Supplementary Information

Phenolic compound-mediated covalent crosslinking as a universal mechanism for cellulase deactivation during lignocellulose saccharification

Ya-Jun Liu (刘亚君)[a,b,c,d,#,*], Chao Chen (陈超)[a,b,c,d,#], Shenglei Wu (吴升磊)[e,#], Guang-Lei Liu (刘光磊)[e,*], Yingang Feng (冯银刚)[a,b,c,d,*]

[a] CAS Key Laboratory of Biofuels, Shandong Provincial Key Laboratory of Synthetic Biology, Shandong Engineering Laboratory of Single Cell Oil, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao 266101, China
[b] Shandong Energy Institute, Qingdao, 266101, China
[c] Qingdao New Energy Shandong Laboratory, Qingdao, 266101, China
[d] University of Chinese Academy of Sciences, Beijing 100049, China
[e] MOE Key Laboratory of Evolution and Marine Biodiversity, College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China

[#] The authors contribute equally to this work.
* Corresponding authors: Ya-Jun Liu (liuyj@qibebt.ac.cn); Guanglei Liu (liugl@ouc.edu.cn); Yingang Feng (fengyg@qibebt.ac.cn)

Table S1. The strains and plasmids used in this study

Strains/Plasmids	Relevant characteristic	Sources
Strains		
E. coli		
DH5α	*F- 80d lacZΔM15, Δ(lacZYA-argF) U169, deoR, recA1, endA1, hsdR17(r_k^-, m_k^+), phoA, supE44, l-, thi-1, gyrA96, relA1*	Transgen
BL21(DE3)	*ompT gal dcm lon hsdSB(rB − mB −) l (DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*	Transgen
BL21(DE3)::pET28aNS-rCaBglA	BL21 (DE3) derivative to express recombinant CaBglA with N-terminal His tag	[1]
BL21(DE3)::pET30a-rCaBglA	BL21 (DE3) derivative to express recombinant CaBglA with C-terminal His tag	This work
BL21(DE3)::pET28aNS-rCglT	BL21 (DE3) derivative to express recombinant CglT	[1]
BL21(DE3)::pET28a-rTnBGL	BL21 (DE3) derivative to express recombinant TnBGL (accession number ADA66698.1)	[2]
BL21(DE3)::pET28a-SMT3-rTtBGL	BL21 (DE3) derivative to express recombinant TtBGL (accession number YP_004660190.1)	[2]
BL21(DE3)::pET30a-rCel9D	Rosetta (DE3) derivative containing pET28a-SMT3-rCel9D	This work
BL21(DE3)::pET30a-rCel9K	Rosetta (DE3) derivative containing pET30a-rCel9K	This work
Rosetta(DE3)	*F- ompT $hsdS_B$ (r_B^-, m_B^-) gal dcm lacY1(DE3) pRARE (argU, argW, ileX, glyT, leuW, proL) (Cam^r)*	Tsingke Bio
Rosetta(DE3)::pET30a-rCel1A	Rosetta (DE3) derivative containing pET30a-rCel1A to express recombinant Cel1A	This work
Rosetta-gami (DE3)::pET20b-Cel5A	Rosetta-gami (DE3) derivative to express recombinant *Cel5A* (accession number DQ178347.1)	Gifted from Dr. Xiangfei Song
Rosetta-gami	Rosetta-gami (DE3) derivative to express	

(DE3)::pET20b-Cel7B	recombinant *Cel7B* (accession number AAA34212.1)	
C. thermocellum		
Δ*pyrF*	Derived from the wild-type strain DSM1313, with deleted *pyrF* gene	[1]
Δ*pyrF*::Cel48S_CD-His_{12}	Derived from Δ*pyrF*, expressing the catalytic domain of Cel48S bearing a His_{12} tag	[3]
Δ*pyrF*::Cel9K-CaBglA (WCB2)	Derived from Δ*pyrF*, expressing the genomic fusion protein Cel9K-CaBglA	[4]
Δ*pyrF*::pHK-P_{2638}-CaBglA (WCB3)	Derived from Δ*pyrF*, capable of plasmid-based expression of CaBglA	[5]
Plasmids		
pET30a	Expression vector with C-terminal His affinity tag	Transgen
pET30a-CaBglA	pET30a derivate for CaBglA expression (residues 4-455)	This work
pET30a-Cel9D	pET30a-SMT3 derivative for Cel9D expression (residues 45-578)	This work
pET30a-Cel9K	pET30a derivative for Cel9K expression (residues 208-824)	This work
pET30a-Cel1A	pET30a derivative for Cel1A expression (residues 2-466)	This work

