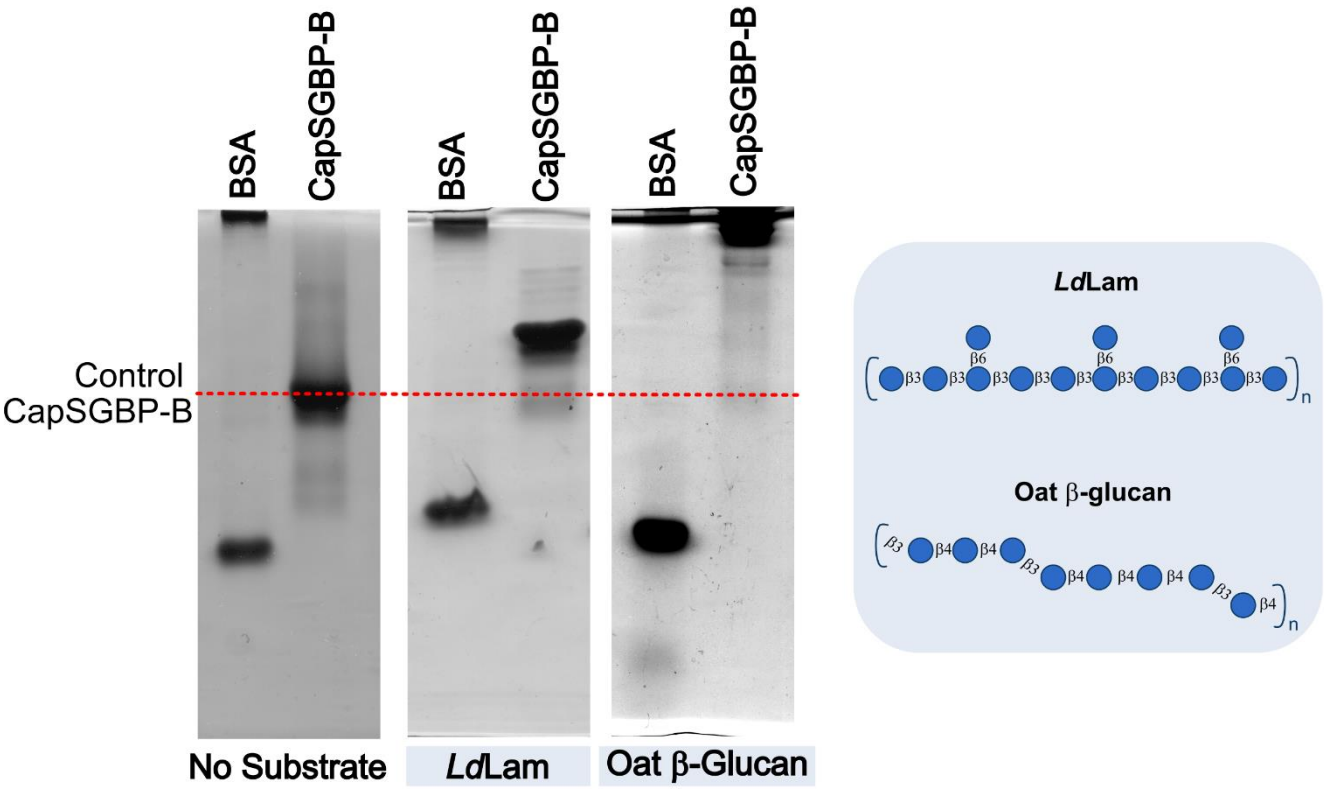
SUPPLEMENTARY INFORMATION

A functionally augmented carbohydrate utilization locus from herbivore gut microbiota fueled by dietary β-glucans

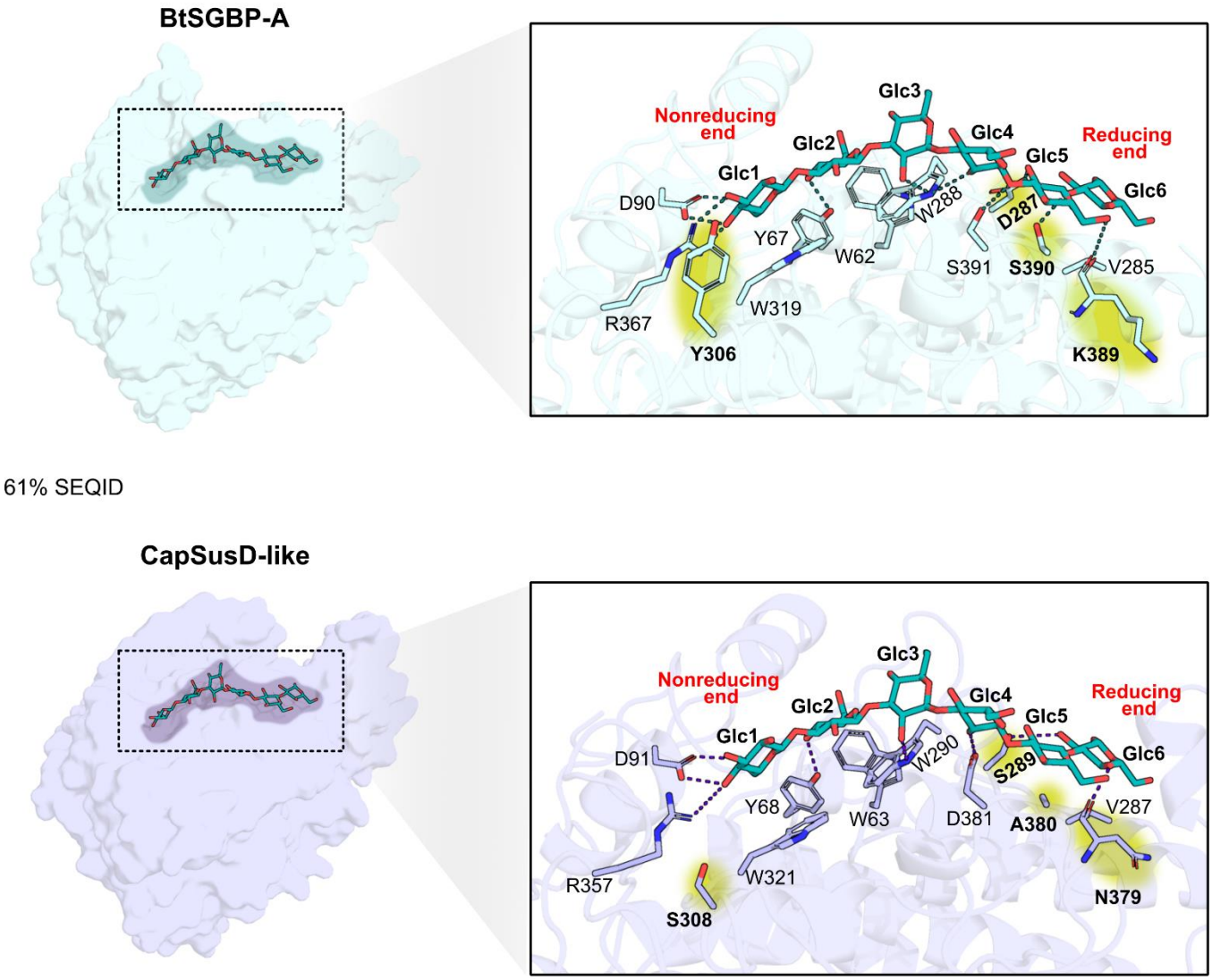
Fernanda Mandelli[1,#], Marcele Pandeló Martins[1,#], Mariana Chinaglia[1,#], Evandro Antonio de Lima[1], Mariana Abrahão Bueno Morais[1], Tatiani Brenelli Lima[1], Lucélia Cabral[1], Renan Augusto Siqueira Pirolla[1], Felipe Jun Fuzita[1,] Douglas Antônio Alvaredo Paixão[1], Maxuel de Oliveira Andrade[1], Lucia Daniela Wolf[1], Plinio Salmazo Vieira[1], Gabriela Felix Persinoti[1], Mario Tyago Murakami[1]*

[1]Brazilian Biorenewables National Laboratory (LNBR), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, São Paulo 13083-970, Brazil.

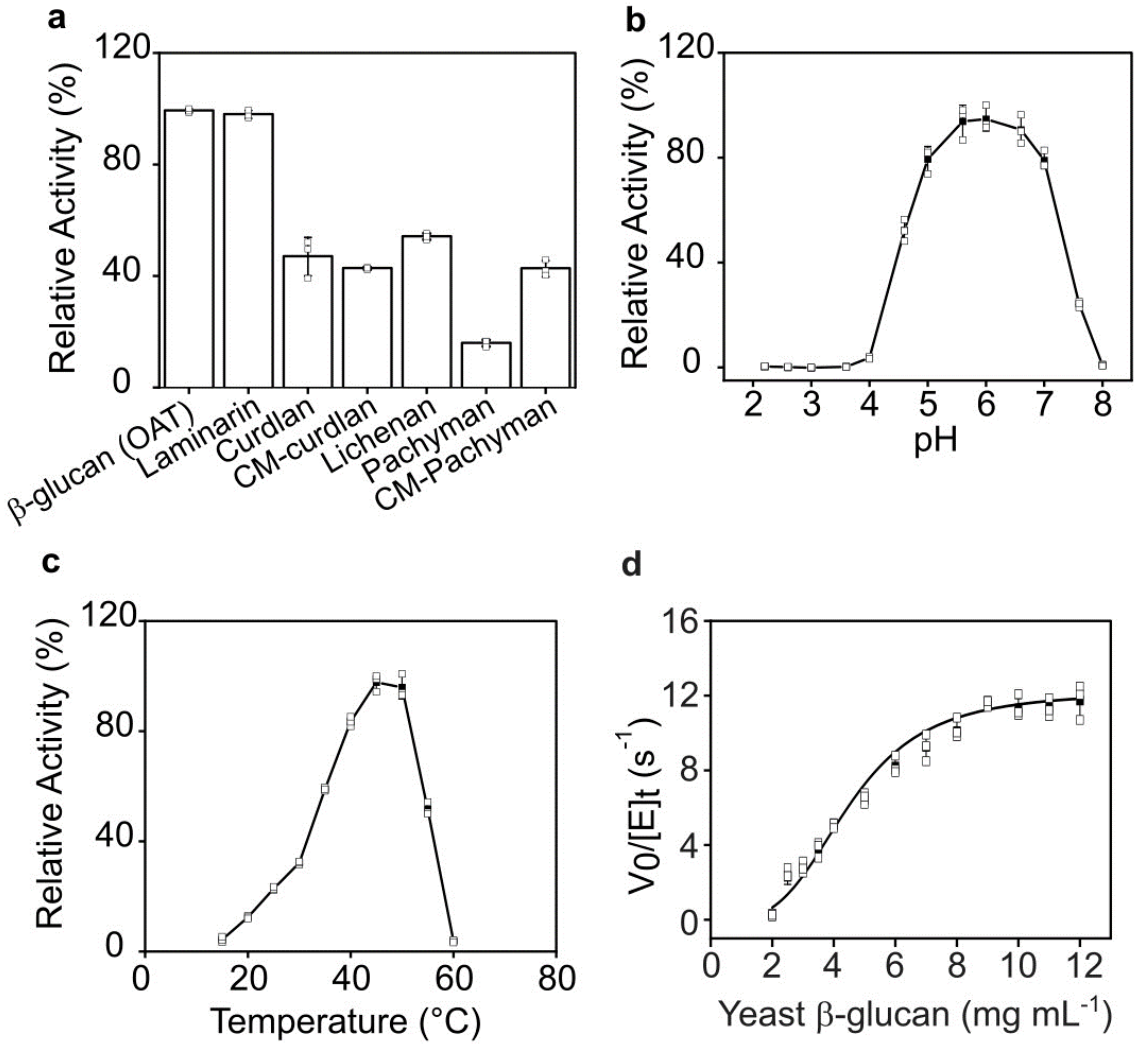
*Correspondence and requests should be addressed to M.T.M (mario.t.murakami@gmail.com)



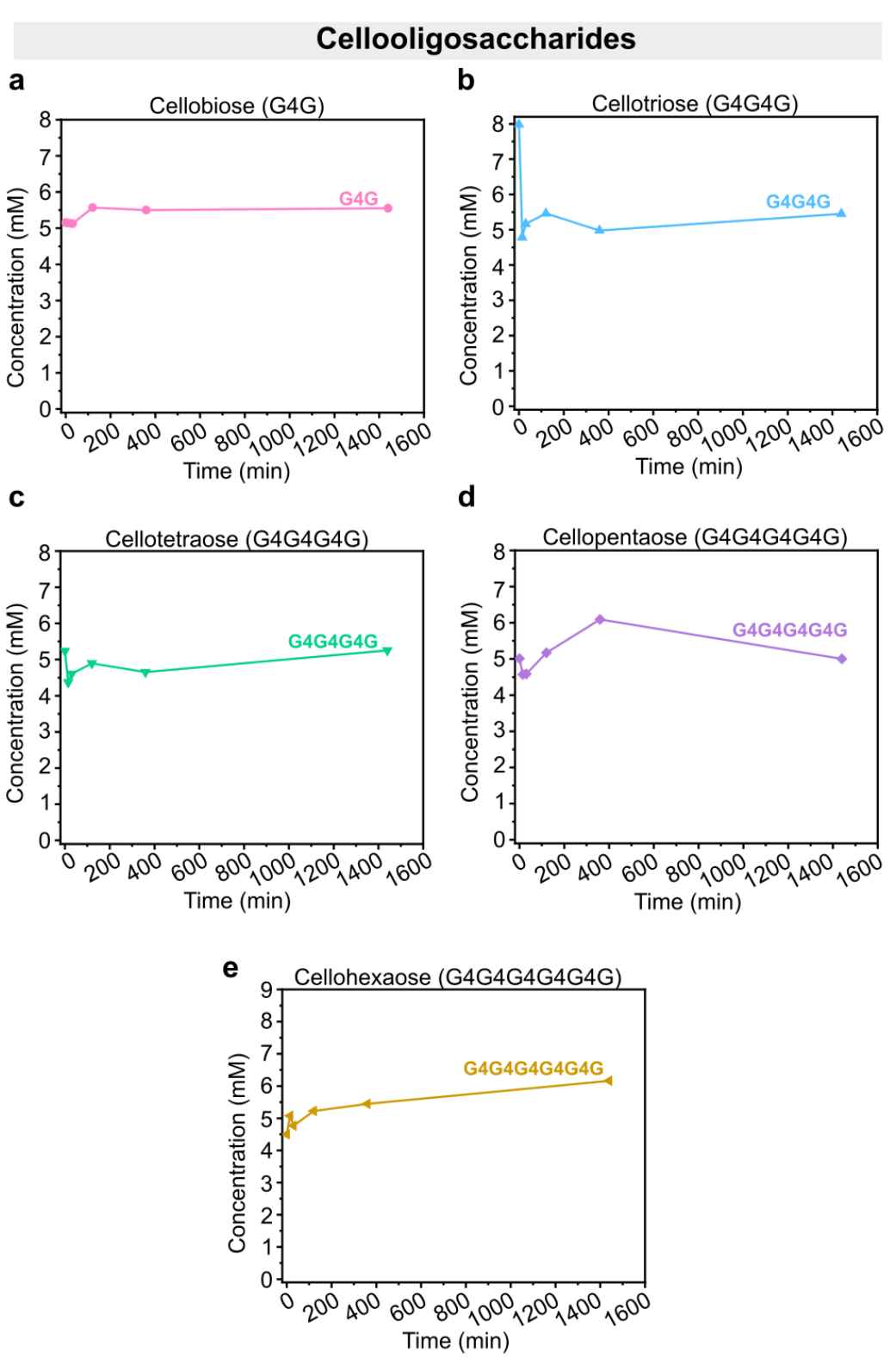
Supplementary Figure 1: CapSGBP-B binds to β(1,3)-glucans and mixed linked β-glucans. The protein binding was evaluated by affinity gel electrophoresis (AGE) against the Laminarin from *Laminaria digitata* (*Ld*Lam) and the mixed-linkage β-glucan from oat. Bovine serum albumin (BSA) was used as a control. In contrast to *Bo*SGBP$_{MLG}$-B, the lower migration of CapSGBP-B in AGE containing *Ld*Lam indicates its binding to β(1,3)-glucans[1].



