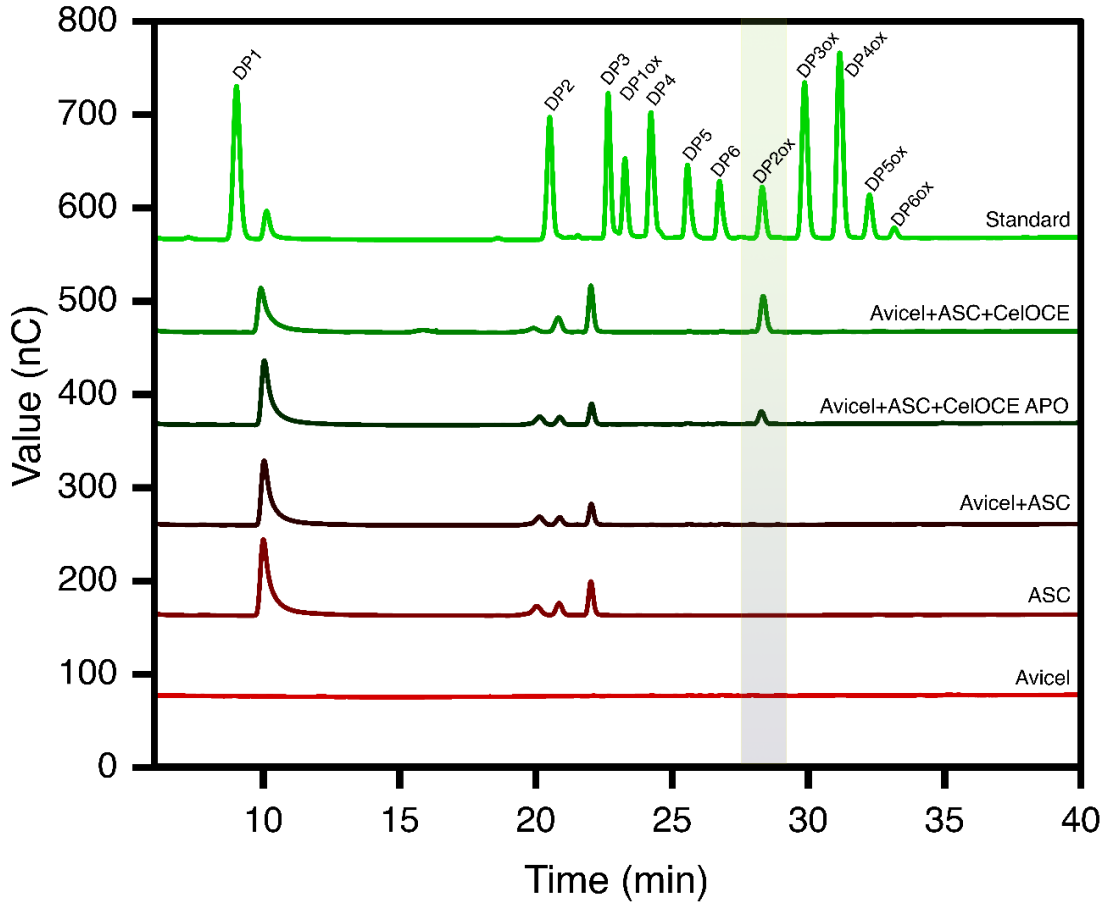
Supplementary Fig. 10. Synchrotron X-ray fluorescence spectroscopy (XRF) for accurate identification of copper binding to CelOCE. (**a**) XRF spectra of CelOCE, (**b**) buffer solution, (**c**) copper foil, and (**d**) EDTA-chelated CelOCE. The dashed line in panels A and C indicates the copper *k*-edge at 8047.94 eV, a characteristic energy for copper absorption. The presence of this peak in the CelOCE spectrum (**a**) but not in the control or chelated enzyme spectra (B and D) supports the specific binding of copper to CelOCE. Data consists of three independent experiments ($n = 3$).



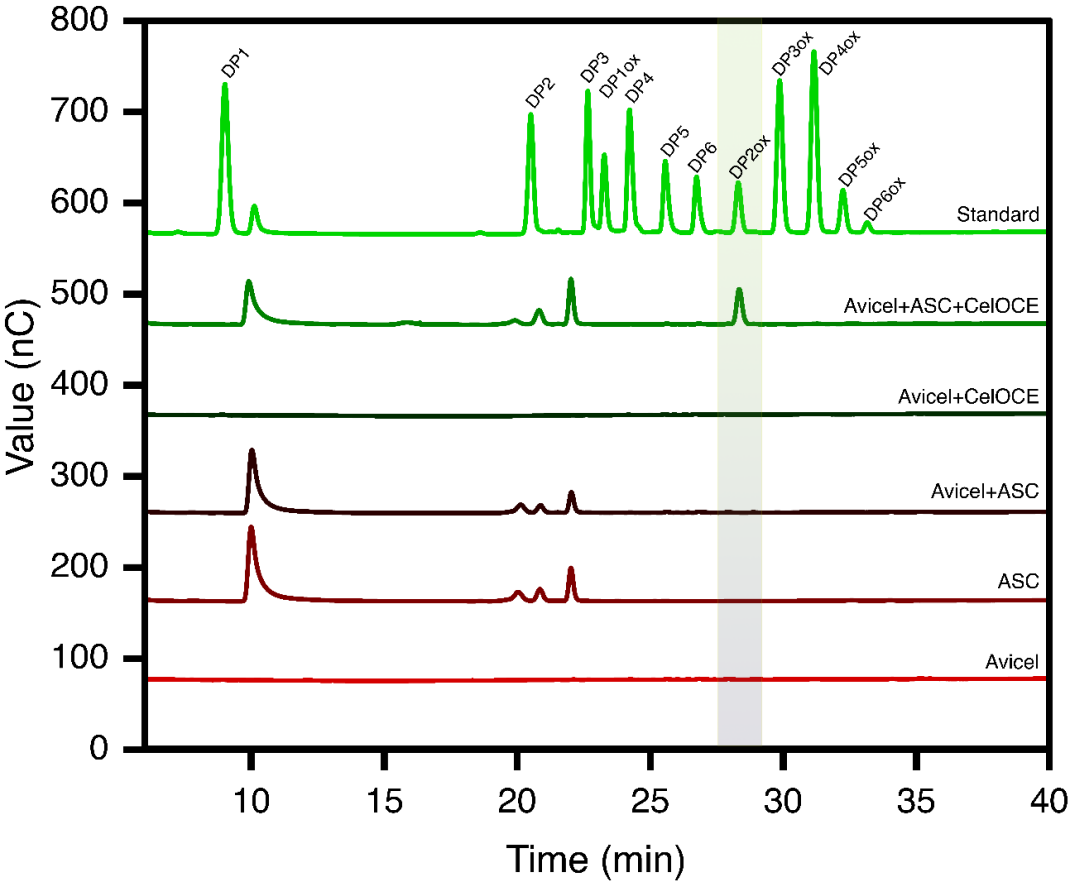
Supplementary Fig. 11. CelOCE synchrotron X-ray absorption spectroscopy (XAS) for copper identification. XAS spectra of CelOCE samples at the k_α absorption edge of copper. The data collection for CelOCE without copper addition is represented by the green line. Further XAS measurements were conducted after incubating CelOCE with 1 mM $CuSO_4$ (red line) or 1 mM $NiSO_4$ (blue line) for 4 h at 20 °C. The asterisk (*) marks the copper *k*-edge for the copper metal foil used for beamline calibration. Data consists of three independent experiments ($n = 3$).