Table S2. Determination of phenolic compounds present in water-soluble extractives from rice straw and wheat straw by GC-MS.

	Rice straw (μg/g)		
	RN	R60	R160
Vanillin	71.91±7.97	49.92±0.07	33.00±1.93
p-hydroxy benzaldehyde	195.71±23.12	70.82±0.86	nd
Coniferyl aldehyde	56.95±7.23	19.78±1.62	nd
Ferulic acid	56.18±8.72	86.98±0.68	8.19±0.42
Vanillic acid	107.98±14.78	49.19±2.34	28.71±1.34
Syringic acid	89.26±16.81	38.61±1.12	12.59±0.50
p-coumaric acid	494.29±23.27	338.33±14.52	17.60±0.27
p-hydroxybenzoic acid	80.94±8.14	nd	nd
	Wheat straw (μg/g)		
	WN	W60	W160
Vanillin	53.86±8.07	46.03±0.79	42.15±6.82
p-hydroxy benzaldehyde	73.87±15.55	38.38±0.80	nd
Coniferyl aldehyde	73.32±15.62	23.34±1.30	nd
Ferulic acid	45.46±6.05	75.04±1.17	7.97±1.63
Vanillic acid	80.23±9.86	47.99±1.21	51.97±9.86
Syringic acid	74.96±13.17	41.29±1.31	43.94±10.71
p-coumaric acid	403.94±68.30	283.72±2.19	6.14±5.66
p-hydroxybenzoic acid	31.63±8.88	nd	nd

nd. Not detected

Data are presented as mean values ± SD from three independent experiments. Source data are provided as a Source Data file.

Table S3. Determination of phenolic compounds present in acid-soluble lignin by GC-MS.

	Acid-soluble lignin (μg/g)
Vanillin	259.08±65.18
Coniferyl aldehyde	334.01±14.55
Vanillic acid	877.15±98.69
Syringic acid	2252.37±478.65
p-hydroxybenzoic acid	134.99±9.77

Data are presented as mean values ± SD from three independent experiments. Source data are provided as a Source Data file.

Table S4. Primers used in this study

Primers	Sequences (5'-3')*	Notes
CaBglA-F	AAGAAGGAGATATACATATGTCATTCC CGAAAGGATTTTTG	Cloning gene *CaBglA* to construct pET30a-rCaBglA
CaBglA-R	GGTGGTGGTGGTGCTCGAGTGAATTT TCCTTTATATACTG	
Cel1A-F	AAGAAGGAGATATACATATGTTGCCCA AGGACTTTCAGTG	Cloning gene *Cel1A* to construct pET30a-rCel1A
Cel1A-R	GGTGGTGGTGGTGCTCGAGCGCCGCC GCAATCAGCTCGTC	
Cel9K-F	AAGAAGGAGATATACATATGGTACTTC CGCAGCCGGATGT	Cloning gene *Cel9K* to construct pET30a-rCel9K
Cel9K-R	TGGTGGTGGTGGTGCTCGAG GTCTACTCCTCCTGGCGGT	
Cel9D-F	AAGAAGGAGATATACATATGACGGAG AATTATCAATTTGA	Cloning gene *Cel9D* to construct pET30a-rCel9D
Cel9D-R	TGGTGGTGGTGGTGCTCGAGTTGAGC AGAATTATAGTTGAC	
T7	GCCCCAAGGGGTTATGCTAG	For colony PCR and sequencing
T7ter	TAATACGACTCACTATAGG	

* Sequences for homologous recombination were underlined.