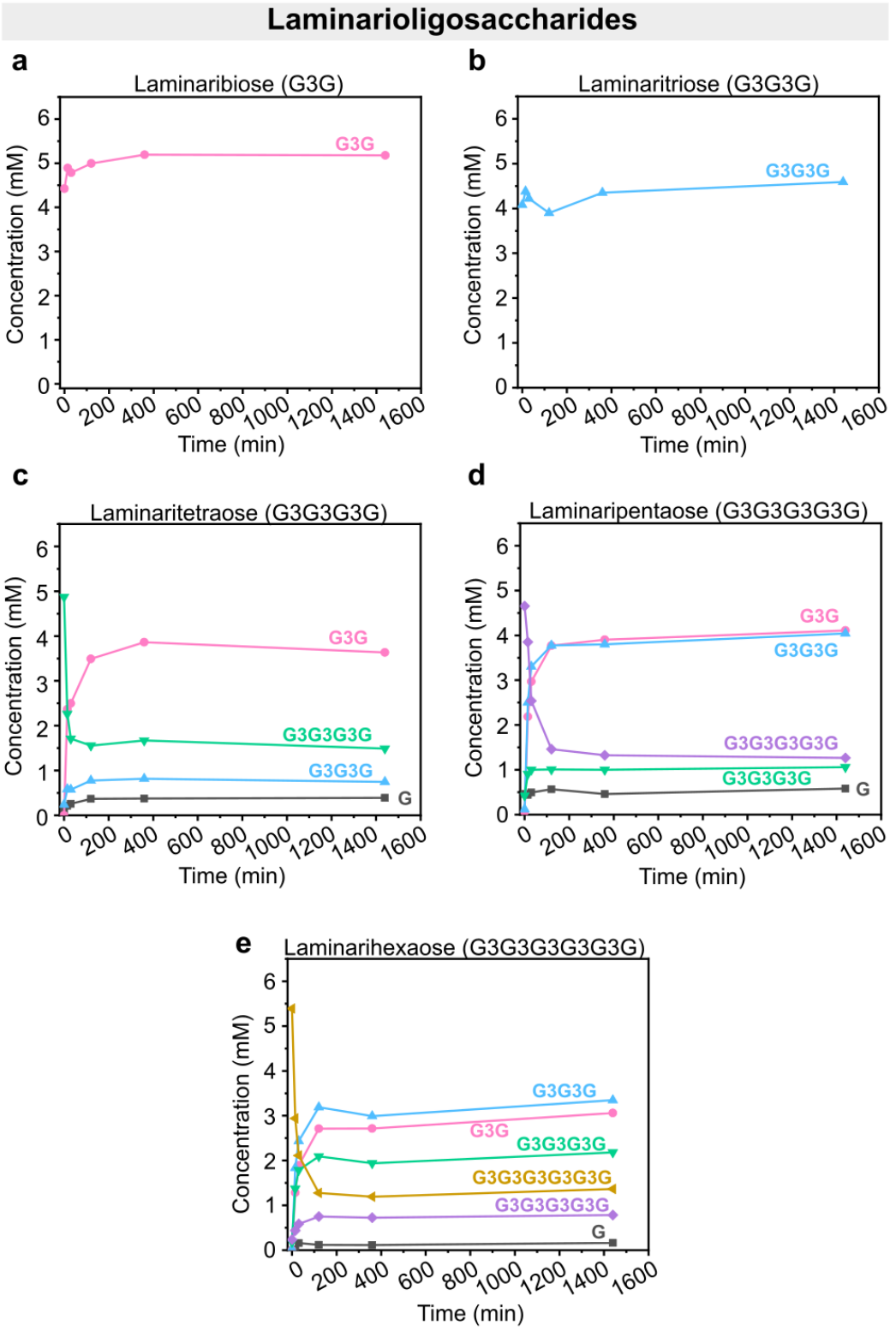
Supplementary Figure 2: Structural comparison between *Bt*SGBP-A (SusD-like, PDBID 7KV3[2]) and CapSusD-like 3D model[3] highlights the conservation of the key residues involved in the interaction with β(1,3)-glucans. Note that there are a few non-conserved residues at the extremities of the glycan binding interface (highlighted in yellow), which may affect the glycan binding affinity of CapSusD-like protein.



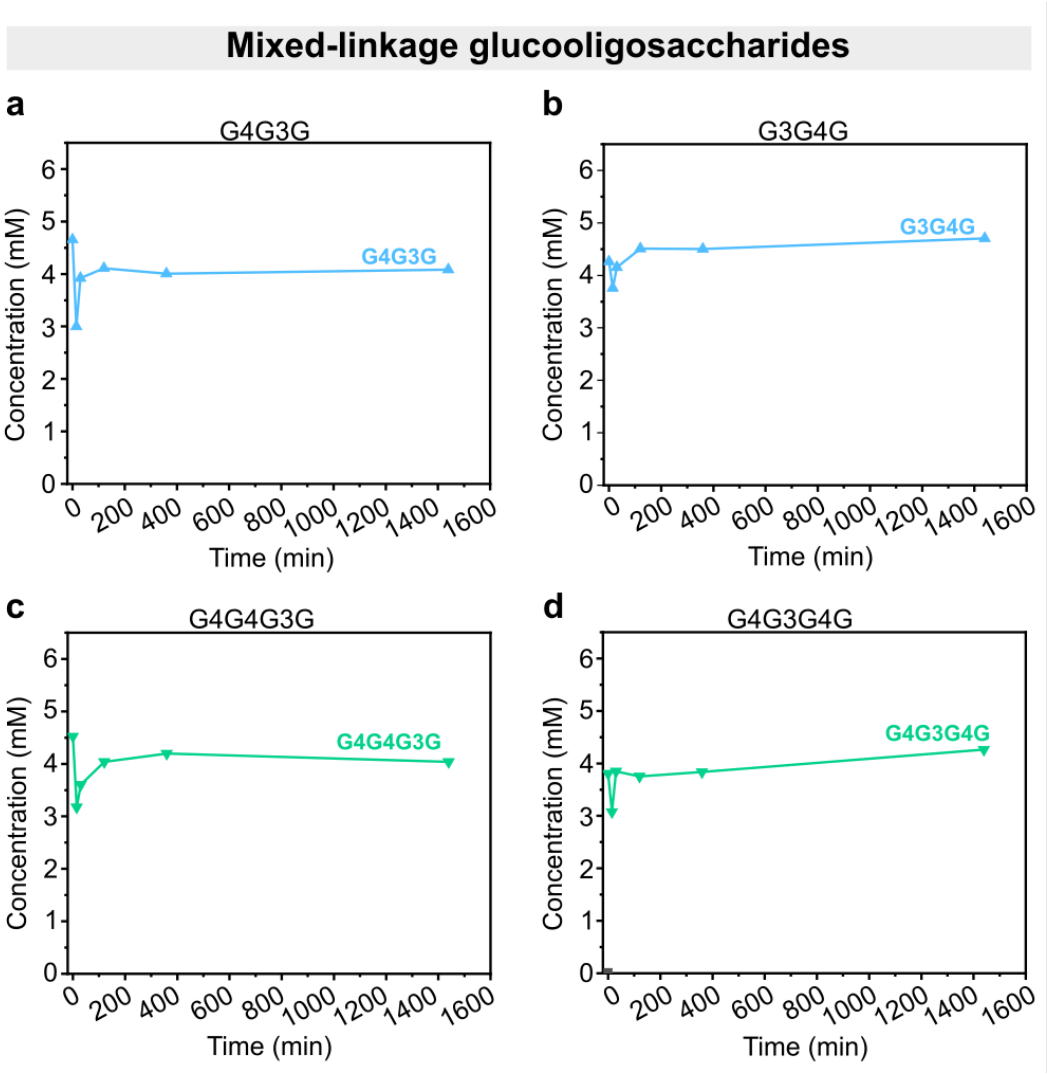
Supplementary Figure 3: Biochemical characterization of CapGH16_3. (a) Relative activity on polymeric substrates. **(b)** Optimum pH and **(c)** temperature. **(d)** Substrate saturation curve with yeast β-glucan assessed at 45°C and pH 6. Results are expressed as mean ± standard deviations from three independent experiments (n=3).



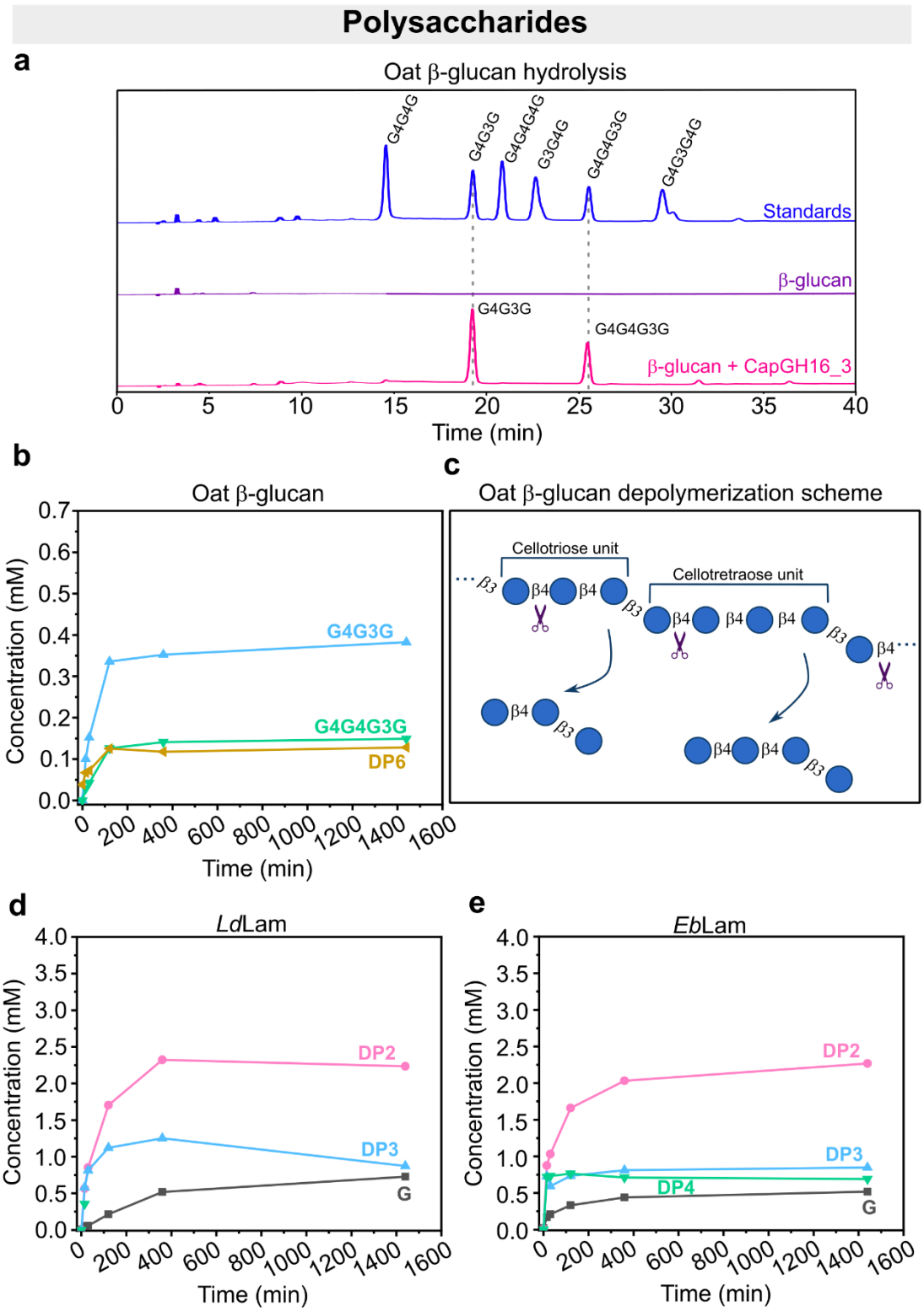
Supplementary Figure 4: Cleavage pattern of cellooligosaccharides by CapGH16_3. Note that this enzyme is not active on only β(1,4)-linked glucooligosaccharides.



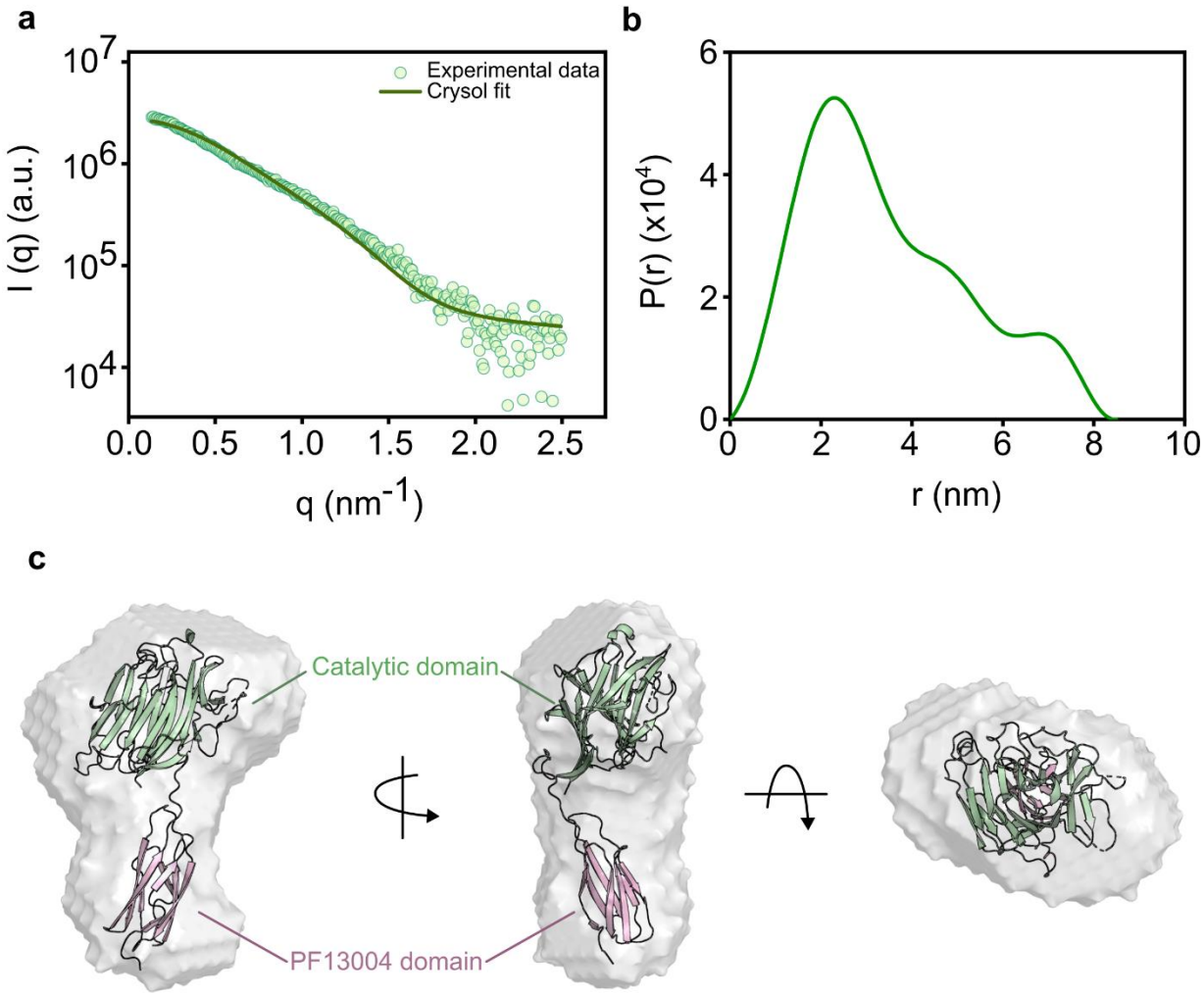
Supplementary Figure 5: Cleavage pattern of laminarioligosaccharides by CapGH16_3. These results indicate that at least four subsites are needed to be occupied for the productive binding of β(1,3)-glucooligosaccharides.