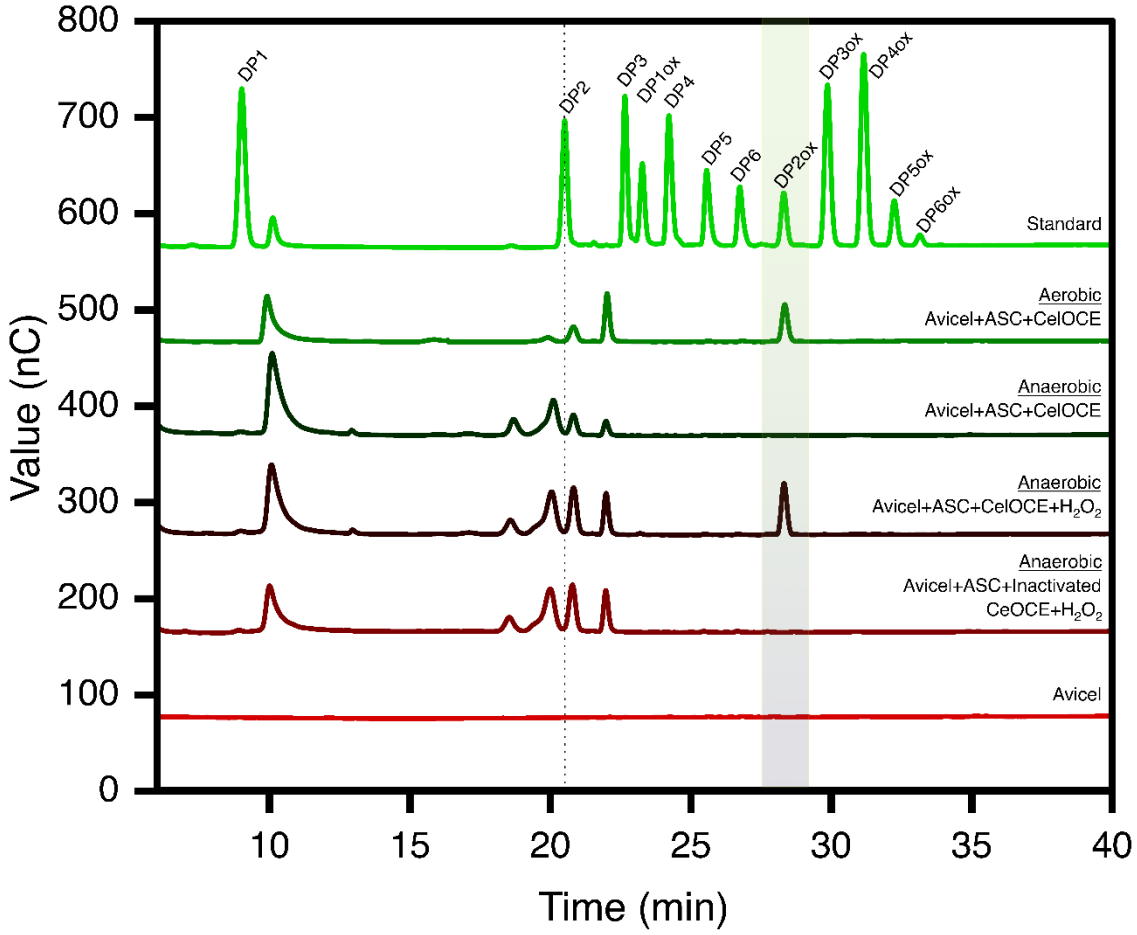
Supplementary Fig. 12. Effect of EDTA on CelOCE EPR Spectrum. Electron Paramagnetic Resonance (EPR) spectra of CelOCE with and without the addition of EDTA (ethylenediaminetetraacetic acid), a chelating agent known to bind metal ions. The presence or absence of the characteristic Cu^{2+} signal in the EPR spectra supports the specific binding of copper to CelOCE and its removal upon EDTA treatment.



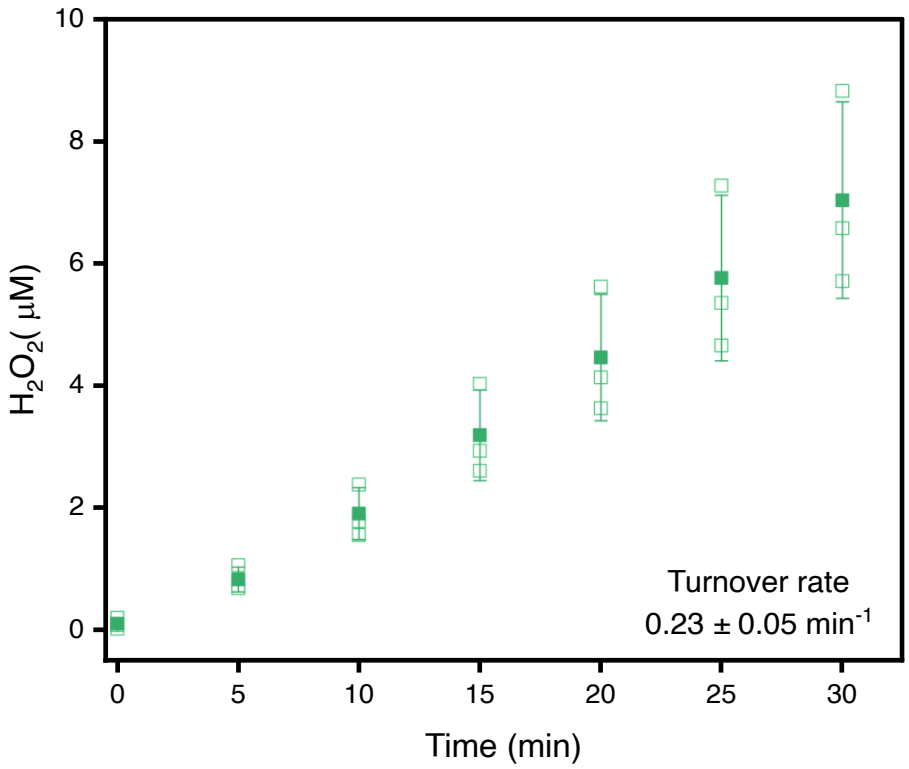
Supplementary Fig. 13. Effect of EDTA on CelOCE activity. HPAEC-PAD chromatograms show the products released from Avicel by CelOCE under different conditions. The near-complete absence of cellulose cleavage products in the presence of the chelating agent EDTA (CelOCE APO) compared to native CelOCE highlights the role of copper in CelOCE activity. Control reactions using Avicel and ascorbic acid (ASC) without enzyme were also included. Standard cellooligosaccharides are represented by the light green line, providing a reference for product identification.