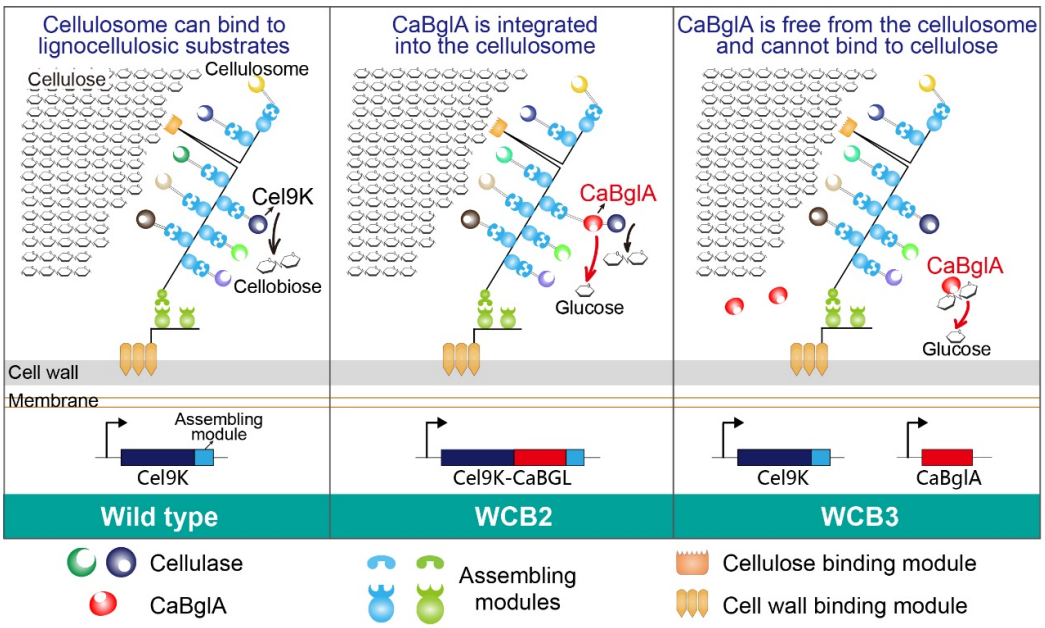

Figure S1. Schematic representation of WCB2 and WCB3 in terms of CaBglA production

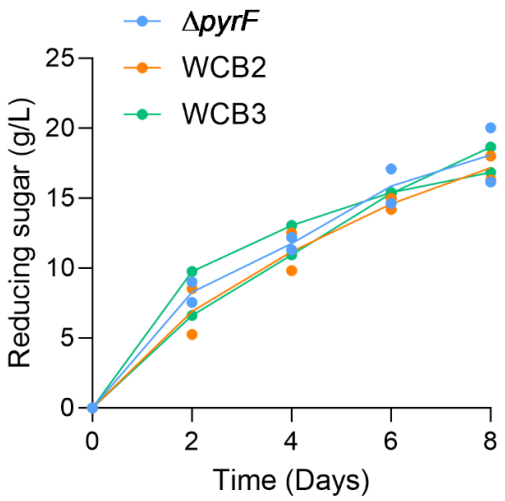

Figure S2. SP saccharification with Δ*pyrF*, WCB2, and WCB3 as the biocatalysts.

4% (w/v) of SP was supplemented as the substrate. The produced reducing sugar was determined by the DNS method. The lines indicate the mean values of two independent replicates of each strain. Source data are provided as a Source Data file.