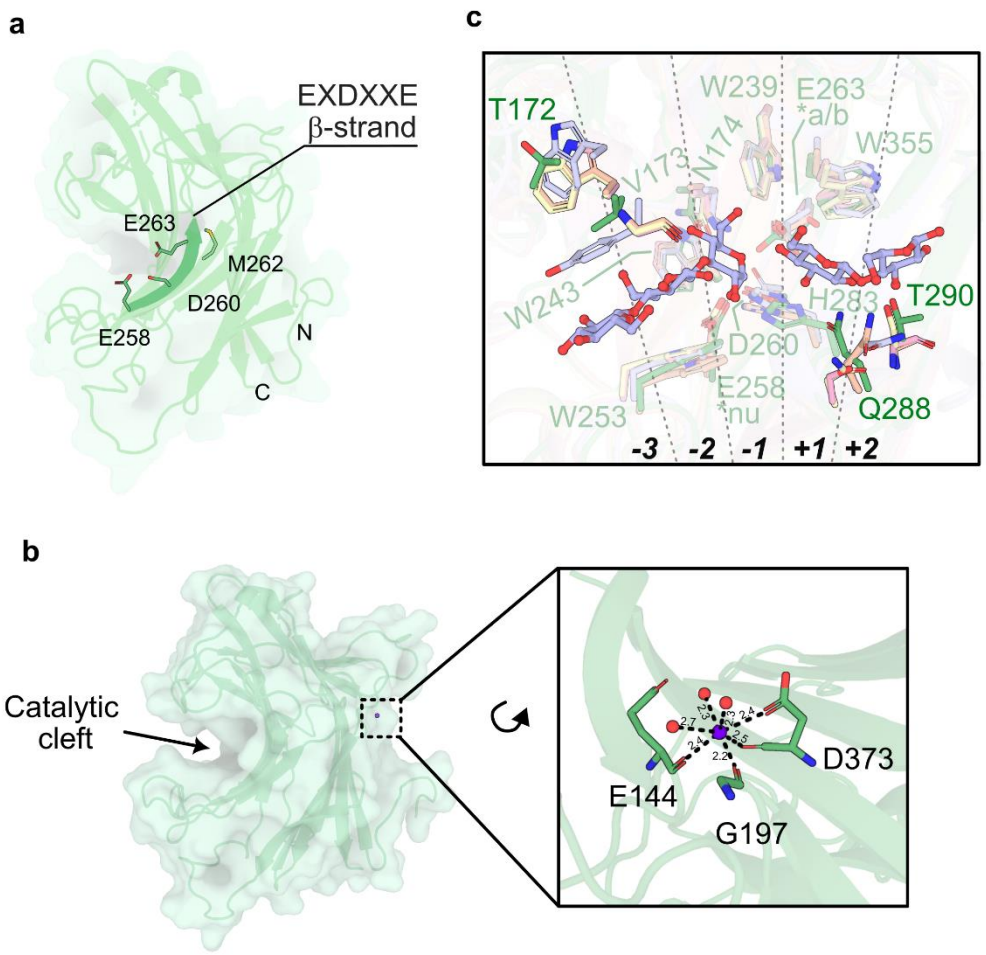
Supplementary Figure 6: Cleavage pattern of mixed-linkage glucooligosaccharides by CapGH16_3. These results highlight that the tested mixed-linkage glucooligosaccharides up to four glucosyl residues are not cleavable. It indicates that longer mixed linkage glucooligosaccharides might be required for productive binding.



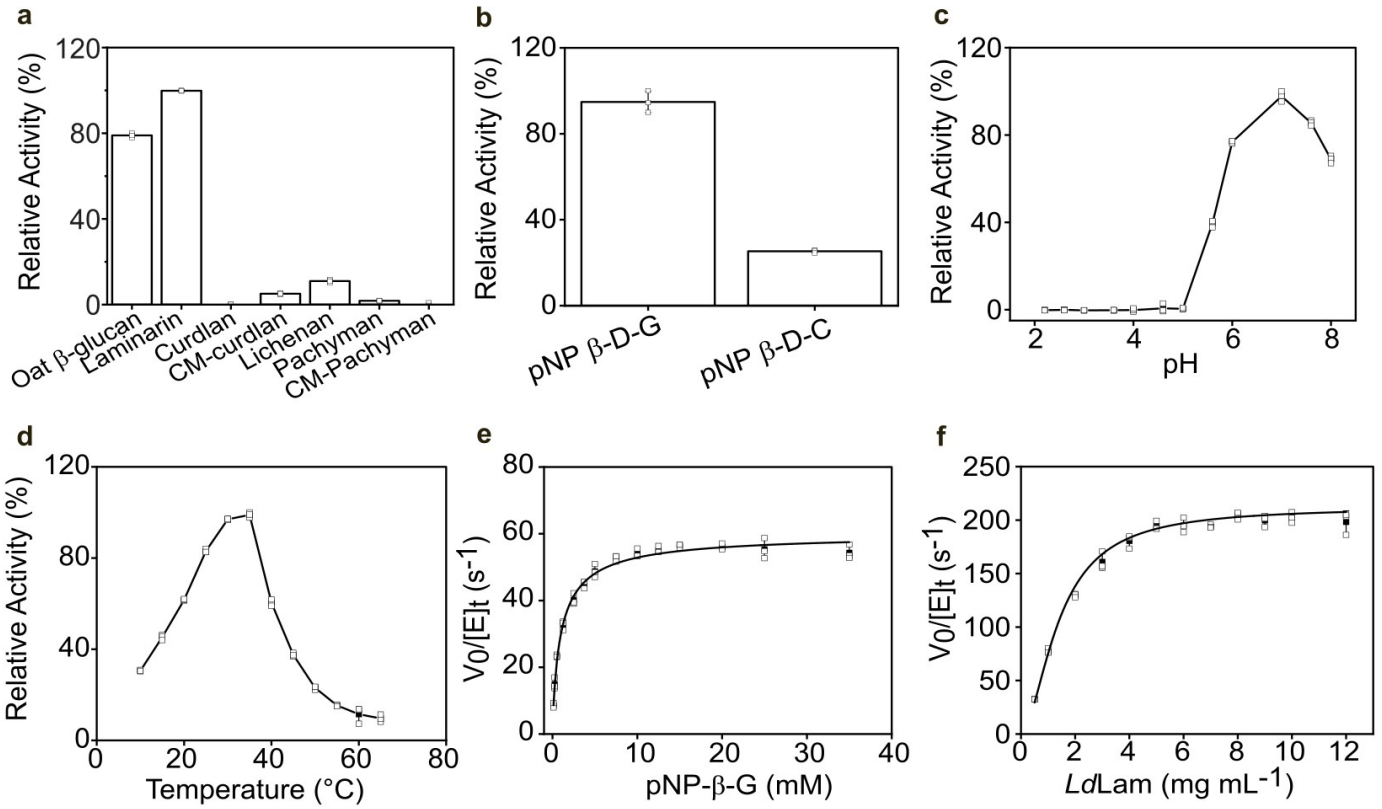
Supplementary Figure 7: Cleavage pattern of polysaccharides by CapGH16_3. (a) HPAEC-PAD analysis of limit-digest products of oat β-glucan by CapGH16_3. **(b)** The released products from oat β-glucan can have both β(1,3) and β(1,4) linkages. **(c)** Oat β-glucan cleavage pattern by CapGH16_3. **(d)** and **(e)** Products released from *Ld*Lam and *Eb*Lam, respectively, which can have both β(1,3) and β(1,6) linkages. This pattern indicates an endo mode of action of CapGH16_3.



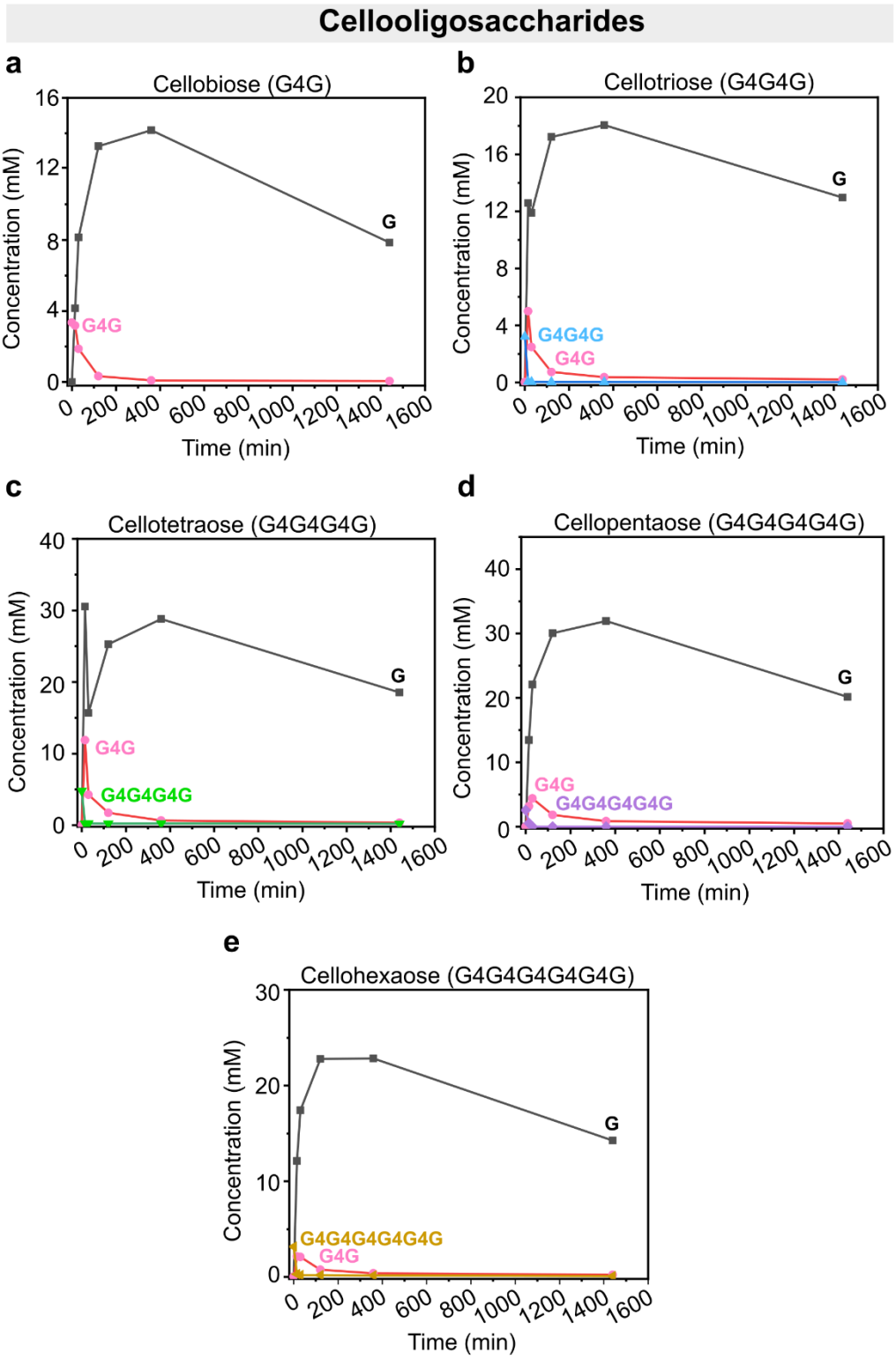
Supplementary Figure 8: SAXS analysis of CapGH16_3. (a) The green circles correspond to the experimental scattering curve obtained by SAXS, while the green line corresponds to the CRYSOL[4] theoretical scattering curve calculated using the crystallographic structure of CapGH16_3 domain along with the modeled PF13004 domain (AlphaFold2[3]). **(b)** The distance distribution curve P(r) was generated with GNOM[5], exhibiting a D_{max} of 8.5 nm. **(c)** The envelope calculated from SAXS data was superimposed onto the CapGH16_3 crystallographic structure (green cartoon) and the modeled PF13004 domain (pink cartoon) in three orientations.



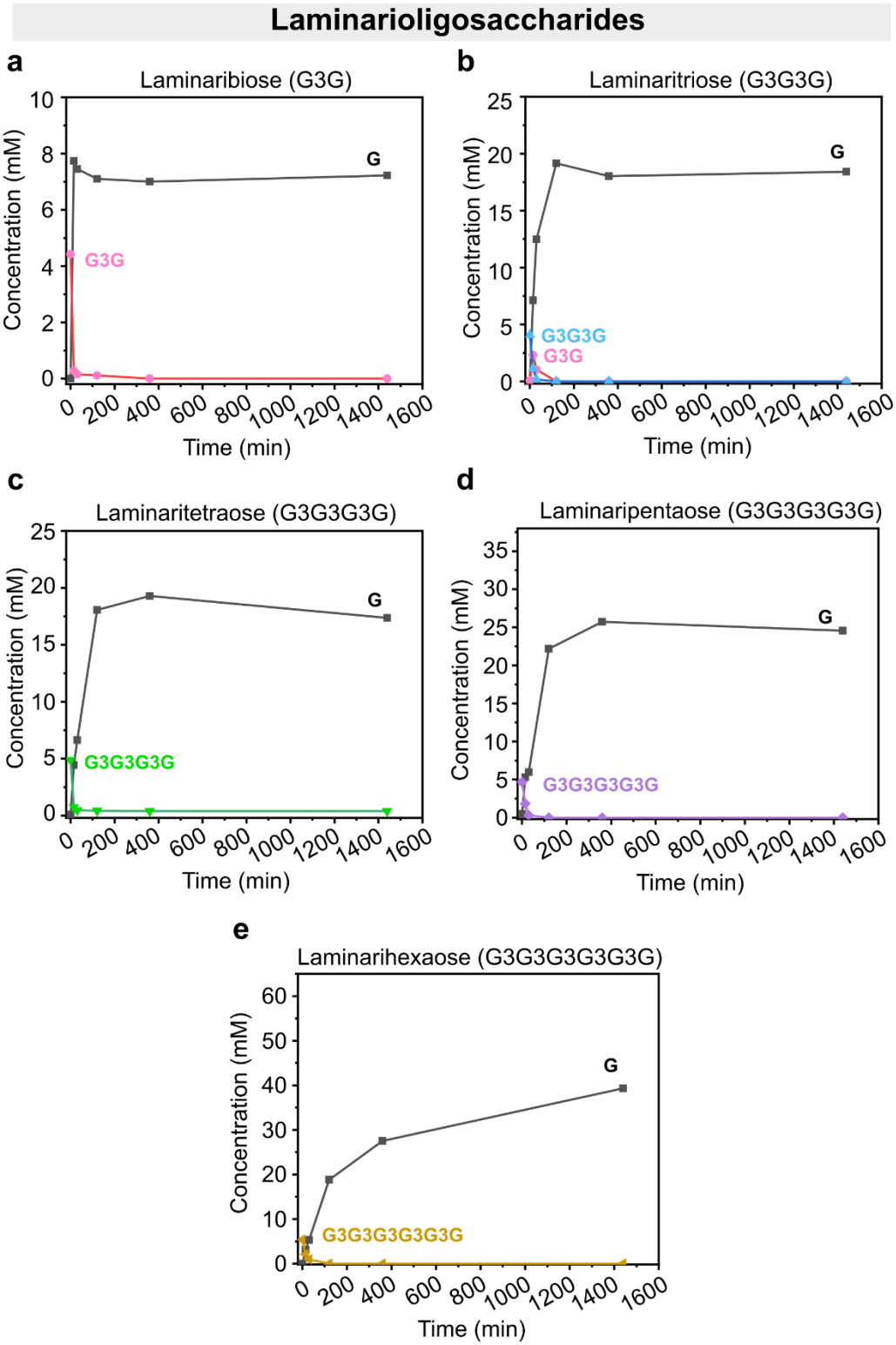
Supplementary Figure 9: Structural overview of CapGH16_3. (a) CapGH16_3 side view highlighting the conserved EXDXXE motif and the presence of the Met262 residue causing the β-strand distortion. **(b)** CapGH16_3 calcium-binding site at the opposite side of the active site. The distances between atoms are represented in angstroms. **(c)** Active site comparison between CapGH16_3 (green) and orthologs including SCLam (PDBID 6XQH[6], light blue), LamR (PDBID 3ILN[7], orange), *Zg*LamC$_{GH16}$ (PDBID 4CRQ[8], light pink), and LamAQ (PDBID 6JH5[9], yellow). The lack of conservation of the distant –3 and +2 subsites is highlighted. *Nu and *a/b represent the nucleophile and acid/base, respectively. The saccharides cellobiose and G4G3G in positive and negative subsites were extracted from the structural superposition with SCLam (PDBIDs 6XQH and 6XQF).