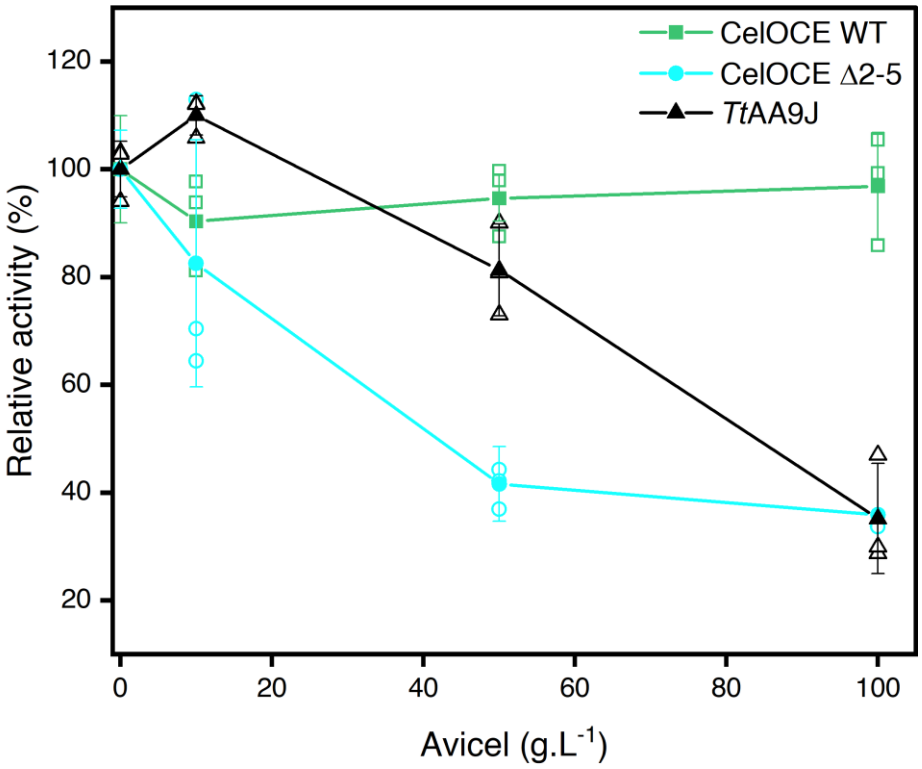
Supplementary Fig. 14. The role of an electron donor in CelOCE activity. HPAEC-PAD profiles illustrating the impact of an electron donor (ascorbate, ASC) on CelOCE activity when using Avicel as a substrate. Cellobionic acid, the sole product of CelOCE-mediated cellulose oxidation, was detected exclusively in reactions containing both the enzyme and the electron donor. Control reactions lacking either the enzyme or ascorbate did not yield cellobionic acid. Standard cellooligosaccharides are provided as a reference (light green line) for peak identification.



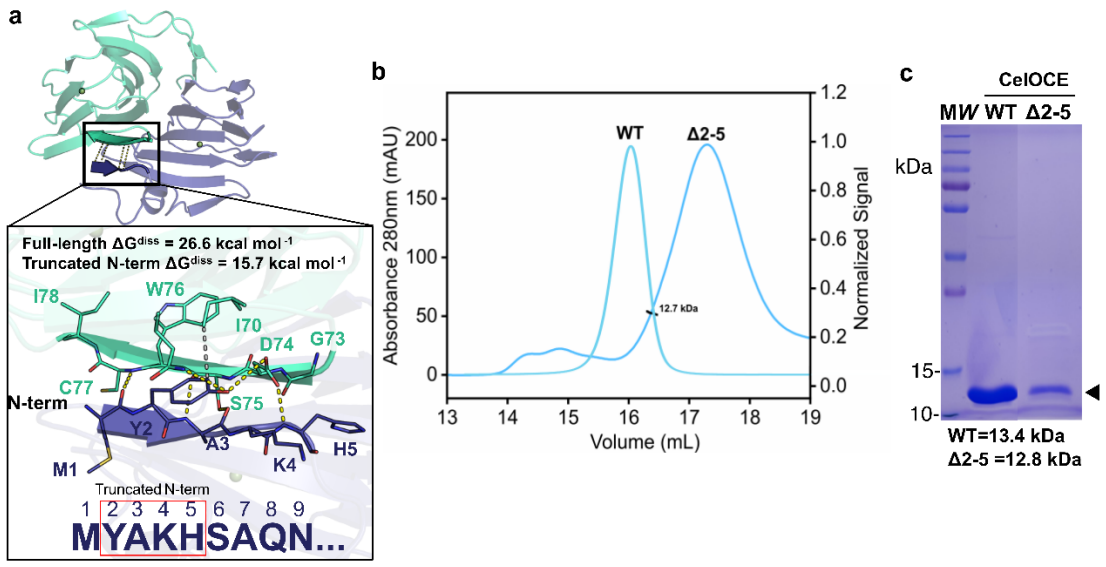
Supplementary Fig. 15. CelOCE activity under anaerobic conditions. Product profile of CelOCE acting on Avicel under anaerobic conditions, as assessed by HPAEC-PAD. In the absence of oxygen, no cellulose degradation products were detected, indicating that CelOCE is inactive under strictly anaerobic conditions. However, the addition of exogenous hydrogen peroxide restores enzymatic activity, leading to the release of detectable products. This observation supports that CelOCE functions as a peroxygenase, requiring hydrogen peroxide as a co-substrate for catalysis.



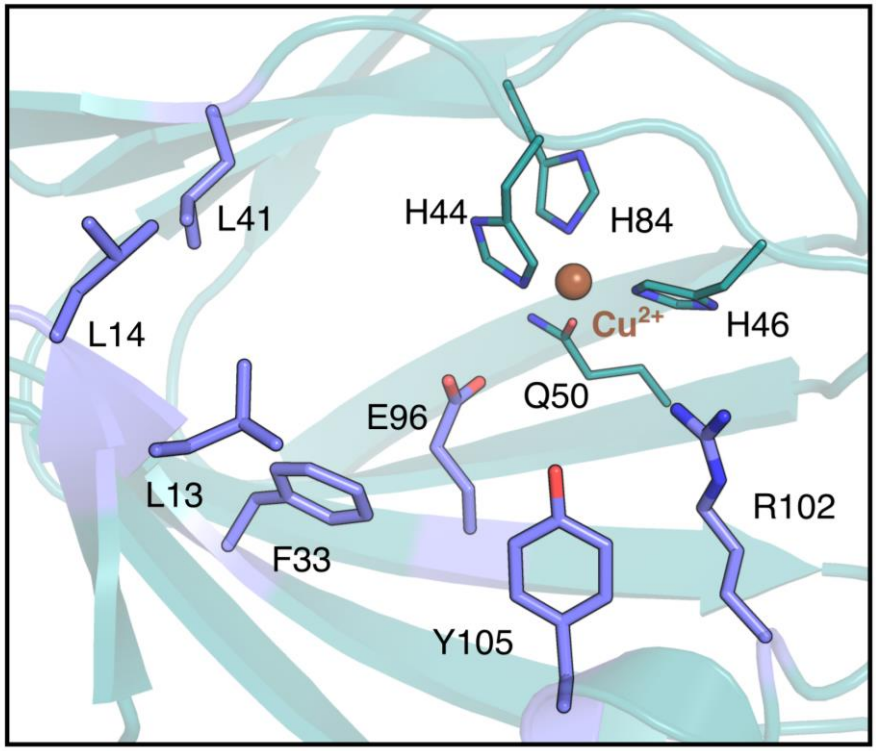
Supplementary Fig. 16. Hydrogen peroxide generation by CelOCE. 1 µM CelOCE was incubated with 1 mM ascorbic acid, and H_2O_2 generated was detected with Amplex Red/horseradish peroxidase. Results are expressed as mean ± standard deviation from three independent experiments (n = 3).