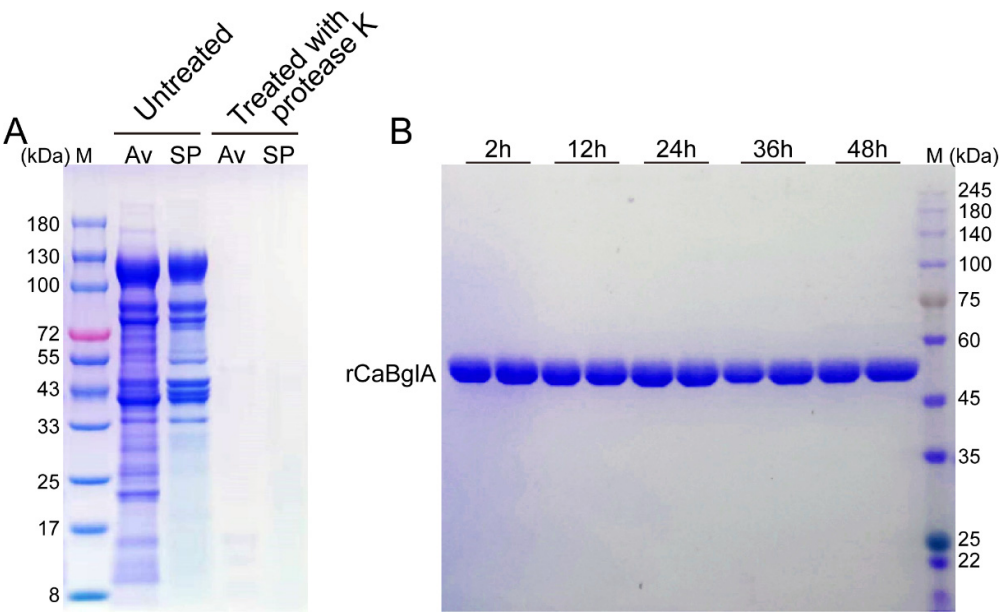

Figure S3. Evaluation of protease K treatment.

A, SDS-PAGE analysis of Avicel (Av) or SP hydrolysate supernatants before and after protease K treatment, suggesting effective digestion of intergenic proteins originated from *C. thermocellum*. **B**, SDS-PAGE analysis of rCaBglA protein incubated with inactivated protease K for 2 to 48 hours, indicating the stability of CaBglA under the experimental conditions. Results are representative of three independent replicates in A and B. Source data are provided as a Source Data file.

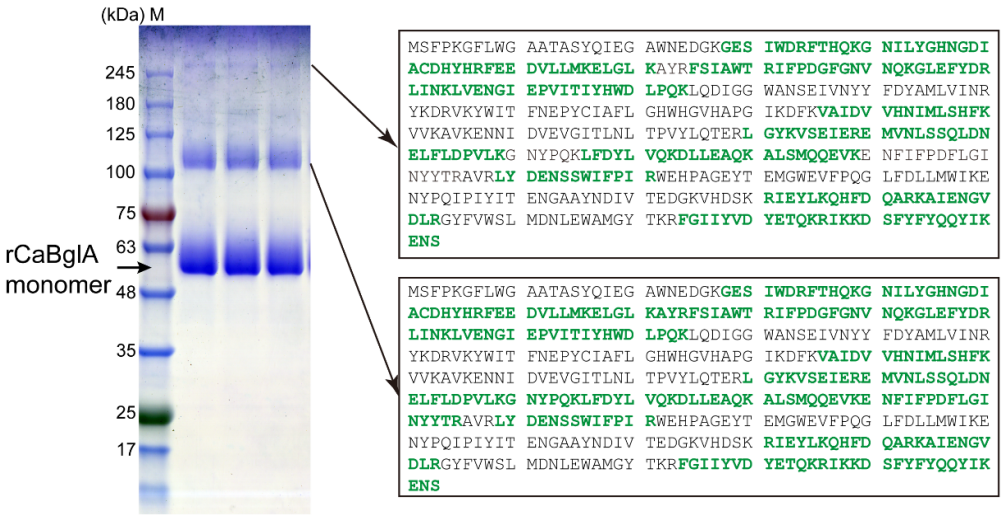

Figure S4. Mass spectrometry identification of CaBglA protein.

rCaBglA protein was incubated with SP supernatant for 3 days. The peptides detected by mass spectroscopy are in green. The bands of over 100 kDa and 245 kDa were selected for analysis, both of which were determined to be CaBglA, indicating the covalent crosslinking. SDS-PAGE results are representative of two independent experiments. Source data are provided as a Source Data file.