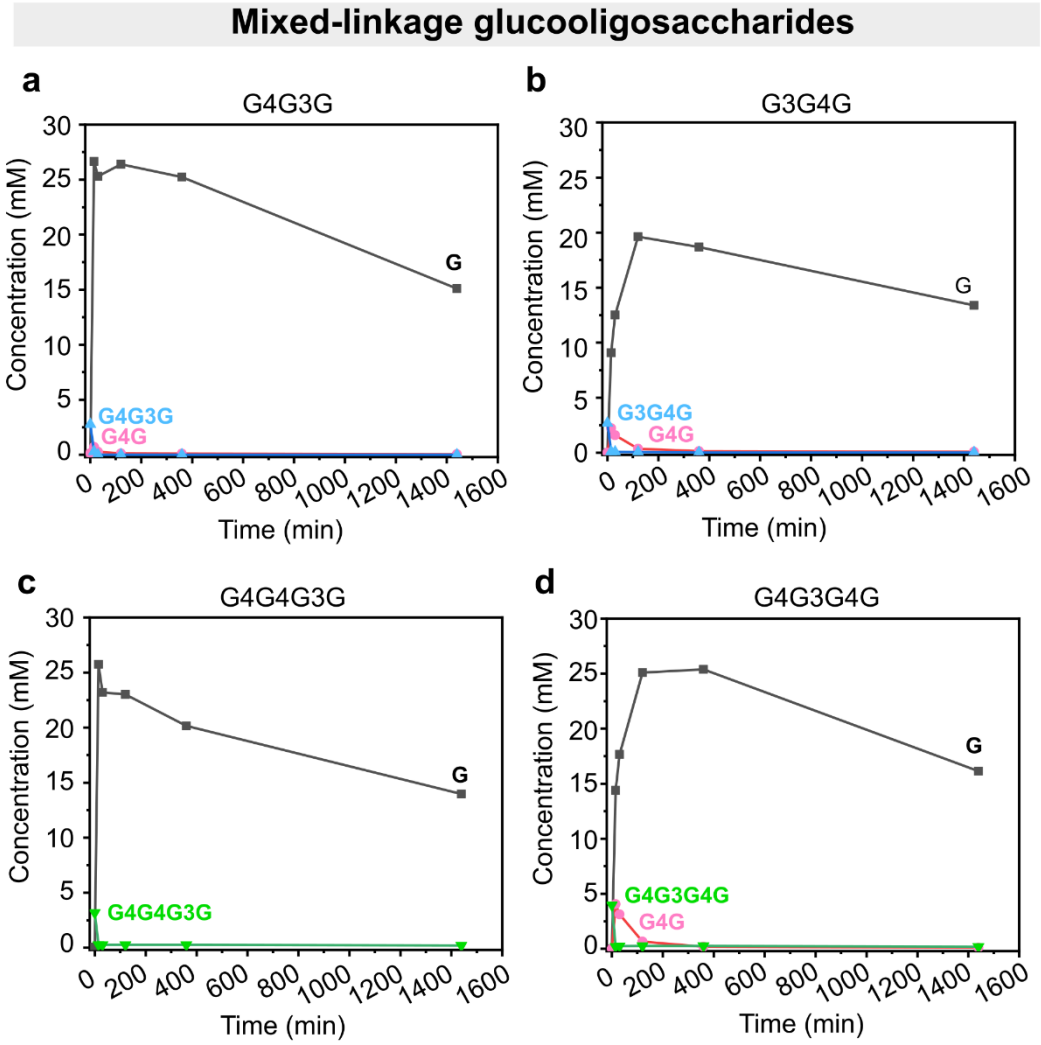
Supplementary Figure 10: Biochemical Characterization of CapGH3b. (a) Relative activity of CapGH3b on polymeric substrates **(b)** and p-nitrophenol-linked substrates (pNP). **(c)** Optimum pH and **(d)** temperature. Kinetics curves with **(e)** p-nitrophenol-β-D-glucopyranoside and **(f)** *Ld*Lam assessed at 35°C and pH 7. Results are expressed as mean ± standard deviations from three independent experiments (n=3).



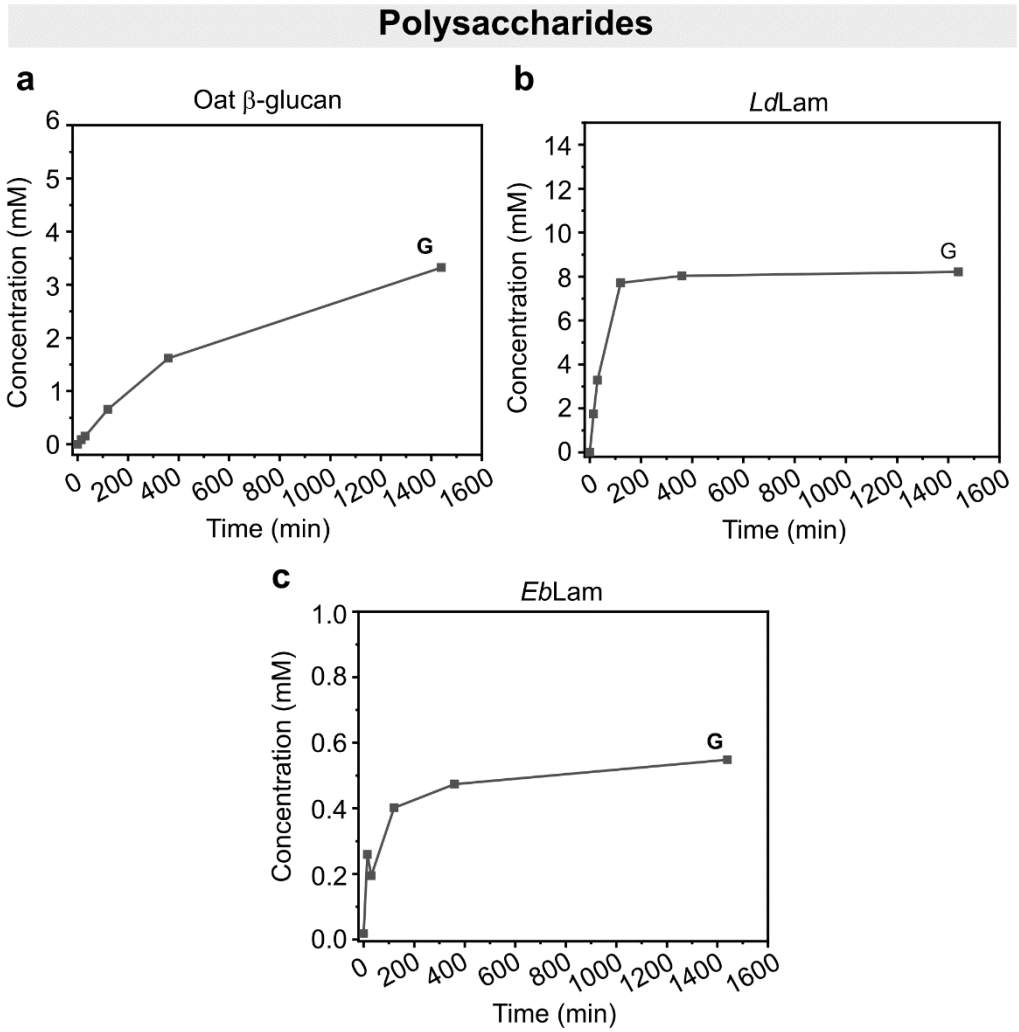
Supplementary Figure 11: Cleavage pattern of cellooligosaccharides by CapGH3b. The enzyme was able to completely depolymerize both short and long β(1,4)-glucooligosaccharides.



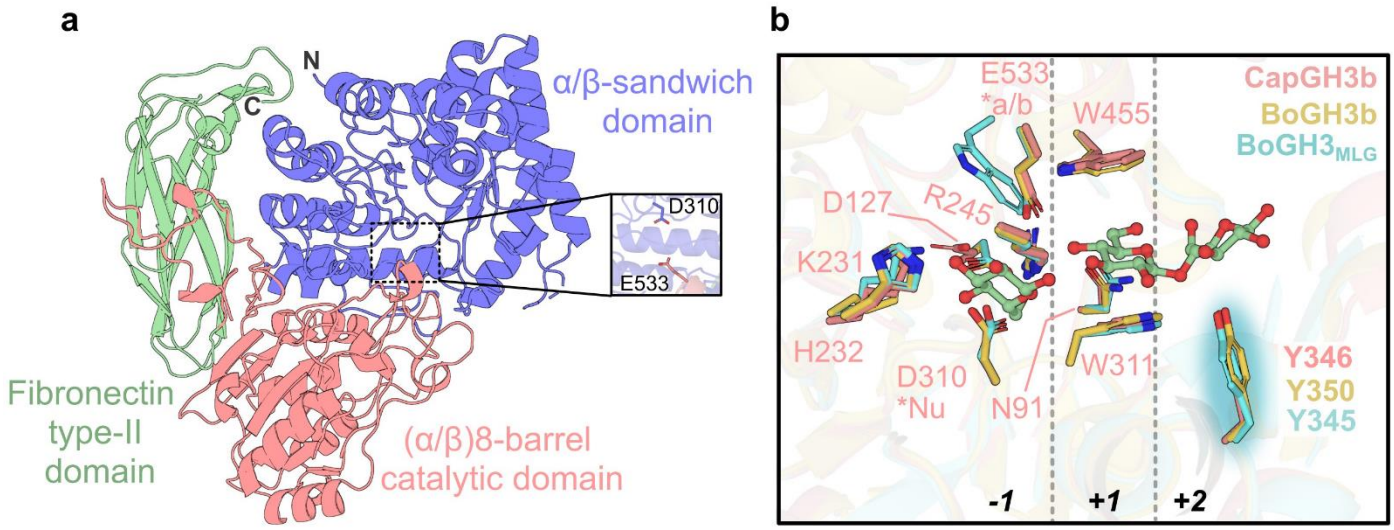
Supplementary Figure 12: Cleavage pattern of laminarioligosaccharides by CapGH3b. The enzyme was also able to utterly depolymerize both short and long β(1,3)-glucooligosaccharides.



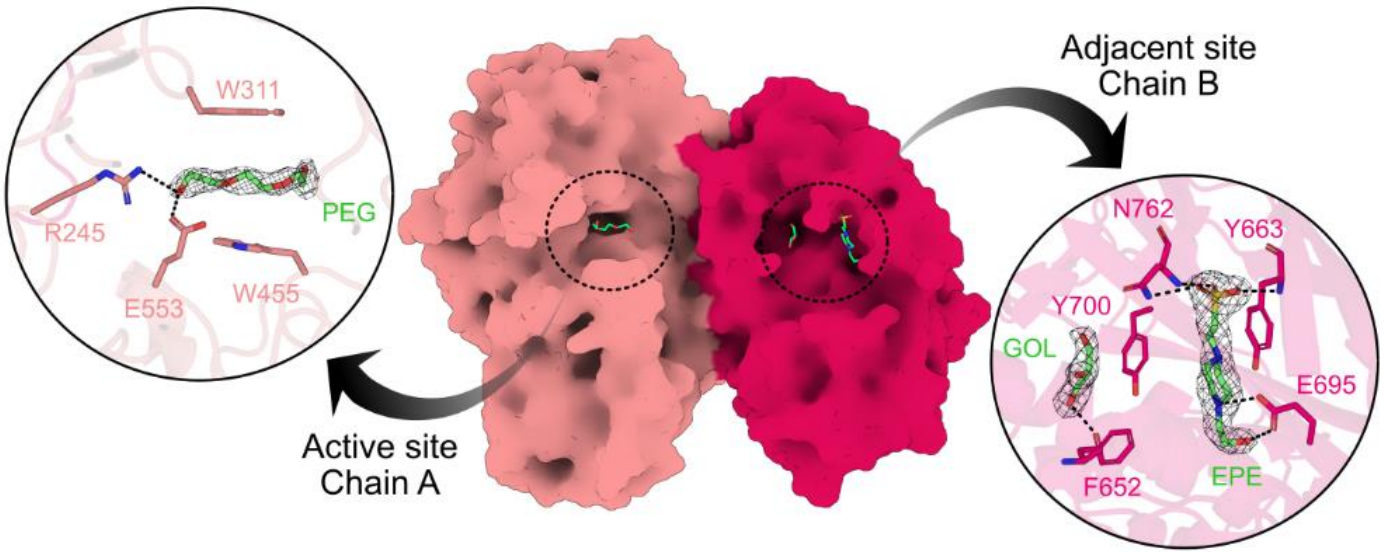
Supplementary Figure 13: Cleavage pattern of mixed-linkage oligosaccharides by CapGH3b. The depolymerization of mixed-linkage glucooligosaccharides shows the versatility of CapGH3b towards distinct β(1,3) and β(1,4) linkages present in glucooligosaccharides.



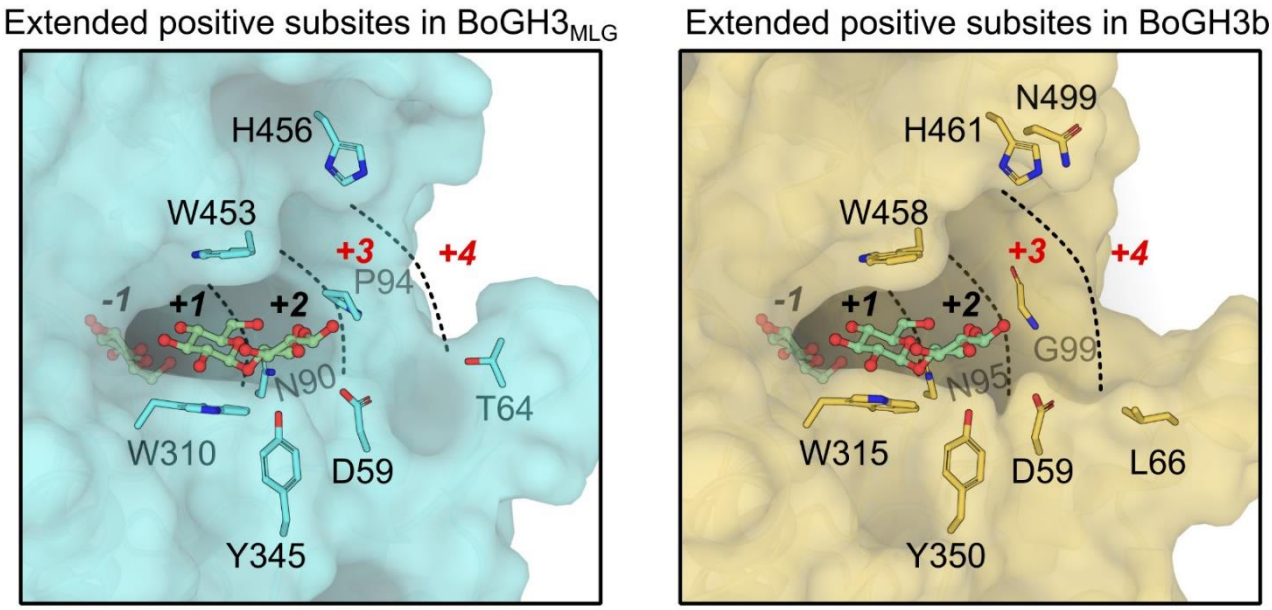
Supplementary Figure 14: Cleavage pattern of polysaccharides by CapGH3b. The only glucose release indicates an exo mode of action of CapGH3b.