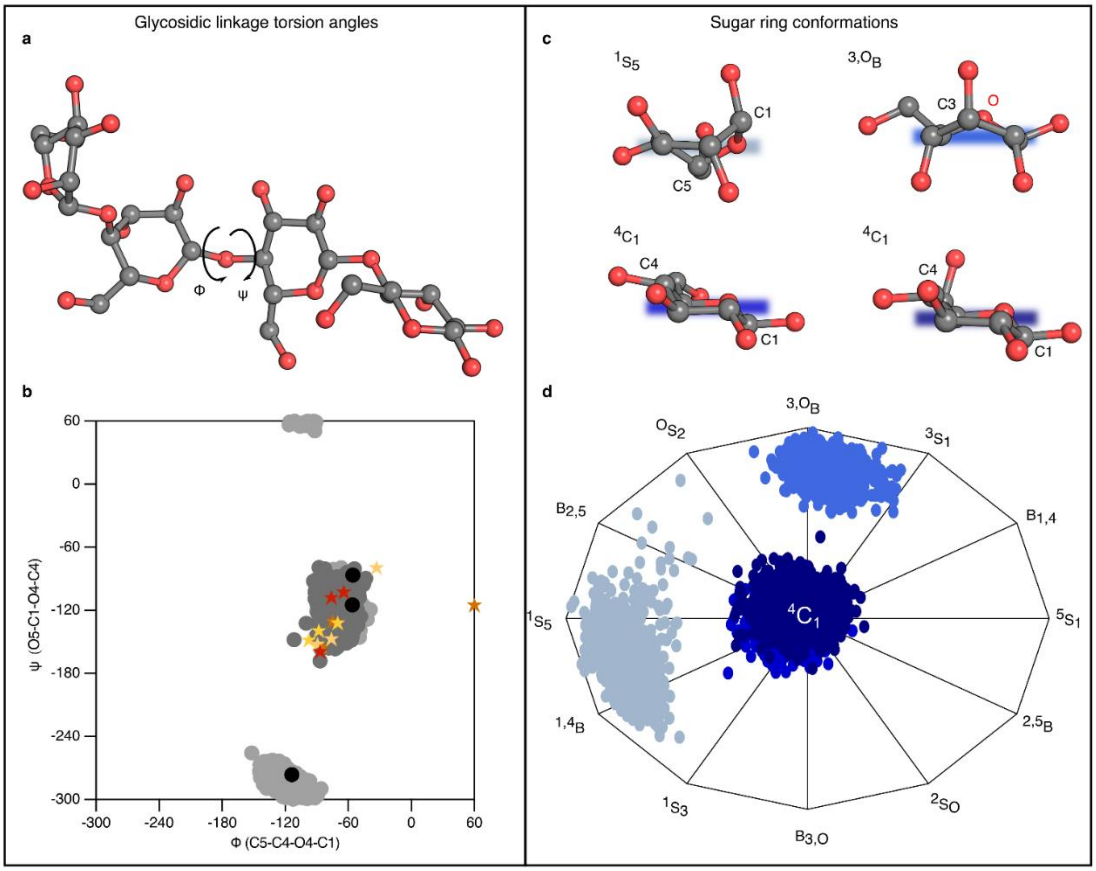
Supplementary Fig. 17. Influence of Avicel concentration on hydrogen peroxide production by native (dimeric) and monomeric (Δ2-5 mutant) forms of CelOCE, and *Tt*AA9J. While *Tt*AA9J and the monomeric CelOCE variant exhibited decreased H_2O_2 production with increasing Avicel levels, the dimeric CelOCE showed no such dependence. This result supports the importance of dimerization for *in situ* hydrogen peroxide generation. Results are expressed as mean ± standard deviation from three independent experiments (n = 3).



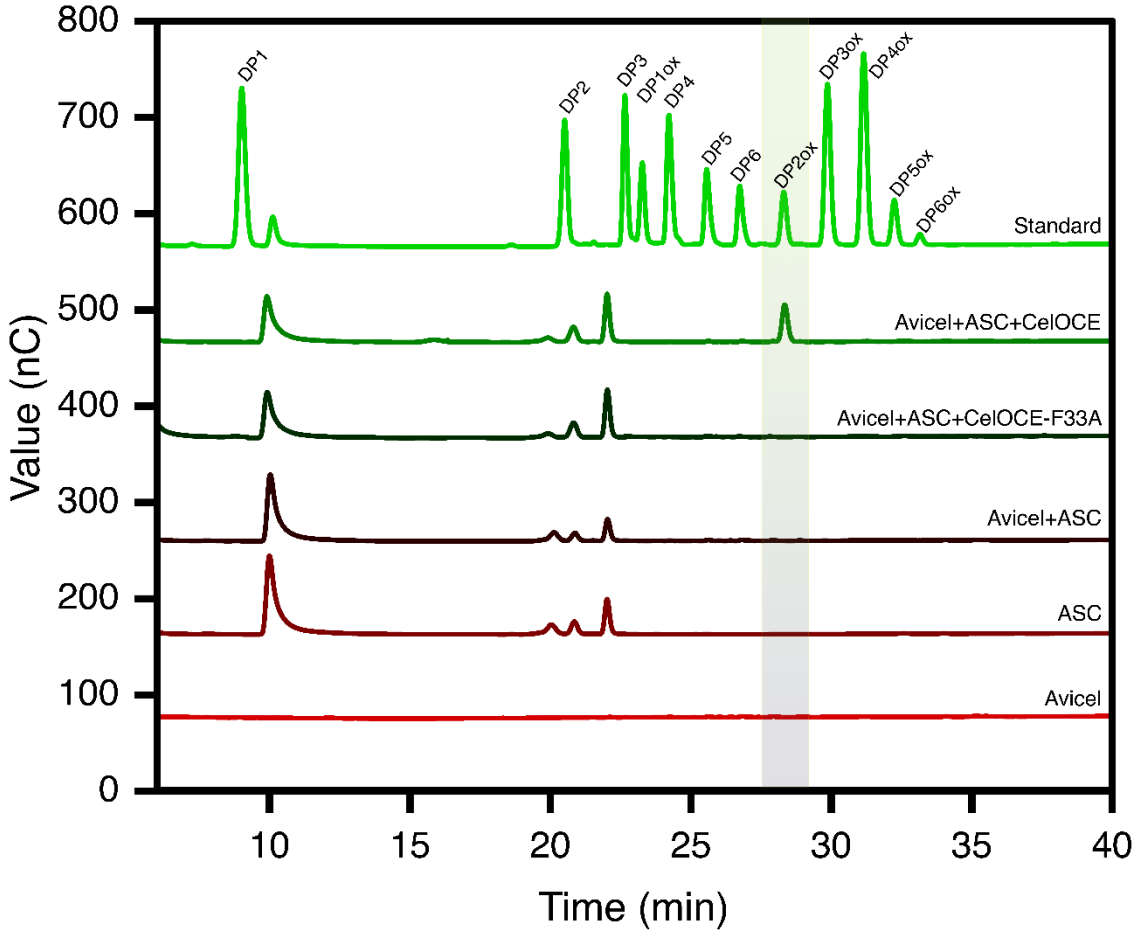
Supplementary Fig. 18. Disruption of CelOCE Dimerization through N-terminal Truncation. (a) Structural basis for CelOCE dimerization. The dimeric structure is primarily stabilized by polar and hydrophobic interactions between side chains of residues within the 6-strand β-sheet of each subunit, and the formation of an antiparallel β-sheet between the N-terminal β-strand of one subunit and the β7-strand of the other. To disrupt this dimeric assembly, the N-terminal β-strand (amino acids 2 to 5) was truncated. The calculated ΔG^{diss} value indicates a substantial decrease in dimer stability upon truncation. **(b)** SEC-MALS analysis confirms the monomerization of the enzyme resulting from the N-terminal truncation. **(c)** SDS-PAGE gel comparing the purified samples of wild-type CelOCE (13.4 kDa) and the Δ2-5 variant (12.8 kDa), further supporting the successful disruption of dimer formation. Uncropped gel image is presented in Supplementary Fig. 30b.