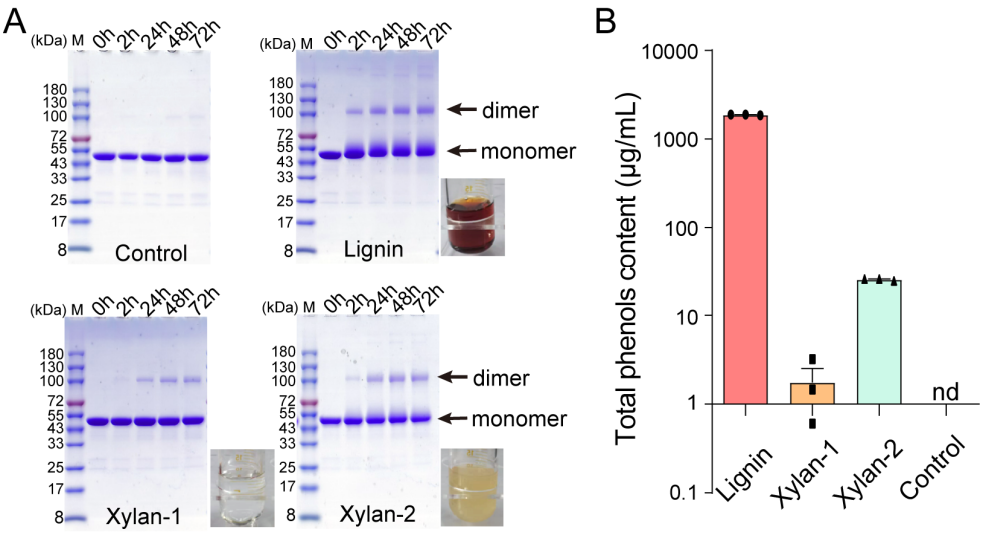

Figure S5. SDS-PAGE analysis of rCaBglA incubated with acid-soluble lignin and beechwood xylan solutions (A) and TPC analysis of the solutions (B).

A, The images of lignin and xylan solutions (16.7 g/L) used are shown on the right of the corresponding panels. Xylan was purchased from Sigma-Aldrich Co. Ltd (Xylan-1) or Megazyme Ltd (Xylan 2). The incubations lasted for 2 to 72 hours. Results are representative of two independent experiments. B, the control experiment was conducted using the supernatant of the solution of insoluble Avicel. Three independent experiments were conducted for calculation, and the data are presented as mean values ± SD. Source data are provided as a Source Data file.

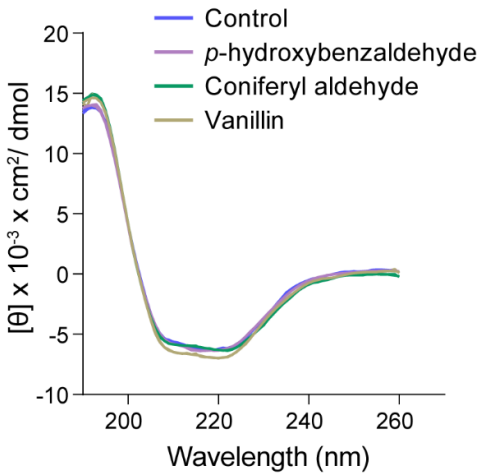

Figure S6. CD spectra of rCaBglA protein incubated with selected phenolic aldehydes.

Data are presented as the mean of three technical replicates. Source data are provided as a Source Data file.