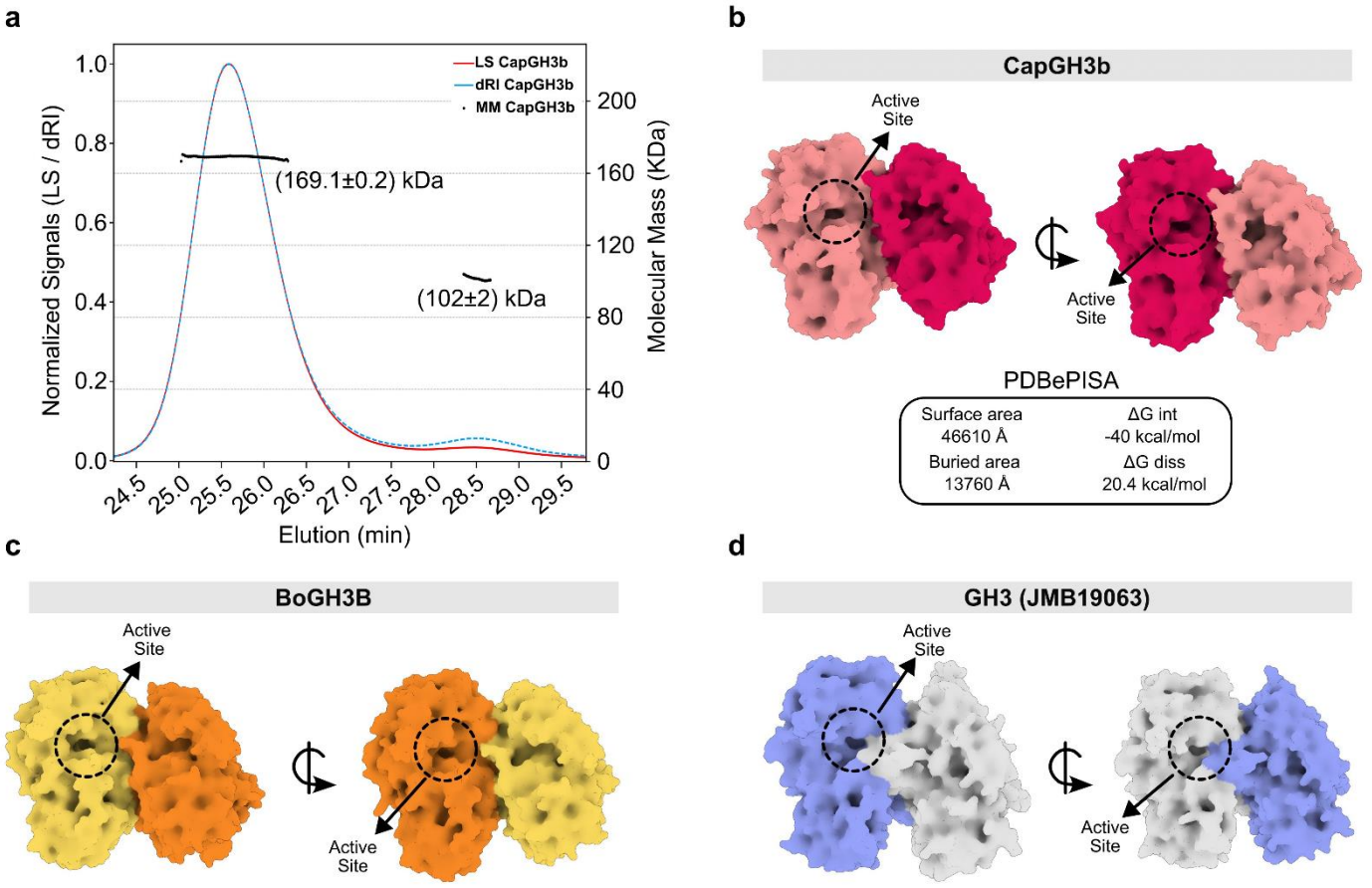
Supplementary Figure 15: Structure overview of CapGH3b and its active site. (a) CapGH3b domain organization highlighting the catalytic residues. **(b)** Active site comparison between CapGH3b and the orthologs *Bo*GH3B (PDID 5JP0[10]) and *Bo*GH3$_{MLG}$ (AlphaFold modeled[11], entry A7LY29). *Nu and *a/b represent the nucleophile and acid/base, respectively, inferred by structural superposition. The saccharide represented in the active site was extracted from the structural superposition with the GH3 β-glucosidase JMB19063 (PDBID 3U4A[12]).



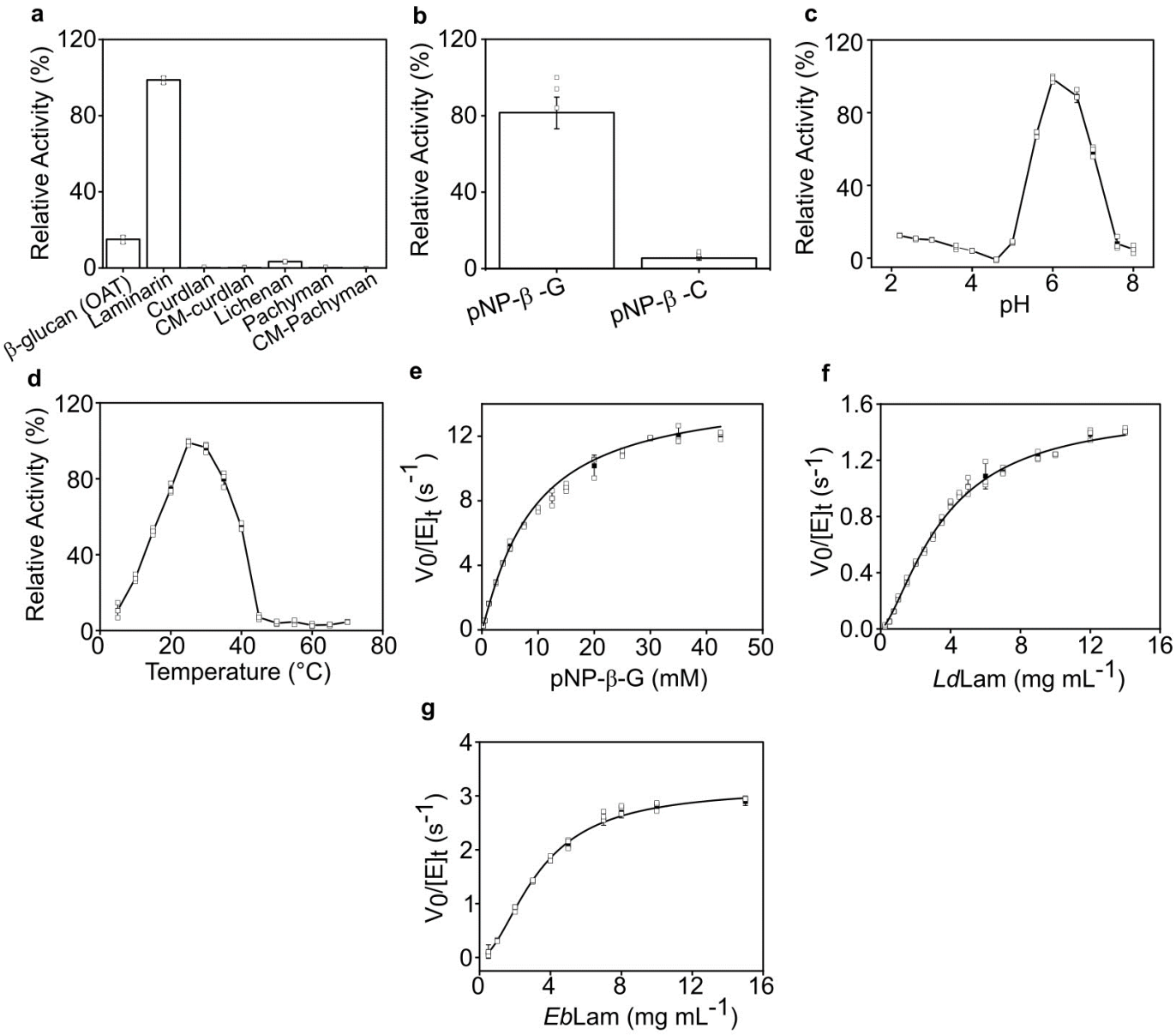
Supplementary Figure 16: Electron density maps of CapGH3b ligands observed in the crystallographic structure. The PEG molecule and its respective interactions in the active site of subunit A are represented, as the Hepes and glycerol molecules (EPE and GOL, respectively) interactions with the subunit B. The electron-density maps (2Fo - Fc) around ligands are represented at 1.0 sigma contour level.



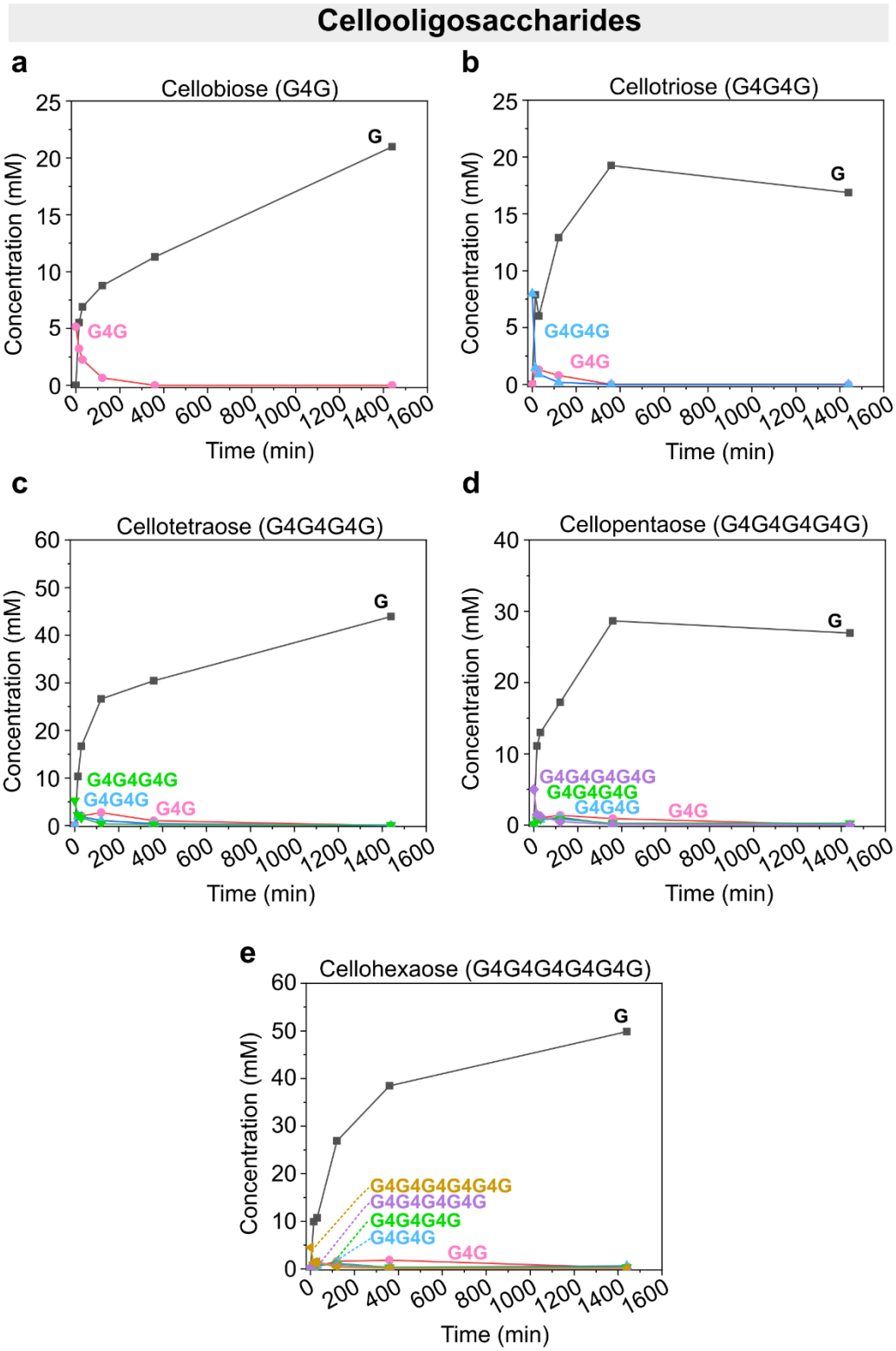
Supplementary Figure 17: Extension of the positive subsites observed for the CapGH3b orthologs *Bo*GH3MLG (AlphaFold modeled[11], entry A7LY29) and *Bo*GH3b (PDBID 5JP0[10]). The saccharide represented in the active site was extracted from the structural superposition with GH3 β-glucosidase JMB19063 (PDBID 3U4A[12]).



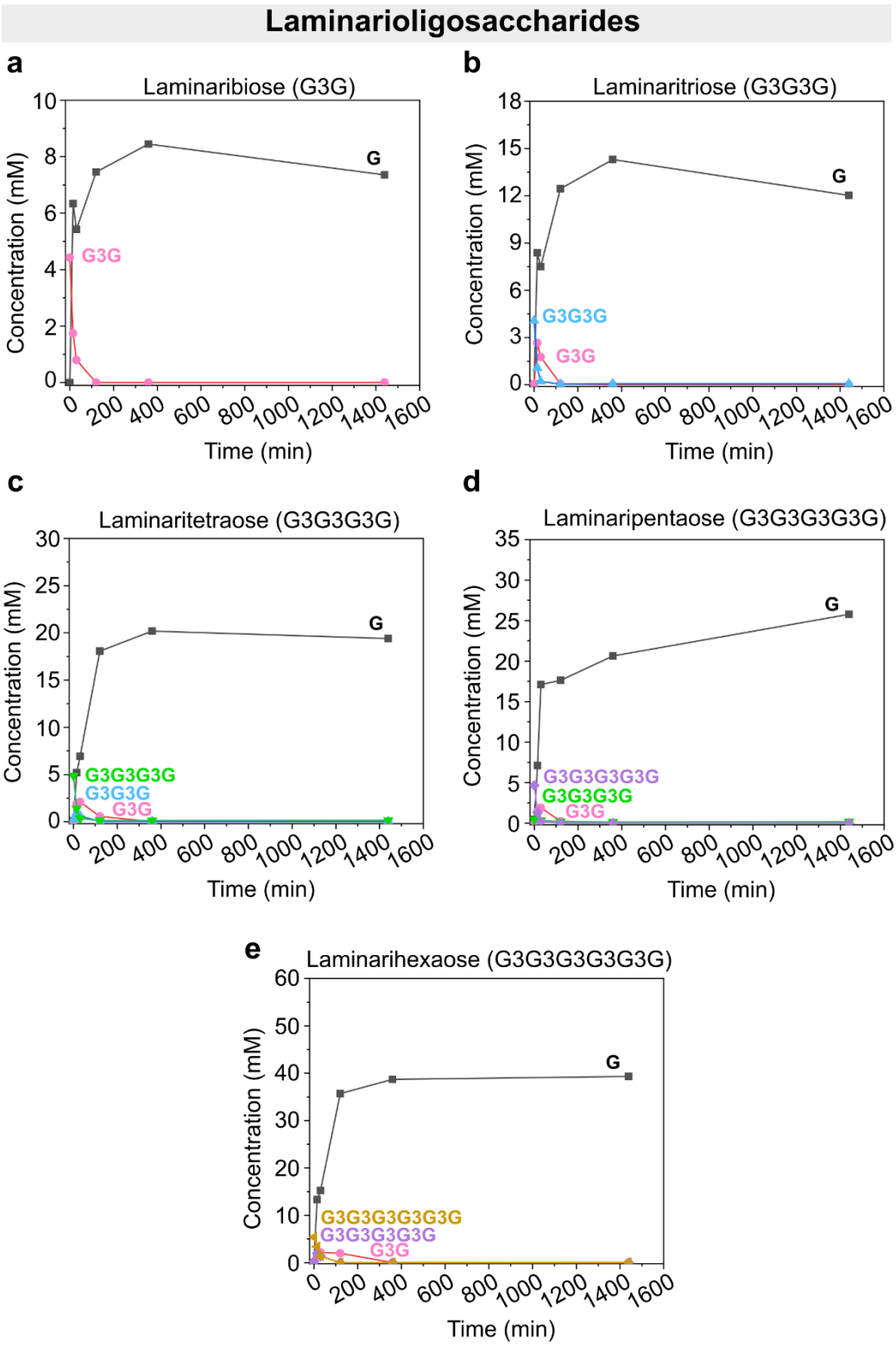
Supplementary Figure 18: CapGH3b oligomerization. (a) Size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) profile of CapGH3b. LS, dRI and MM are laser light scattering, refractive index signal and molecular mass, respectively. **(b)** CapGH3b dimer surface representation and PDBePISA[13] interface analysis. **(c)** *Bo*GH3B dimer surface representation (PDBID 5JP0[10]). **(d)** JMB19063 (GH3 recovered from metagenome) dimer surface representation (PDBID 3U4A[12]).