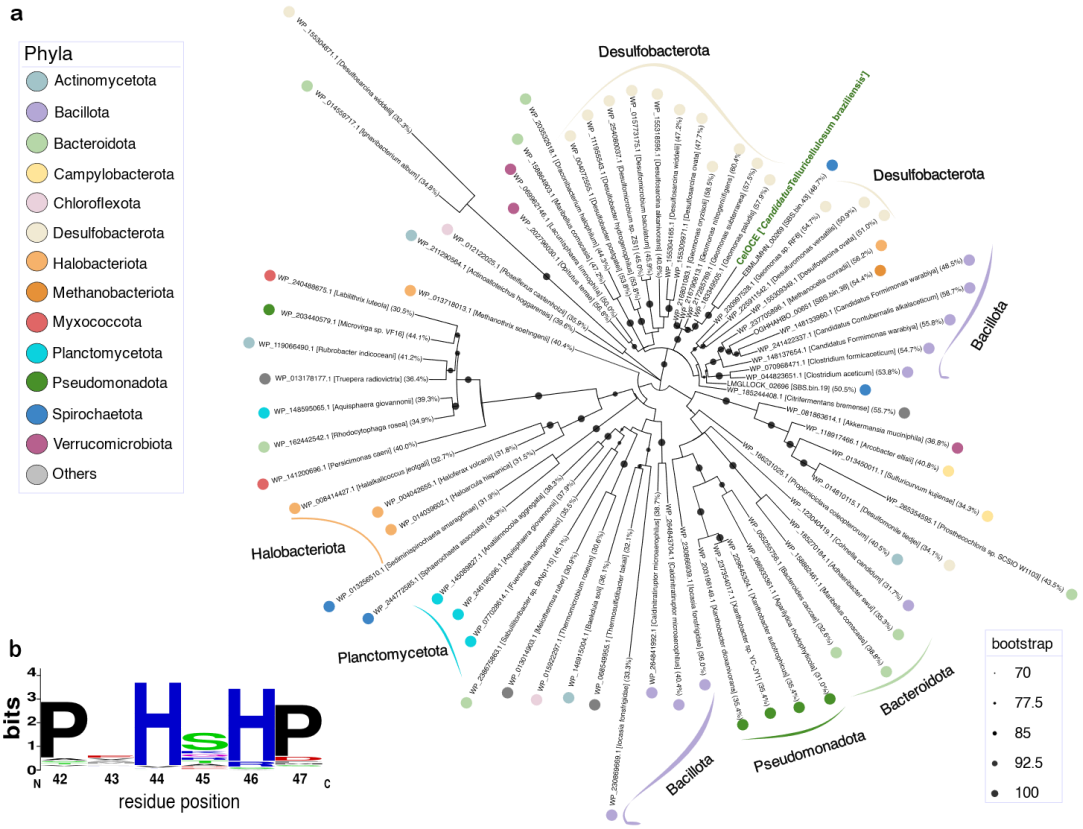
Supplementary Fig. 19. Detailed view of CelOCE active site, exhibiting the residues that constitute its microenvironment. These residues include the copper-coordinating residues (H44, H46, H84, and Q50), as well as additional residues like Y105, F33, R102, L13, L14, L41, and E96. Protein residues are depicted as sticks, and the copper is shown as an orange sphere.



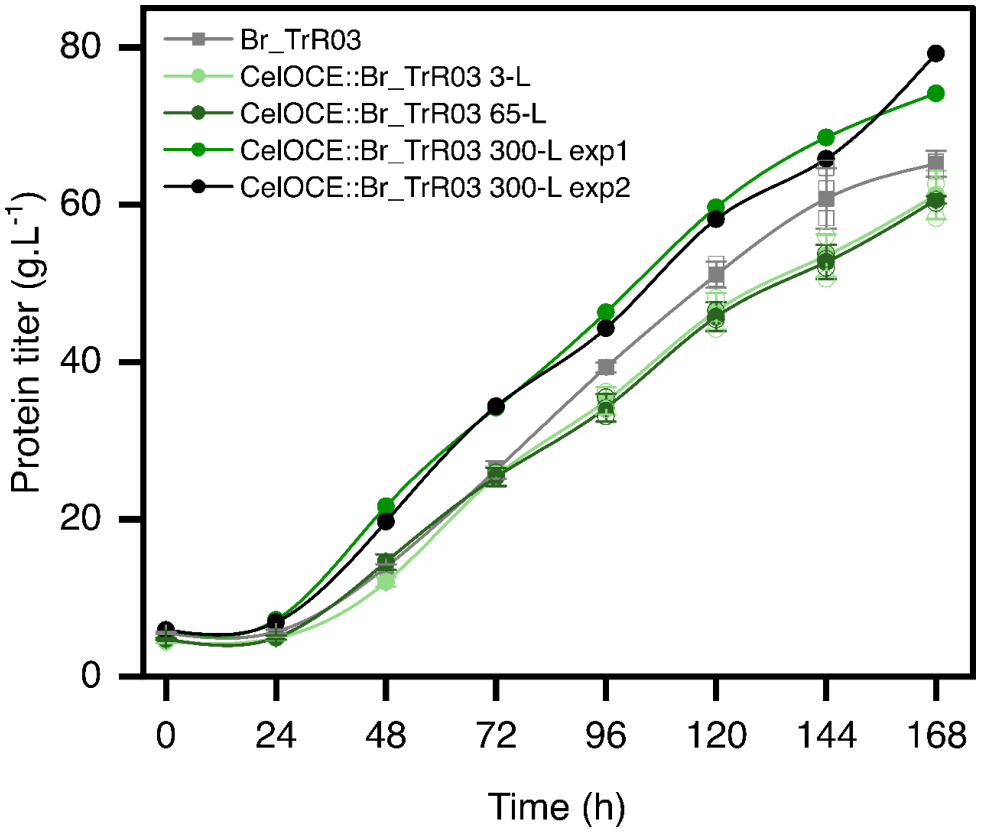
Supplementary Fig. 20. Glycosidic linkage torsion angles and sugar ring conformations for the saccharide docked in CelOCE. (a) Representative frame from the equilibrium procedure, highlighting the torsion angles (O5-C1-O4-C4) and (C5-C4-O4-C1). **(b)** Torsion angles explored during the molecular dynamics trajectory post-docking, considering the three linkages in the saccharide. Stars indicate experimental data retrieved from PDB structures, and the black circles represent the angles present in the frame shown in (a). **(c)** Sugar conformations of each subunit in the saccharide, as represented in (a). **(d)** Stoddart diagram depicting the sugar conformations explored during the same trajectory as in panel b.