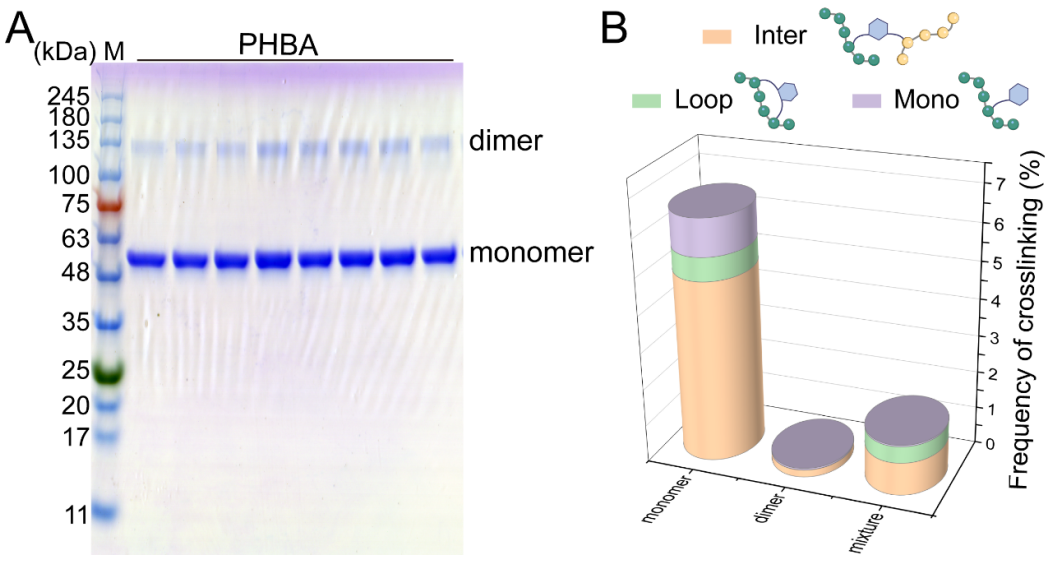

Figure S7. Mass spectrometry analysis of the CaBglA protein after incubation with *p*-hydroxybenzaldehyde.

A, SDS-PAGE analysis of rCaBglA incubated with *p*-hydroxybenzaldehyde (PHBA). Results are representative of two independent experiments. B, Frequency of identified CaBglA crosslinking modifications at lysine and cysteine residues. LC-MS/MS analysis was performed on three samples: the monomer and dimer bands excised from the gel in (A), along with the unfractionated mixture of PHBA-incubated rCaBglA ("mixture"). Three types of protein crosslinking are proposed, and the predicted molecule structures are shown in the figure. Inter, inter-crosslinking between two peptides. Loop, intra-crosslinking within a single peptide. Mono, modification of an amino acid residue within the peptide. The identified crosslinking was mediated by one or two molecules of *p*-hydroxybenzaldehyde. The LC-MS/MS analysis was performed once. Source data are provided as a Source Data file.

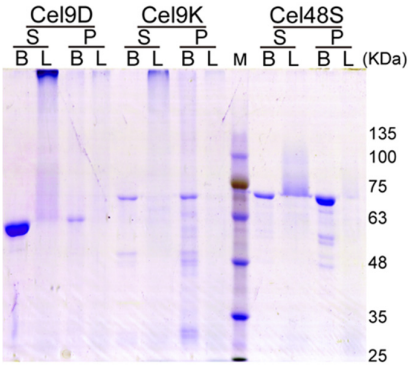

Figure S8. SDS-PAGE analysis of Cel9K, Cel9D, and Cel48S.

The cellulases were first mixed with Avicel to allow substrate binding. The mixtures were subsequently added with acid-soluble solution (L) or GS-2 buffer (B) for 3 days and were concentrated. Both supernatants (S) and pellets (P), which contain unbounded and bounded cellulases, respectively, were analyzed. Results are representative of two independent experiments. Source data are provided as a Source Data file.