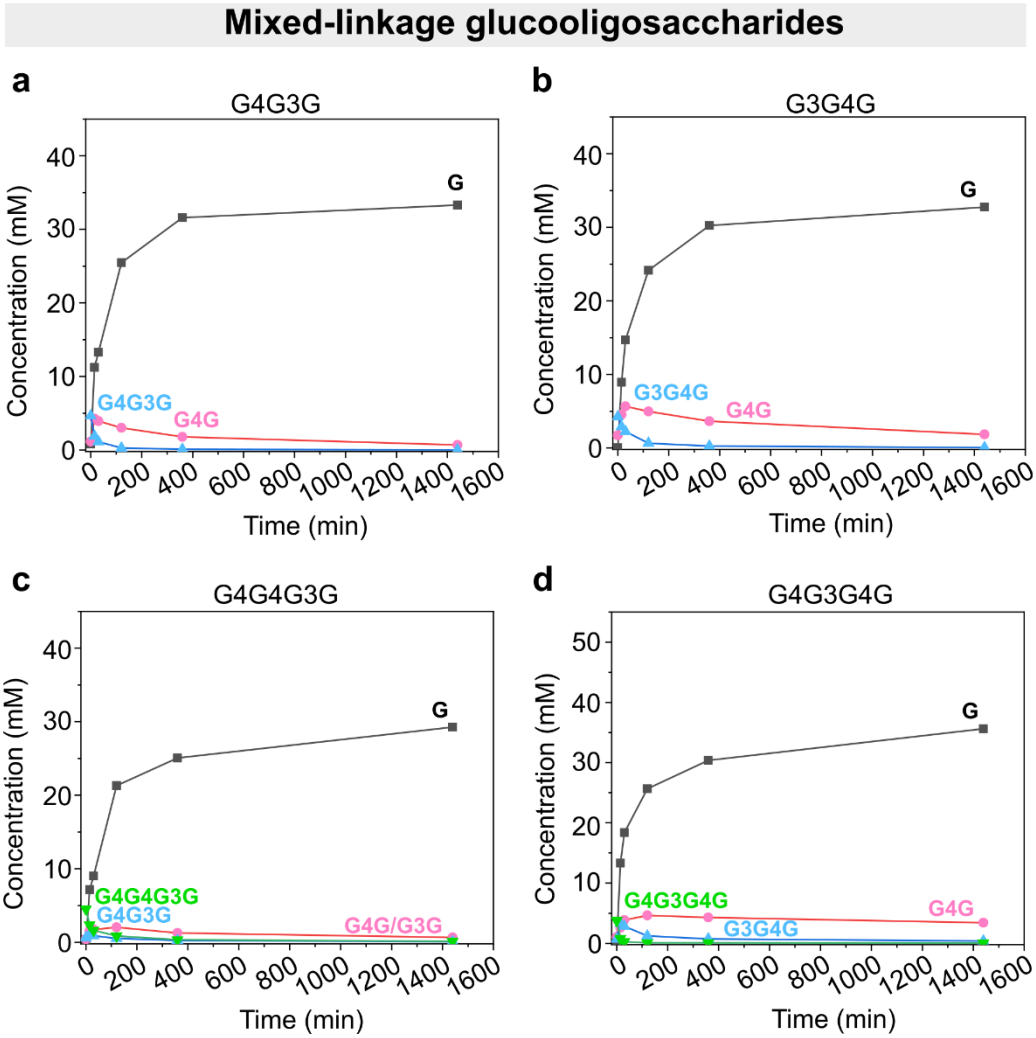
Supplementary Figure 19: Biochemical characterization of CapGH3a. (a) Relative activity of CapGH3a on polymeric substrates and **(b)** p-nitrophenol-linked substrates (pNP-G: 4-Nitrophenyl β-D-glucopyranoside and pNP-C: 4-Nitrophenyl β-D-cellobioside). Determination of **(c)** optimum pH and **(d)** temperature. Kinetics curves were obtained with **(e)** pNp-β-D-glucopyranoside, **(f)** *Ld*Lam, and **(g)** *Eb*Lam, assessed at 25°C and pH 6. Results are expressed as mean ± standard deviations from three independent experiments (n=3).



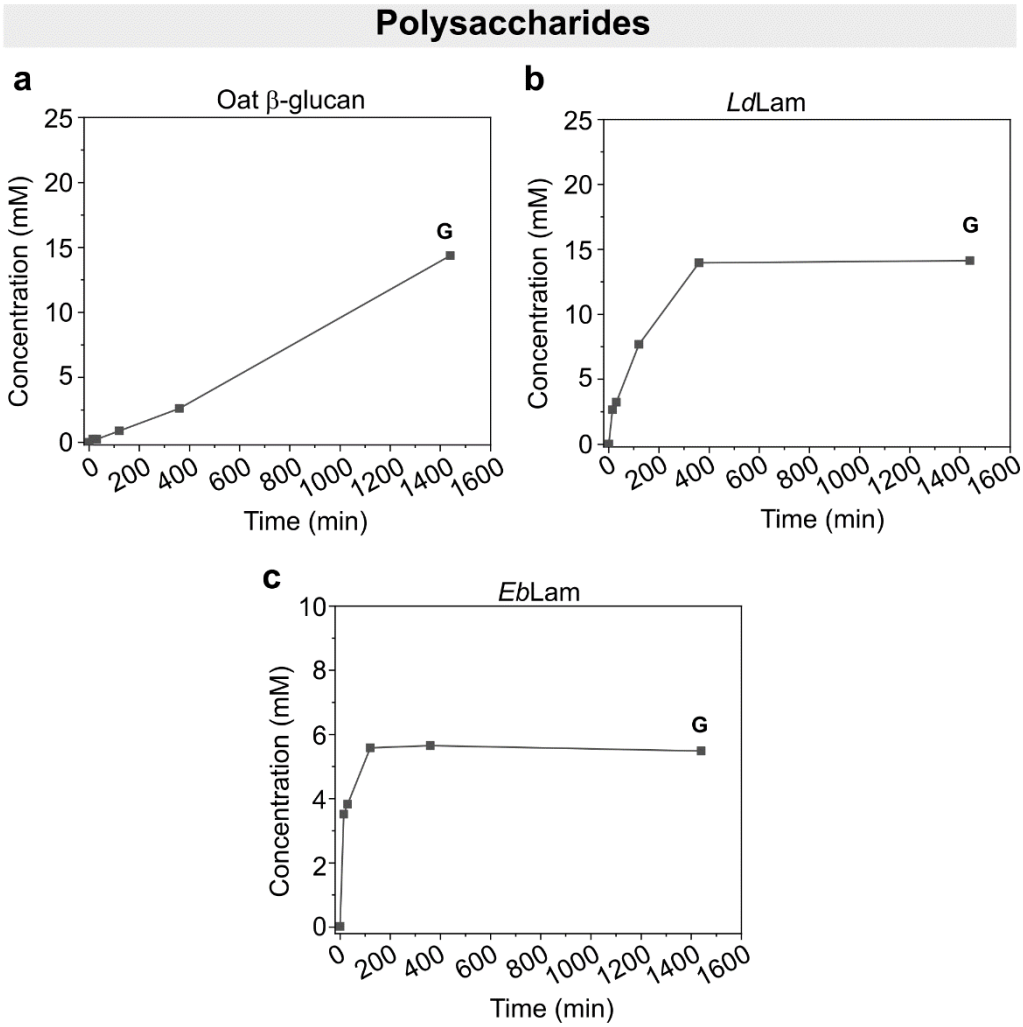
Supplementary Figure 20: Cleavage pattern of cellooligosaccharides by CapGH3a. Results indicate that CapGH3a is active on both short and long β(1,4)-glucooligosaccharides.



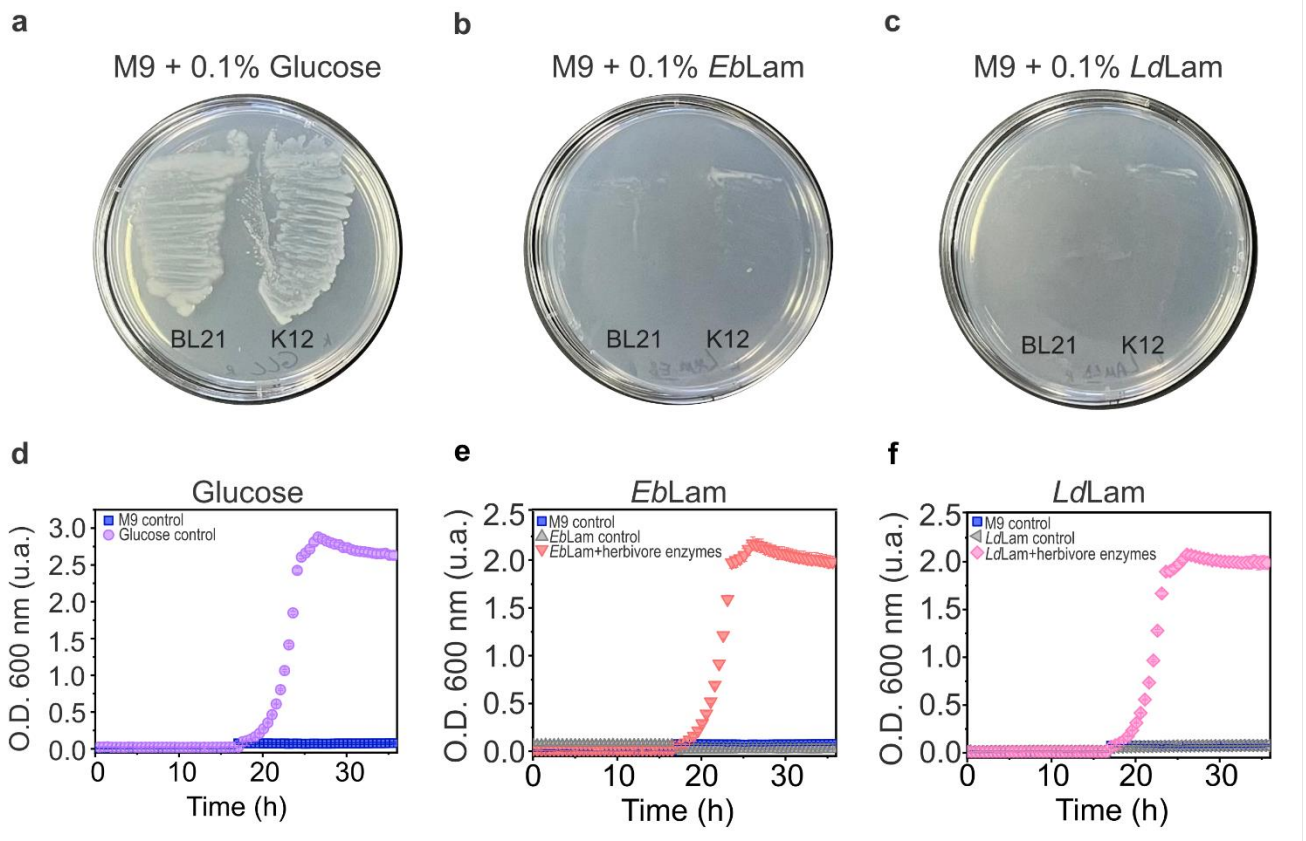
Supplementary Figure 21: Cleavage pattern of laminarioligosaccharides by CapGH3a. Results indicate that CapGH3a is active on both short and long β(1,3)-glucooligosaccharides.



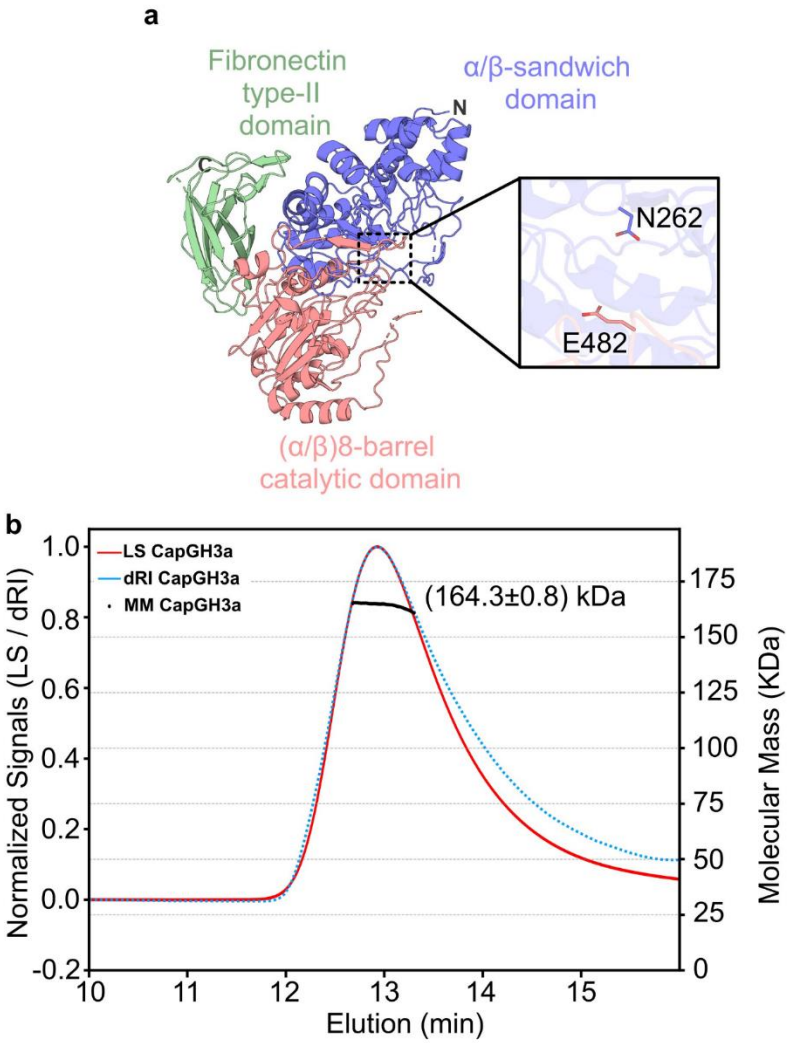
Supplementary Figure 22: Cleavage pattern of mixed-linkage glucooligosaccharides by CapGH3a. These results highlight that the enzyme can cleave different types of mixed-linkage glucooligosaccharides commonly generated by an endo-glucanase like CapGH16_3.