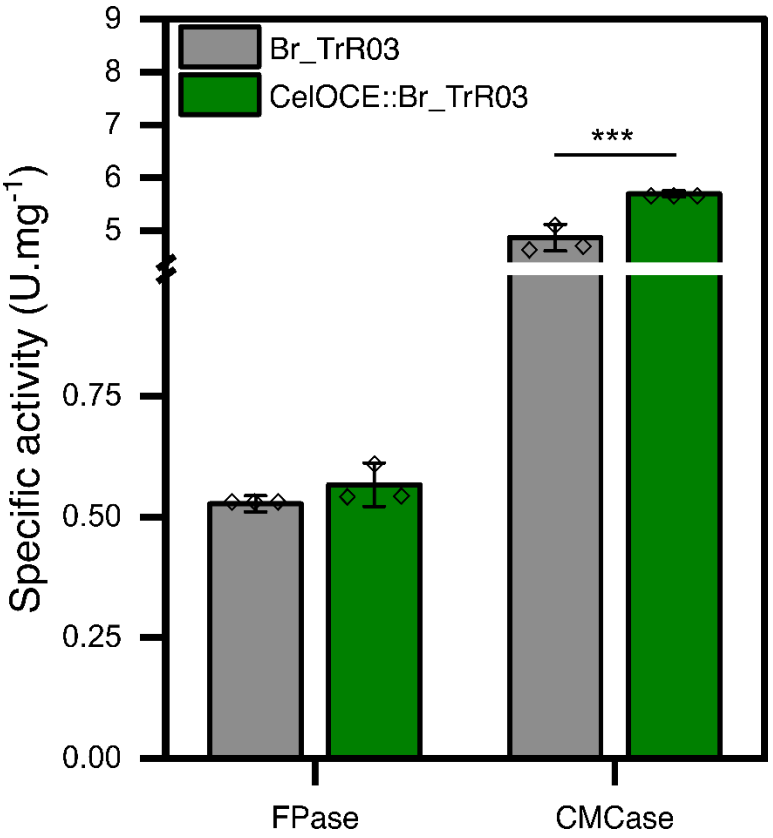
Supplementary Fig. 21. Impact of F33A mutation on CelOCE Activity. HPAEC-PAD profiles comparing the products released from cellulose by wild-type CelOCE and the CelOCE-F33A variant. Structural-guided studies of CelOCE indicate that the -1 glucosyl residue stacks with F33, contributing to proper substrate binding and cleavage. Supporting this, experimental data showed that the F33A mutation abolished the CelOCE catalytic activity.



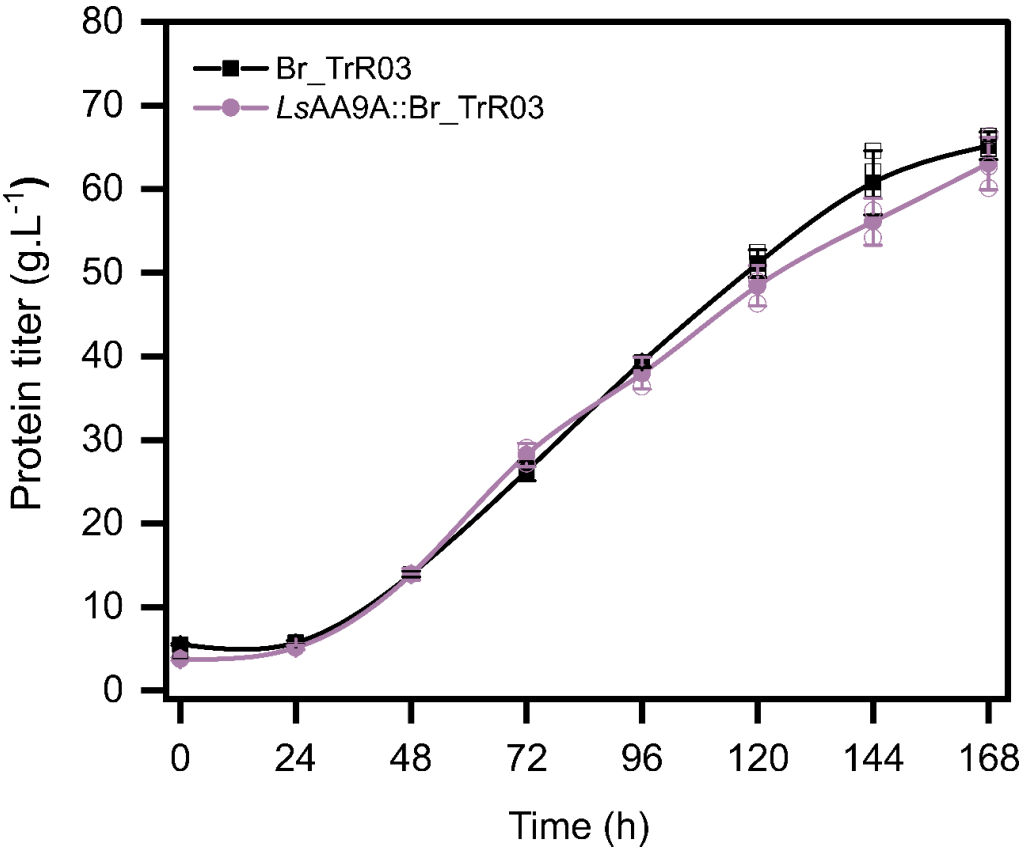
Supplementary Fig. 22. Phylogenetic analysis of CelOCE orthologs. (a) A phylogenetic tree was constructed using 72 representative sequences from complete genomes available in the NCBI RefSeq database, including sequences recovered from metagenome-assembled genomes (MAGs) in this study. The sequence identity of each ortholog to CelOCE is indicated in parentheses. **(b)** This panel highlights the conservation of proline residues in CelOCE orthologs, particularly within the motif PxHxHP. This motif contains two histidine residues (H44 and H46) that are crucial for copper coordination in the active site.