Figure S3A

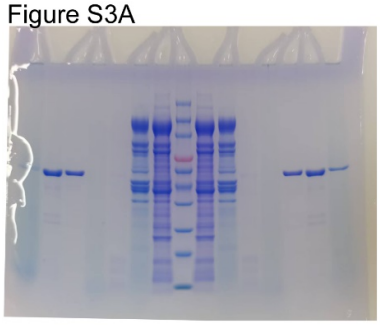

Figure S3B

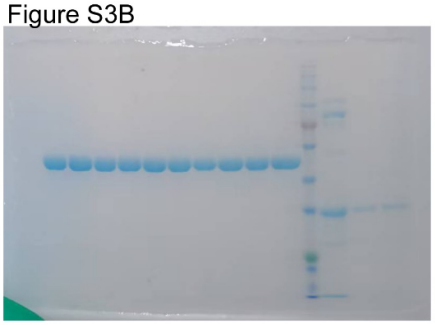

Figure S4

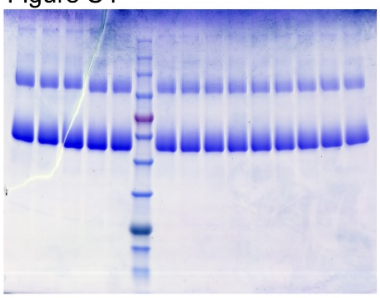

Figure S5A Control

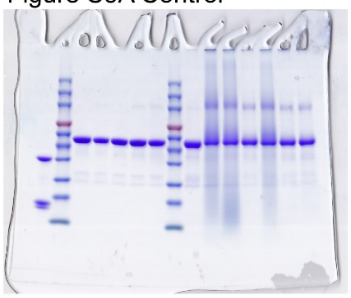

Figure S5A Lignin and Xylan-1

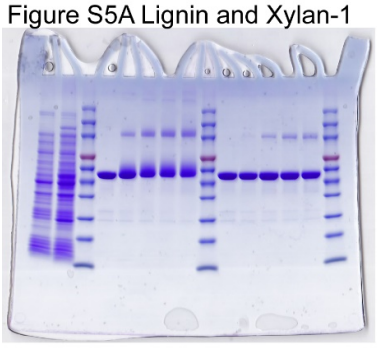

Figure S5A Xylan-2

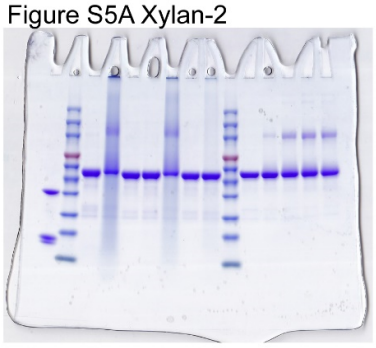

Figure S7A

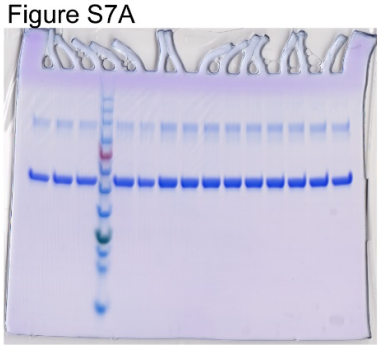

Figure S8

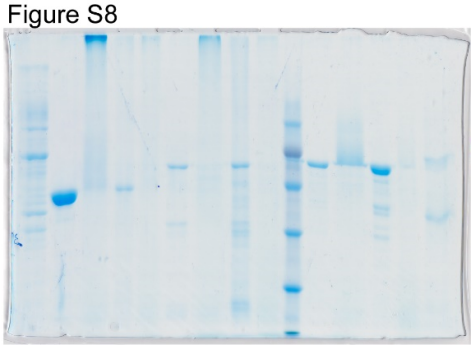

Figure S9. Original pictures of gels used in the Supplementary Information.

Supplementary References

1. Zhang, J. et al. Efficient whole-cell-catalyzing cellulose saccharification using engineered *Clostridium thermocellum*. *Biotechnol. Biofuels* **10**, 124 (2017).
2. Zheng, X. Master Degree Thesis, University of Chinese Academy of Sciences (2021).
3. Liu, Y.-J. et al. Determination of the native features of the exoglucanase Cel48S from *Clostridium thermocellum*. *Biotechnol. Biofuels* **11**, 6 (2018).
4. Liu, S., Liu, Y.-J., Feng, Y., Li, B. & Cui, Q. Construction of consolidated bio-saccharification biocatalyst and process optimization for highly efficient lignocellulose solubilization. *Biotechnol. Biofuels* **12**, 35 (2019).
5. Qi, K. et al. Coordinated β-glucosidase activity with the cellulosome is effective for enhanced lignocellulose saccharification. *Bioresour. Technol.* **337**, 125441 (2021).