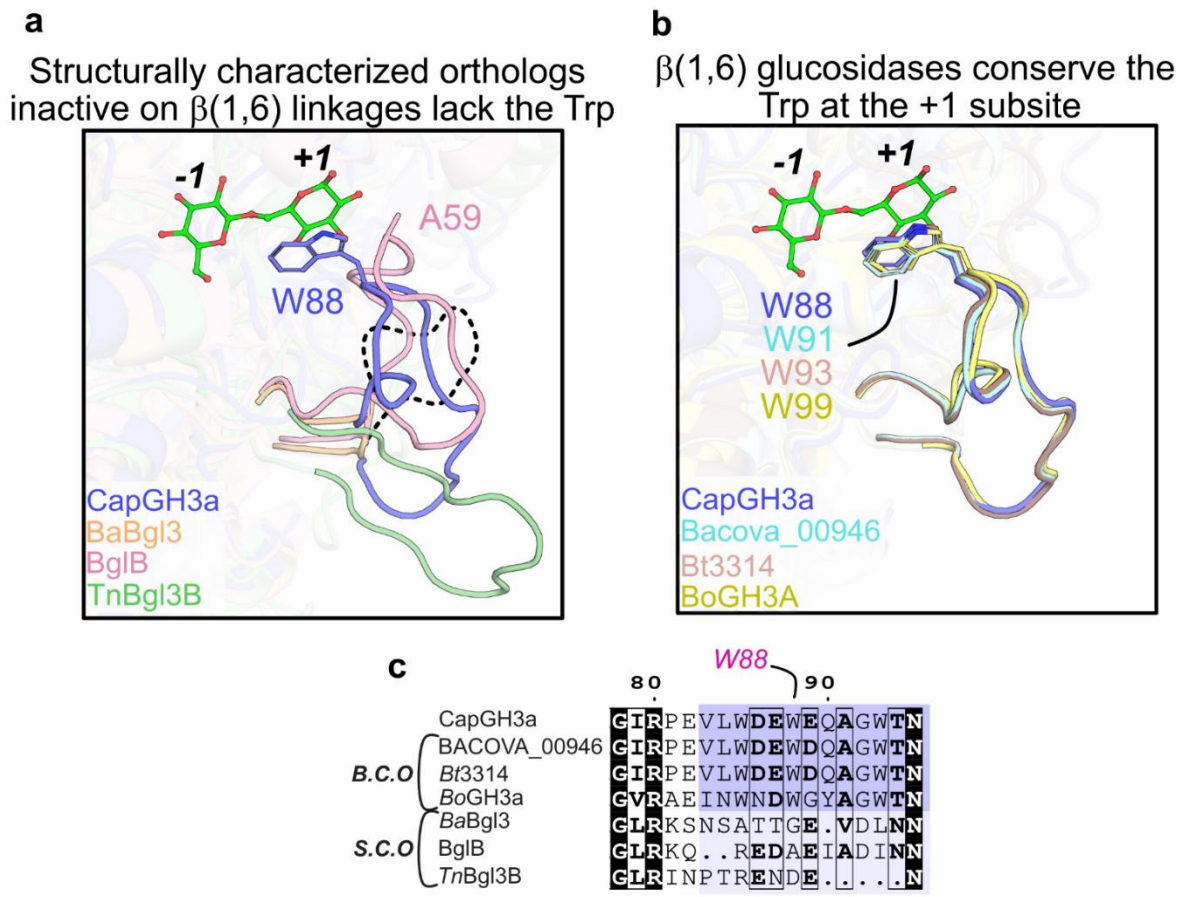
Supplementary Figure 23: Cleavage pattern of polysaccharides by CapGH3a. The exclusive release of glucose during hydrolysis indicates an exo mode of action of CapGH3a.



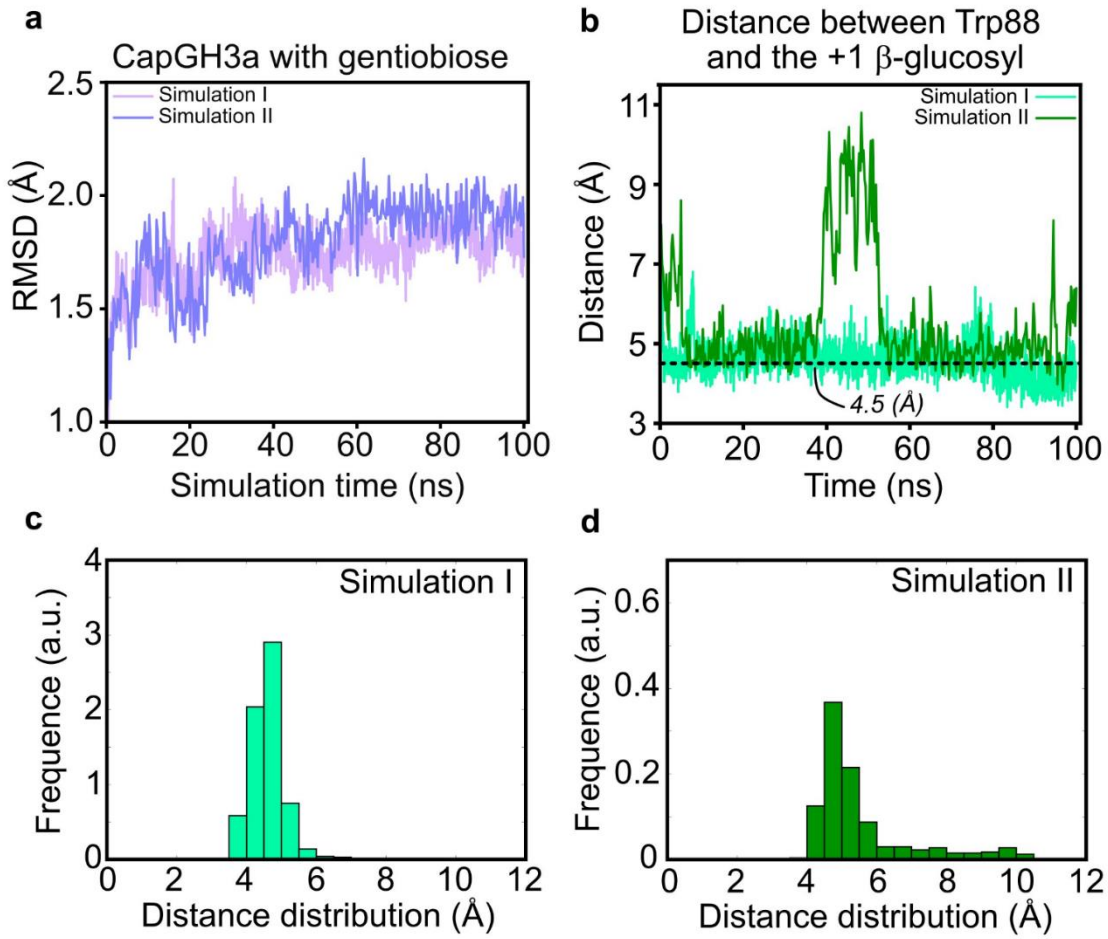
Supplementary Figure 24: Evaluation of *E. coli* growth in different carbon sources. *E. coli* BL21 (DE3) and K12 strains were able to grow in M9 medium supplemented with 0.1% **(a)** glucose, but not with the polysaccharides **(b)** *Eb*Lam and **(c)** *Ld*Lam. Cells were incubated at 37 ºC for 36 h. Since the PUL was obtained from an unculturable bacterium, growth curves were obtained in the presence of **(d)** 0.1% glucose, **(e)** 0.1% *Eb*Lam and **(f)** 0.1% *Ld*Lam, as carbon source, using the model bacterium *E. coli* K12 that naturally is not able to utilize β(1,3)-glucans, reiterating the capacity of this enzymatic system to release glucose from both linear and substituted β(1,3)-glucans to promote bacterial growth. Growth control was performed using M9 medium without carbon source (M9 control).



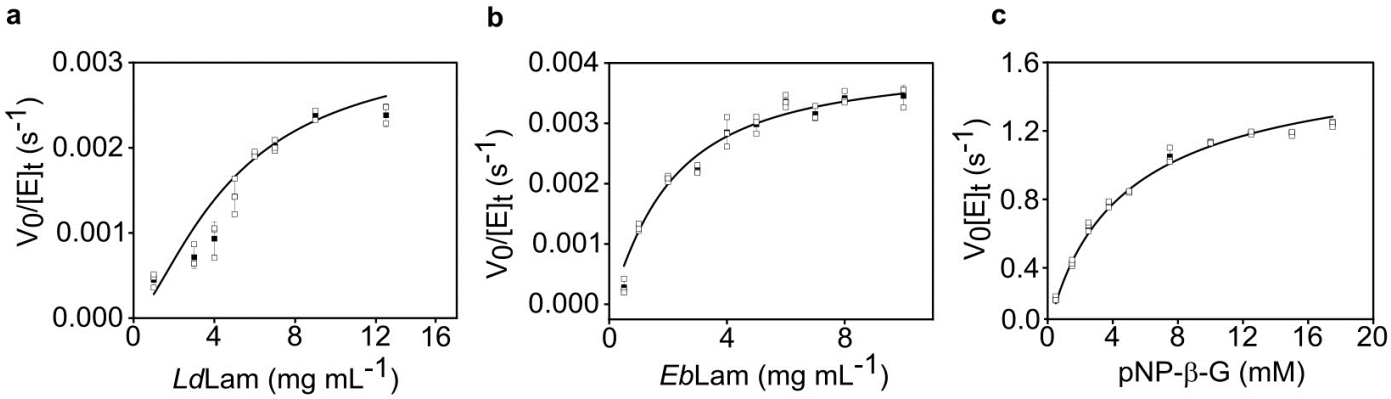
Supplementary Figure 25: CapGH3a domains and oligomerization. (a) Canonical GH3 family domains highlighting the catalytic residues, nucleophile Asp262 and acid/base Glu482, inferred by structural homology. **(b)** Size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) profile of CapGH3a. LS, dRI and MM are laser light scattering, refractive index signal and molecular mass, respectively. The monomeric and dimeric molecular weight of CapGH3a are 82 and 164 kDa, respectively.



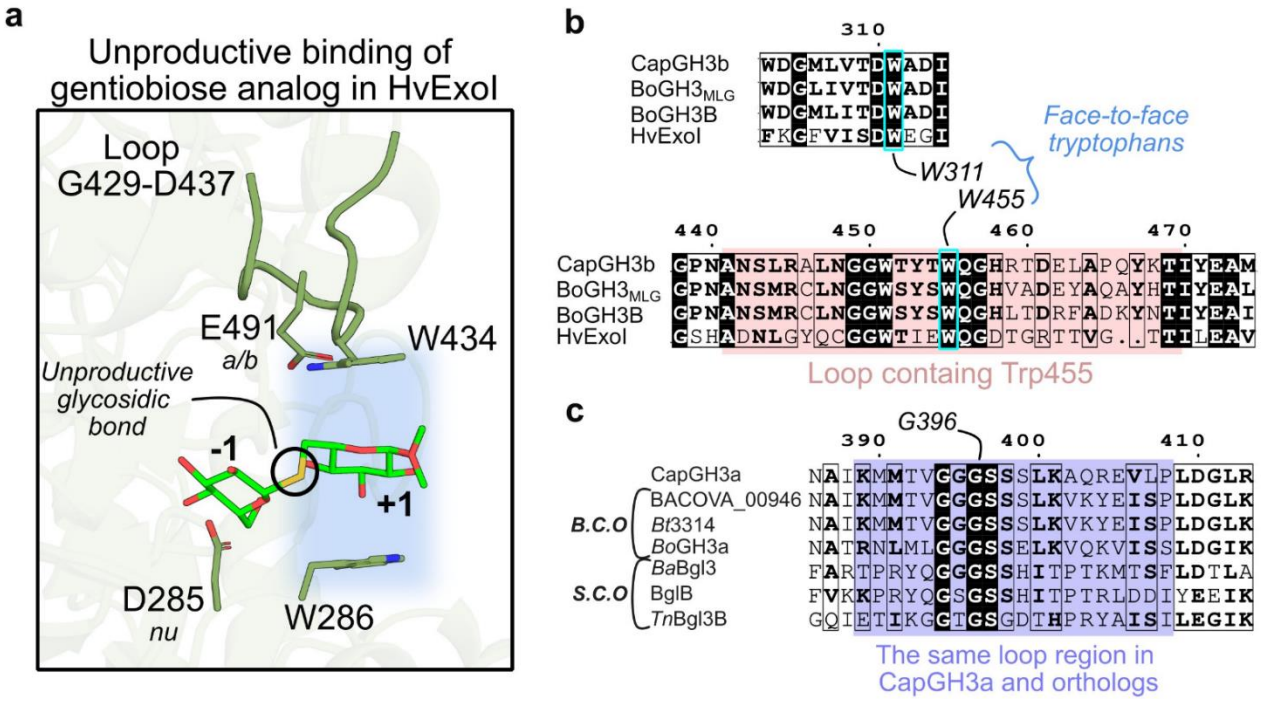
Supplementary Figure 26: The loop containing the Trp88 residue in CapGH3a has a pivotal role in the regioselectivity for β(1,6)-linkages. (a) Structural alignment of CapGH3a and its structurally characterized orthologs, *Ba*Bgl3 (PDBID 5WAB[14]), BglB (PDBID 7MS2[15]), and *Tn*Bgl3B (PDBID 2X40[16]). **(b)** Structural alignment of CapGH3a and biochemically characterized orthologs, Bacova_00946, *Bt*3314, and *Bo*GH3A (all modeled by AlphaFold[11] entries A7LT07, Q8A2J1, and A7LXS8, respectively). Gentiobiose from (a,b) was docked into the CapGH3a active site **(c)** Multiple sequence alignment of CapGH3a and biochemically characterized orthologs (B.C.O.) and structurally characterized orthologs (S.C.O). The region corresponding to the loop containing Trp88 is highlighted in blue. The alignment figure was generated with ESPript 3.0[17].



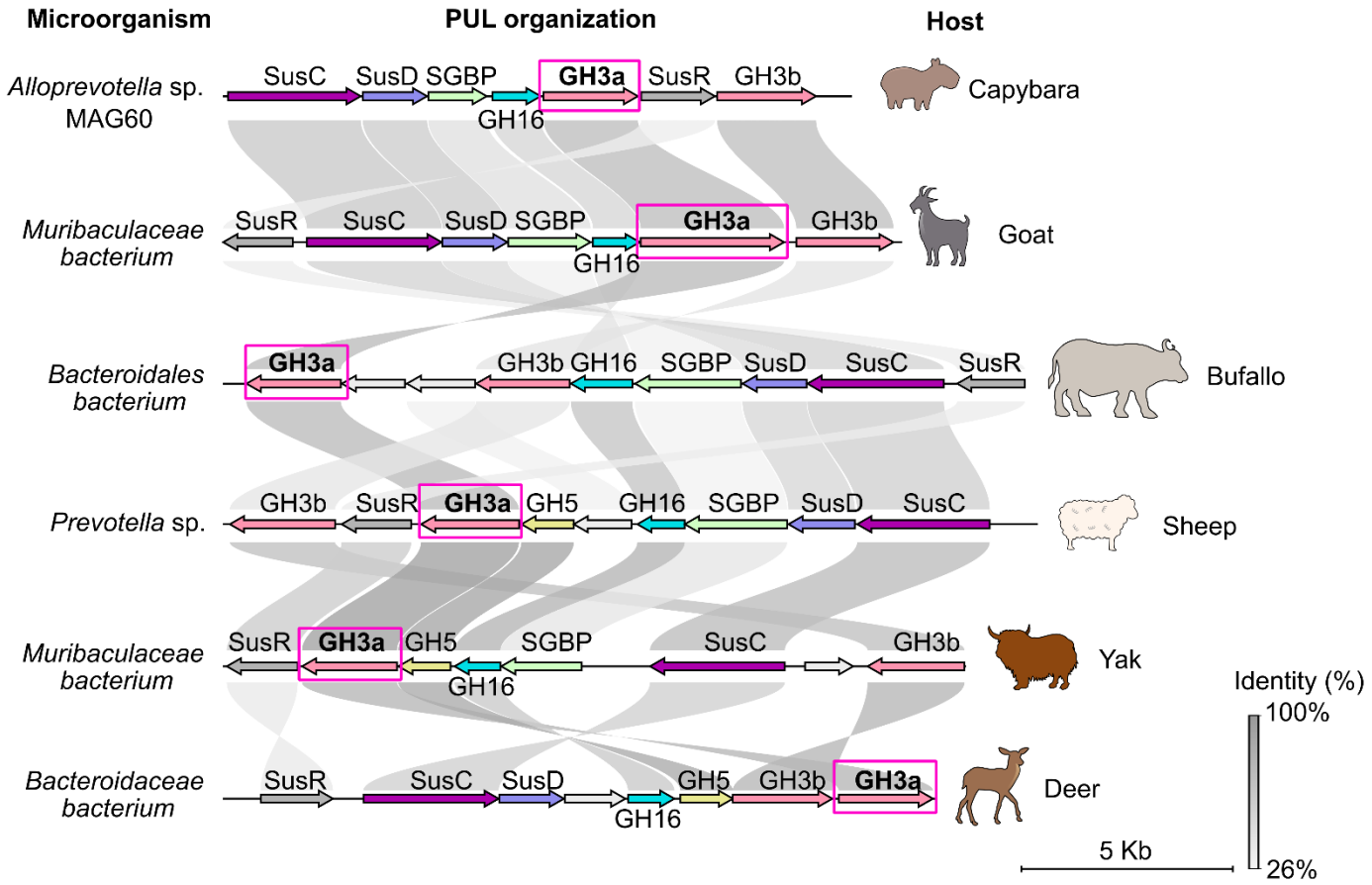
Supplementary Figure 27: Molecular dynamics simulations of CapGH3a with gentiobiose indicate the stabilization of the +1 glucosyl moiety by Trp88. (a) Root-mean-square deviations (RMSD) of the production trajectories of CapGH3a with gentiobiose. **(b)** Representation of the distance between the atom CG of CapGH3a Trp88 and the C5 of gentiobiose throughout 100 ns of production. **(c-d)** Distance distributions obtained from the graph represented in "b" for each system (I, c) and (II, d). All data was taken from two independent molecular dynamics simulations.