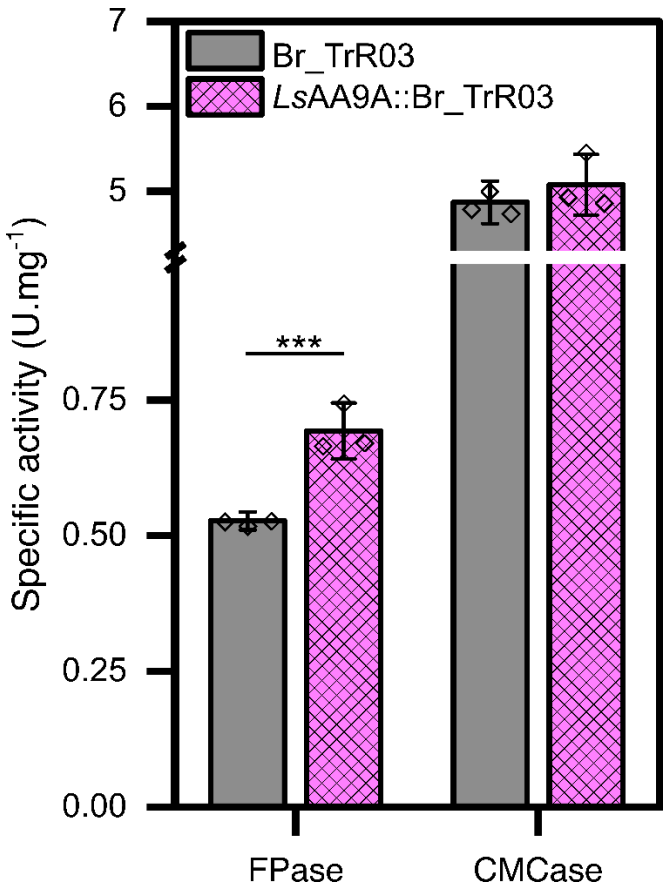
Supplementary Fig. 23. Extracellular protein production in 3-L, 65-L and 300-L bioreactors by engineered strain co-expressing CelOCE. Comparison of the extracellular protein production of the parental strain Br_TrR03 and the engineered CelOCE::Br_TrR03 strain during bioreactor cultivations. The data from 3-L and 65-L bioreactors consist of three independent experiments (n = 3) and are expressed as mean ± standard deviation. The data from 300-L bioreactors derived from two independent experiments (n = 2).



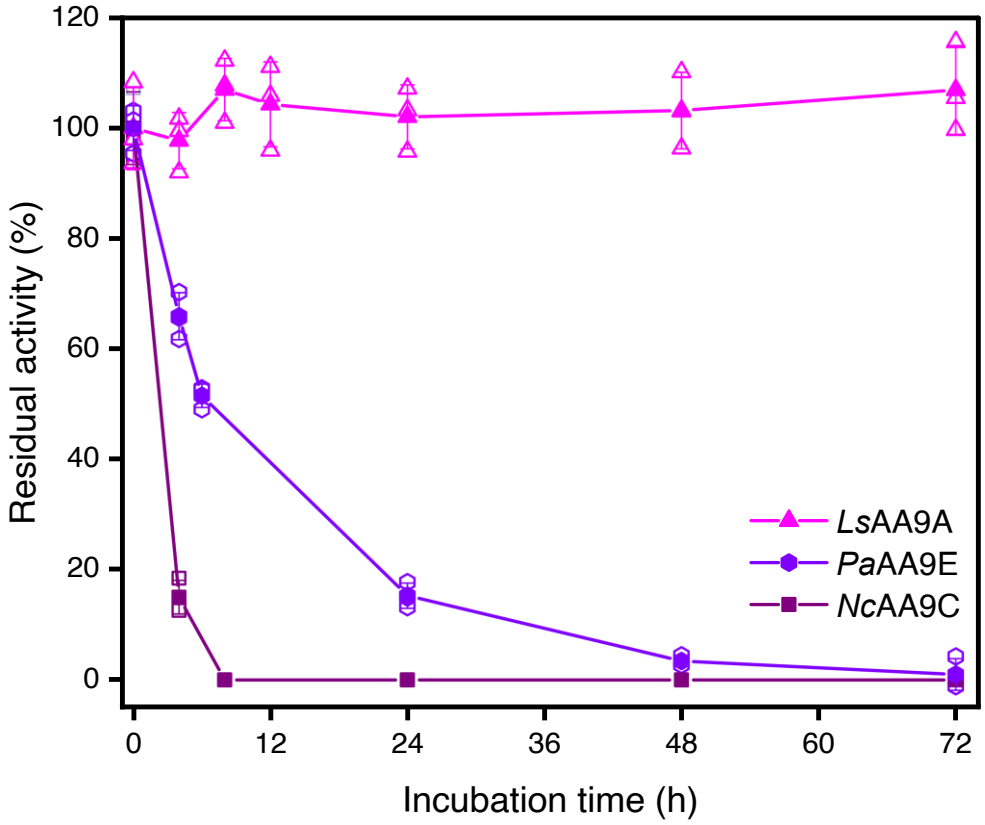
Supplementary Fig. 24. FPase and CMCase activities of enzyme cocktails produced by the engineered and parental strains. Specific activity levels of filter paper activity (FPase) and carboxymethyl cellulase (CMCase) are presented. The data are presented as mean ± standard deviation, derived from three independent experiments (n = 3). Tukey post hoc tests were conducted to identify statistically significant differences between the two strains (***$p < 0.001$).



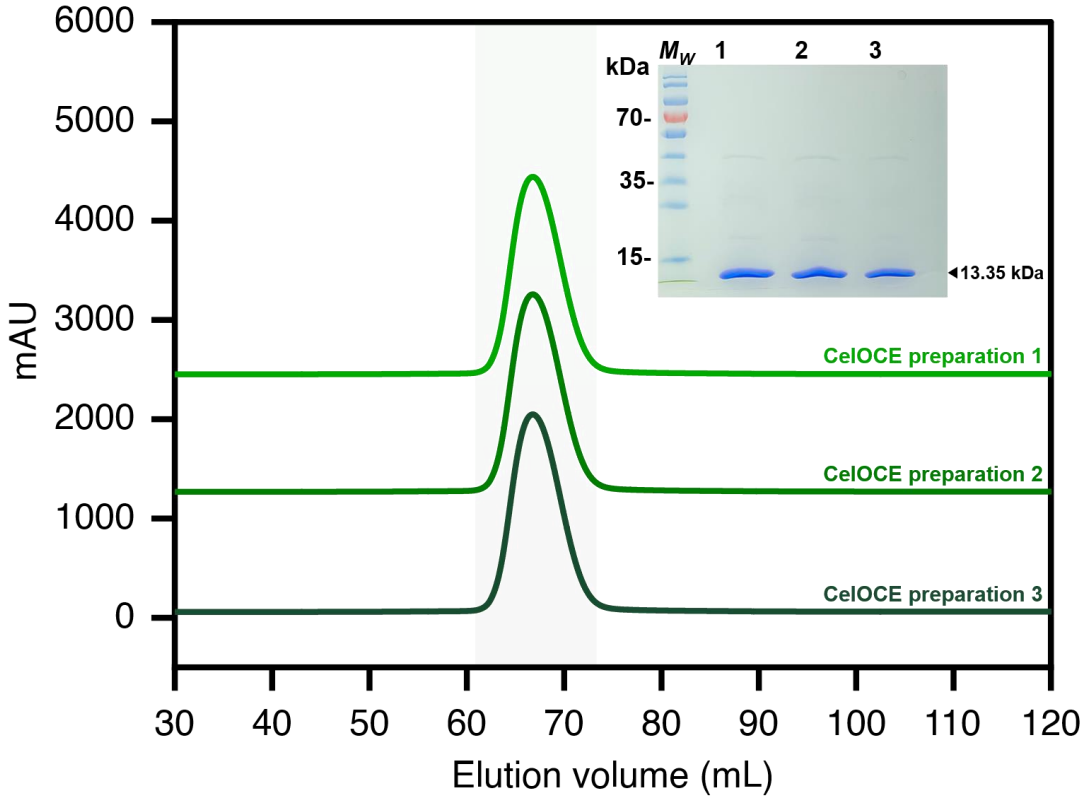
Supplementary Fig. 25. Extracellular protein production in 3-L bench bioreactor by engineered strain co-expressing *Ls*AA9A. Comparison of the extracellular protein production of the parental strain Br_TrR03 and the engineered *Ls*AA9A::Br_TrR03 strain during bioreactor cultivations. The data, expressed as mean ± standard deviation, are derived from three independent experiments (n = 3).