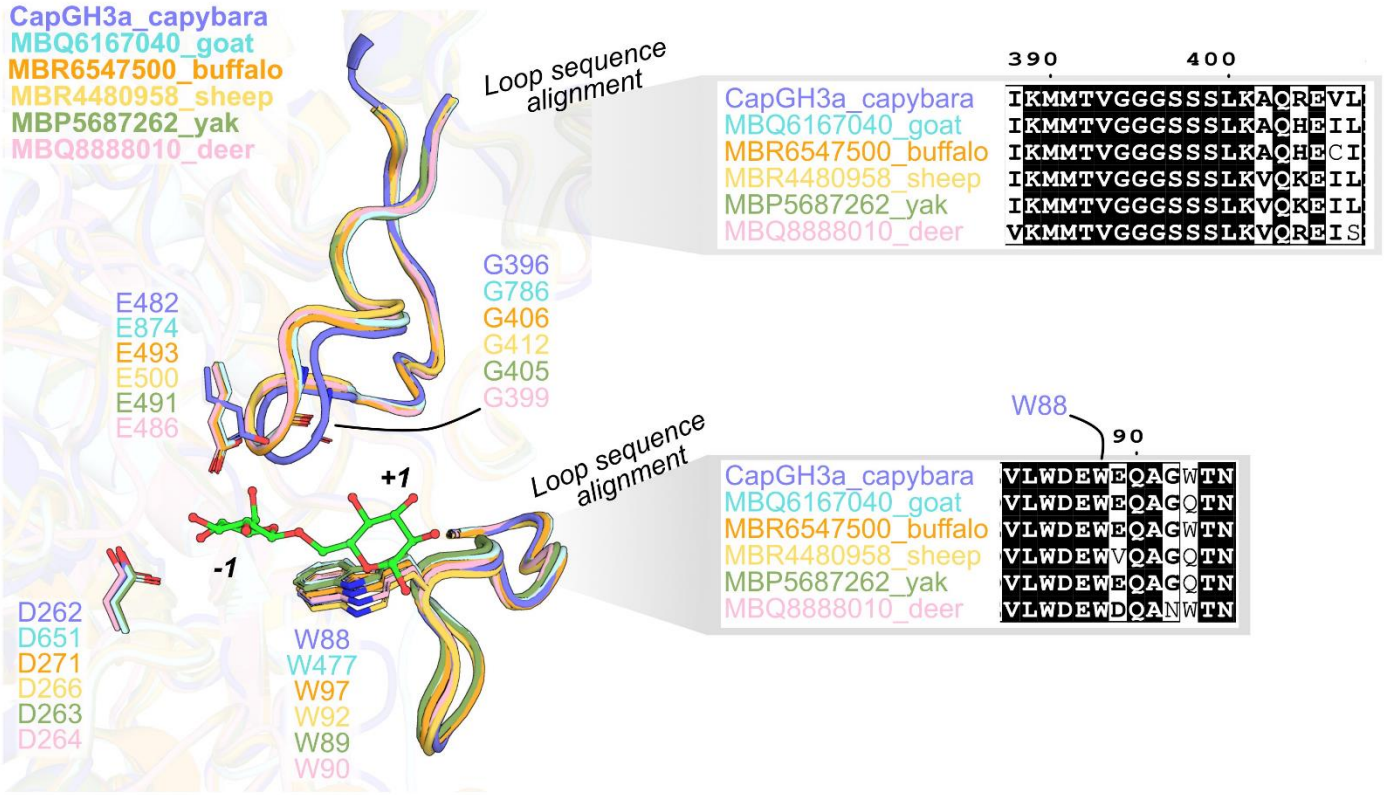
Supplementary Figure 28: Biochemical characterization of the CapGH3a W88A mutant. Kinetics curves obtained with **(a)** *Ld*Lam, **(c)** *Eb*Lam, and **(c)** pNp-β-D-glucopyranoside assessed at 25°C and pH 6. Results are expressed as mean ± standard deviations from three independent experiments (n=3).



Supplementary Figure 29: The two face-to-face tryptophan residues in the positive subsite region observed in CapGH3b and orthologs, not conserved in CapGH3a and its respective orthologs. (a) Unproductive binding of thio-gentiobioside in *Hv*ExoI (PDBID 3WLP[18]). **(b)** Multiple sequence alignment of CapGH3b with the orthologs *Bo*GH3$_{MLG}$, *Bo*GH3B, and *Hv*ExoI (32% SEQID with CapGH3b). The two face-to-face tryptophan residues forming the +1 subsite in CapGH3b and orthologs are highlighted. **(c)** Multiple sequence alignment of CapGH3a and its biochemically characterized orthologs (B.C.O.) and structurally characterized orthologs (S.C.O). The same loop region described in "b" for CapGH3b and orthologs is colored in blue, highlighting the presence of a glycine residue (Gly396 in CapGH3a) instead of the tryptophan described above. The alignment figure was generated with ESPript 3.0[17].



Supplementary Figure 30: PULs from herbivores harboring a GH3 equivalent to CapGH3a from capybara PUL. The links between the genomes are in grey scale according to the sequence identity of the proteins encoded.



Supplementary Figure 31: Structural and multiple sequence alignment of CapGH3a and herbivores orthologs demonstrating the conservation of the +1 subsite associated with the regioselectivity to β(1,6)-linkages. Structural models were obtained with AlphaFold2[3]. The alignment figure was generated with ESPript 3.0[17].

Supplementary Table 1: Average nucleotide identity (ANI) and alignment fraction (AF) between *Alloprevotella* sp. MAG60 and other closely related *Alloprevotella* strains.

Query Genome	Reference Genome	ANI	AF
Alloprevotella sp. MAG60	GCF_003859795.1 *Alloprevotella* sp. OH1205 COT-284	76.82	0.35
Alloprevotella sp. MAG60	GCF 014385435.1 *Alloprevotella* sp. Lung230	77.71	0.43
Alloprevotella sp. MAG60	GCA 015257125.1 *Prevotellaceae bacterium*	76.62	0.37
Alloprevotella sp. MAG60	GCA 905369775.1 *Prevotellaceae bacterium* UBA1746	76.73	0.37
Alloprevotella sp. MAG60	GCA 905371275.1 *Prevotellaceae bacterium*	75.75	0.22
Alloprevotella sp. MAG60	GCF 000318095.2 *Alloprevotella* sp. oral str. F0040	76.20	0.28
Alloprevotella sp. MAG60	GCA 015259235.1 *Alloprevotella* sp.	76.01	0.28
Alloprevotella sp. MAG60	GCA 900095835.1 *Prevotellaceae bacterium* Marseille-P2826	76.18	0.30

Supplementary Table 2: Set of polysaccharides used in enzyme assays.

Substrate
Arabinan (Sugar Beet)
Arabinogalactan (Larch Wood)
Arabinoxylan (Rye Flour - High viscosity 33cSt)
Avicel® PH-101
Carboxymethylcellulose sodium salt low viscosity (CMC)
CM-Curdlan (1,3-β-D-Glucan)
CM-Pachyman (1,3-β-D-Glucan)
Chitosan
Curdlan (1,3-β-D-Glucan)
Galactomannan (Carob; Low Viscosity)
Galactan (Potato)
Laminarin from *Laminaria digitata* (1,3:1,6-β-D-Glucan / 7:1)
Lichenan (Icelandic Moss) (1,3:1,4-β-D-Glucan / 1:2)
Linear 1,5-α-L-Arabinan (Sugar Beet) (500 mg)
Locust bean gum from *Ceratonia siliqua* seeds
Mannan (Ivory Nut)
Pachyman (1,3-β-D-Glucan)
Pectic Galactan (Potato)
Pectin from citrus peel
Polygalacturonic Acid (PGA)
Pullulan
Rhamnogalacturonan I (Potato)
Starch from potatoes
Xanthan gum from *Xanthomonas campestris*
Xylan from beechwood
Xyloglucan (Tamarind Seed)
β-Glucan (Oat; Medium Viscosity)
2-Hidroxyetil celulose (HEC)

Supplementary Table 3: Data collection and refinement statistics.

	CapGH16_3	**CapGH3a**	**CapGH3b**
PDB code	8VA4	8VA7	8VA3
Data collection			
Space group	P 43 21 2	C 1 2 1	P 21 21 21
Cell dimensions			
a; b; c (Å)	76.70; 76.70; 131.12	135.61; 132.87; 99.22	94.06; 105.35; 190.41
α; β; γ (o)	90; 90; 90	90; 116.27; 90	90; 90; 90
Resolution (Å)	41.79-1.96 (2.03-1.96)	47.22-2.6 (2.69-2.60)	47.6-1.8 (1.86 -1.80)
$CC_{1/2}$	0.99 (0.65)	0.99 (0.54)	0.99 (0.60)
I/σI	13.77 (1.04)	9.55 (1.02)	14.84 (1.41)
Completeness (%)	96.1 (97.9)	96.2 (98.4)	96.4 (80.2)
R_{meas}	0.22 (5.01)	0.20 (2.12)	0.14 (1.6)
Redundancy	23.2 (24.4)	6.8 (6.5)	10.2 (9.5)
Refinement			
Resolution	41.79-1.96	47.22-2.6	47.07-1.8
Number of reflections	27,631 (2,760)	46,772 (4,747)	168,991 (13,873)
R_{work} / R_{free}	0.222/0.271	0.215/0.263	0.155/0.187
B-factor ($Å^2$)	41.85	57.91	27.70
Macromolecules	41.71	58.05	26.56
Ligands	53.73	62.23	41.27
Water	43.10	46.50	35.63
r.m.s.d.			
Bond lengths (Å)	0.003	0.002	0.012
Bond angles (°)	0.57	0.42	1.09
Ramachandran Plot			
Favored (%)	97.07	94.06	97.28
Outliers (%)	0.00	0.43	0.20

Statistics for the highest-resolution shell are shown in parentheses.

Supplementary Table 4: Degree of synergism (DS) in laminarin depolymerization by mixtures of herbivore PUL enzymes. Reactions were carried out with 0.5 µg of CapGH16_3 and 1 µg of CapGH3a and CapGH3b and incubated at 35°C and pH 6. Activities were assessed by glucose oxidase method.

Substrate	**Reaction time (h)**	**CapGH16_3 + CapGH3b**	**CapGH16_3 + CapGH3a**	**CapGH16_3 + CapGH3a + CapGH3b**
Laminarin from *Laminaria digitada*	1	2.7	4.5	1.8
	2	2.1	3.7	1.5
	3	1.6	1.7	2.1
	19	2.2	1.3	1.7
Laminarin from *Eisenia Biciclys*	1	1.6	1.5	1.6
	2	1.9	1.7	2.1
	3	2.7	2.0	3.0
	19	1.7	2.5	2.7
Oat β-Glucan	1	2.9	0.8	0.8
	2	1.8	1.1	0.9
	3	5.4	2.5	5.1
	19	8.7	3.1	7.7

Supplementary References

1. Tamura, K. *et al.* Surface glycan-binding proteins are essential for cereal beta-glucan utilization by the human gut symbiont *Bacteroides ovatus*. *Cell. Mol. Life Sci.* **76**, 4319–4340 (2019).
2. Tamura, K., Dejean, G., Van Petegem, F. & Brumer, H. Distinct protein architectures mediate species-specific beta-glucan binding and metabolism in the human gut microbiota. *J. Biol. Chem.* **296**, 100415 (2021).
3. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
4. Svergun, D., Barberato, C. & Koch, M. H. CRYSOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
5. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).
6. Liberato, M. V. *et al.* Insights into the dual cleavage activity of the GH16 laminarinase enzyme class on β-1,3 and β-1,4 glycosidic bonds. *J. Biol. Chem.* **296**, 100385 (2021).
7. Bleicher, L. *et al.* Molecular basis of the thermostability and thermophilicity of laminarinases: X-ray structure of the hyperthermostable laminarinase from *Rhodothermus marinus* and molecular dynamics simulations. *J. Phys. Chem. B* **115**, 7940–7949 (2011).
8. Labourel, A. *et al.* Structural and biochemical characterization of the laminarinase ZgLamCGH16 from *Zobellia galactanivorans* suggests preferred recognition of branched laminarin. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71**, 173–184 (2015).
9. Yang, J., Xu, Y., Miyakawa, T., Long, L. & Tanokura, M. Molecular Basis for Substrate Recognition and Catalysis by a Marine Bacterial Laminarinase. *Appl. Environ. Microbiol.* **86**, (2020).
10. Hemsworth, G. R. *et al.* Structural dissection of a complex Bacteroides ovatus gene locus conferring xyloglucan metabolism in the human gut. *Open Biol.* **6**, (2016).
11. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
12. McAndrew, R. P. *et al.* From soil to structure, a novel dimeric β-glucosidase belonging to glycoside hydrolase family 3 isolated from compost using metagenomic analysis. *J. Biol. Chem.* **288**, 14985–14992 (2013).
13. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797 (2007).
14. Florindo, R. N. *et al.* Structural and biochemical characterization of a GH3 β-glucosidase from the probiotic bacteria *Bifidobacterium adolescentis*. *Biochimie* **148**, 107–115 (2018).
15. Almeida, L.R., Muniz, J. R. C. Three-dimensional structure of a GH3 Beta-glucosidase from *Clostridium thermocellum* in complex with glycerol. https://www.rcsb.org/structure/7ms2 (2022).
16. Pozzo, T., Pasten, J. L., Karlsson, E. N. & Logan, D. T. Structural and Functional Analyses of β-Glucosidase 3B from *Thermotoga neapolitana*: A Thermostable Three-Domain Representative of

Glycoside Hydrolase 3. *J. Mol. Biol.* **397**, 724–739 (2010).

17. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, (2014).
18. Streltsov, V. A. *et al.* Discovery of processive catalysis by an exo-hydrolase with a pocket-shaped active site. *Nat. Commun.* **10**, 1–17 (2019).