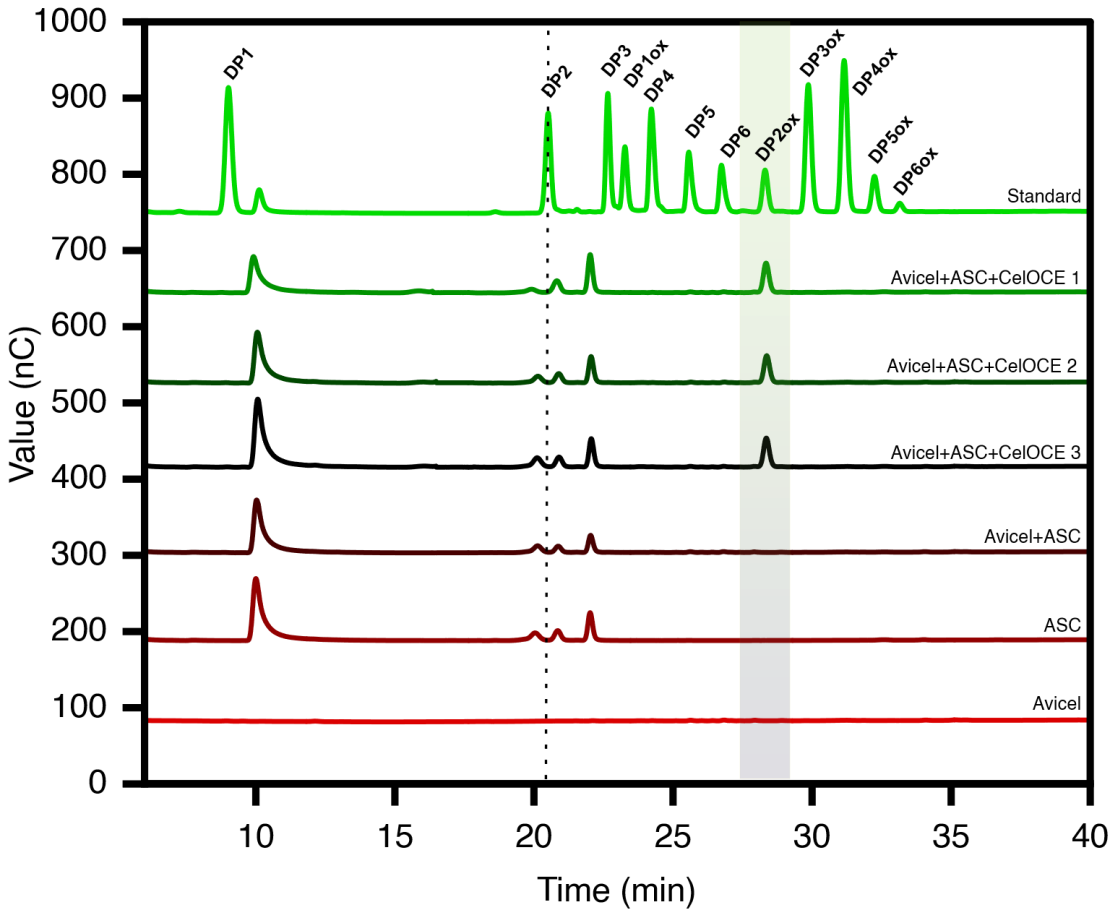
Supplementary Fig. 26. FPase and CMCase activities of enzyme cocktails produced by the engineered *Ls*AA9A::Br_TrR03 and parental strains. Specific activity levels of filter paper activity (FPase) and carboxymethyl cellulase (CMCase) are presented. The data are presented as mean ± standard deviation, derived from three independent experiments (n = 3). Tukey post hoc tests were conducted to identify statistically significant differences between the two strains (***$p < 0.001$), with asterisks marking enzymes exhibiting such differences.



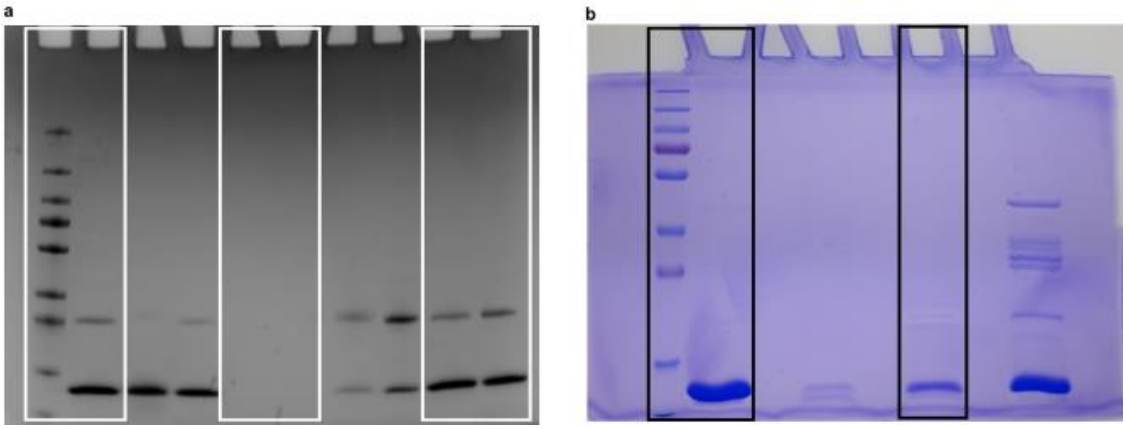
Supplementary Fig. 27. **LPMOs thermostability analysis**. Thermostability for LPMOs from *Lentinus similis* (*Ls*AA9A), *Neurospora crassa* (*Nc*AA9C) and *Podospora anserina* (*Pa*AA9E) using the 2,6-dimethoxyphenol incubated for 0 to 72 h in 50 mM sodium citrate buffer pH 5.0 at 50 °C. The data are presented as mean ± standard deviation, derived from three independent experiments (n = 3).



Supplementary Fig. 28. Quality control of CelOCE enzyme preparations. Size-exclusion chromatograms and SDS-PAGE analyses of three independent CelOCE enzyme preparations obtained from new transformants and purified using previously unused columns. These data confirm the purity and integrity of the enzyme preparations used in subsequent experiments. Note that the SDS-PAGE inset is uncropped.



Supplementary Fig. 29. HPAEC-PAD chromatograms of three independent CelOCE enzyme preparations. The consistent production of cellobionic acid as the sole product across all three preparations highlights the reproducibility and specificity of CelOCE catalytic activity.



Supplementary Fig. 30. Uncropped images of presented SDS data gels. (**a**) Uncropped gel image for Supplementary Fig. 7. (**b**) Uncropped gel image for Supplementary Fig. 18c.