Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Engineering the topology of multienzyme glycoside hydrolase complexes reveals distinct effects on soluble and insoluble polysaccharide degradation

Iker Pardo Larrabeiti
TBI, Université de Toulouse, CNRS, INRAE, INSA
Bastien Annic
INRAE, UR1268 Biopolymères Interactions Assemblages
Mathieu Fanuel
INRAE, UR1268 Biopolymères Interactions Assemblages
Hélène Rogniaux
INRAE, UR1268 Biopolymères Interactions Assemblages
David Ropartz
INRAE, UR1268 Biopolymères Interactions Assemblages
Barinder Bajwa
Agriculture and Agri-Food Canada
D. Wade Abbott
Agriculture and Agri-Food Canada
Sarah Moraïs
Ben-Gurion University of the Negev
Itzhak Mizrahi
Ben-Gurion University of the Negev
Edward A. Bayer
Ben-Gurion University of the Negev
Pierre Roblin
Univ. Toulouse, CNRS, LGC
Claire Dumon
TBI, Université de Toulouse, CNRS, INRAE, INSA
Cédric Y. Montanier
cedric.montanier@insa-toulouse.fr

TBI, Université de Toulouse, CNRS, INRAE, INSA

Research Article

Keywords: multienzyme complexes, glycoside hydrolases, spatial proximity, xylanase, endoglucanase

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11726136/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Engineering the topology of multienzyme glycoside hydrolase complexes reveals distinct effects on soluble and insoluble polysaccharide degradation

Iker Pardo Larrabeiti[1], Bastien Annic[2,3], Mathieu Fanuel[2,3], Hélène Rogniaux[2,3], David Ropartz[2,3], Barinder Bajwa[4], D. Wade Abbott[4], Sarah Moraïs[5], Itzhak Mizrahi[5], Edward A. Bayer[5,6], Pierre Roblin[7], Claire Dumon[1], Cédric Y. Montanier[1*]

[1]TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France

[2]INRAE, UR1268 Biopolymères Interactions Assemblages, Nantes F-44300, France

[3]INRAE, PROBE research infrastructure, BIBS facility, Nantes F-55300, France

[4]Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, 5403 1st Avenue South, Lethbridge, AB T1J 4B1, Canada

[5]Department of Life Sciences, Ben-Gurion University of the Negev, 8410501, Be'er Sheva, Israel

[6]Department of Biomolecular Sciences, The Weizmann Institute of Science, Rehovot 76100 Israel

[7]Univ. Toulouse, CNRS, LGC, Toulouse, France

*Corresponding author:

Cédric Y. Montanier, TBI INSA Toulouse 135 avenue de Rangueil 31077 Toulouse cedex 04 France. cedric.montanier@insa-toulouse.fr, +33 561 55 97 13

Background: Plant cell walls are structurally complex, and certain anaerobic microorganisms cope with this complexity by producing a wide range of glycoside hydrolases (GHs) organized into multienzyme complexes, the cellulosomes, to release oligosaccharides as a carbon source.

Owing to their intrinsic flexibility, determining how the spatial organization within these assemblies influences catalysis remains a major challenge.

Results: We used the Jo–In protein pair, which spontaneously forms a covalent isopeptide bond, to hold fused enzymes in rigid, defined arrangements. Three cellulosomal GHs from *Acetivibrio thermocellus* (*At*Cel8A, *At*Cel9R and *At*Xyn11A) were assembled into a library of 12 tri-enzymatic Jo-In complexes. Toward soluble polysaccharides, the assemblies largely preserved the activity of the free enzymes, whereas their efficiency varied markedly on insoluble substrates such as wheat bran and wheat straw. Mass spectrometry revealed that the oligosaccharide product profiles depended more on the substrate type than on the assembly architecture, although "cellulolytic" and "xylanolytic" assemblies produced distinct oligosaccharides on wheat straw. SAXS revealed similar global dimensions but different interdomain arrangements, with no significant correlation between interdomain distances and product profiles.

Conclusions: Enzyme topology affects the kinetics of insoluble biomass degradation, although it has little effect on activity on soluble substrates. These effects are predicted neither by soluble-substrate activity nor by pairwise interdomain distances alone, indicating that parameters such as enzyme orientation, accessibility and conformational freedom must also be considered. The Jo-In platform provides a framework to separate these contributions in the design of programmable multienzyme complexes.

Keywords: multienzyme complexes, glycoside hydrolases, spatial proximity, xylanase, endoglucanase

Background

The plant cell wall (PCW) is highly recalcitrant to enzymatic deconstruction, displaying a composite structure in which crystalline cellulose is embedded in a network of hemicelluloses, pectins and lignin [1]. To access the oligosaccharides they contain, some lignocellulolytic anaerobic microorganisms such as *Acetivibrio thermocellus* (previously termed *Clostridium thermocellum*) produce a large and diverse repertoire of glycoside hydrolases (GHs), organized into multienzyme complexes called cellulosomes [2]. In these assemblies, enzymes carrying a dockerin module bind to the cohesin modules of a non-catalytic scaffoldin, which also anchors the complex to the insoluble substrate through a carbohydrate-binding module (CBM). This organization is generally thought to promote synergy between complementary activities acting on the same substrate, by bringing enzymes into close proximity at the surface of the cellulosic material [3]. Yet how the enzymes function together within this confined space remains poorly understood. Simple cellulosomes accommodate 2 to 9 enzymes whereas highly structured ones may contain 11 to 160 enzymes [2], as reported from proteomic analyses, for example in *Clostridium clariflavum* [4]. Moreover, much larger entities have been described, including 16-MDa aggregates [5] and polycellulosomes of up to 100 MDa observed by electron microscopy [6]. Even if the real number of grafted enzymes is lower, many of them must operate cooperatively, at high local concentration, in the immediate vicinity of an insoluble substrate. In this context, the flexibility of the cellulosome, arising from the inter-cohesin linkers as well as from the linkers within the grafted GHs, is frequently invoked as a key determinant of its efficiency, as it would allow enzymes to explore the substrate surface and adapt to its heterogeneity [7]. However, this flexibility is more often assumed than measured, and its link with efficiency has rarely been demonstrated directly. In a crowded environment, steric hindrance may also restrict movement, so that an appropriate distance between enzymes could matter more than freedom of motion. The dynamics of such systems therefore remain difficult to characterize, and it remains unclear whether enzyme positioning within the complex and

assembly topology, defined by the spatial arrangement and connectivity of the enzymes, independently influence activity beyond the effects of flexibility. Designer cellulosomes (DCs), which mimic native cellulosomes by artificially assembling enzymes at defined positions on a short scaffoldin [8], have been the main tool to address these questions. Enzyme proximity clearly improves activity on recalcitrant substrates compared with the free enzyme [9–11]. The role of linkers, however, remains debated. Molinier et al. proposed that physical proximity allows enzymes to cleave more adjacent glycosidic bonds on the same cellulose chains, while inter-cohesin linkers of 4 to 128 amino acids had no impact on synergy or activity [12]. In contrast, Vazana et al. found the lowest activity in the absence of linkers, and activity improved with linkers ranging from 5 to 35 amino acids [13]. Although these comparisons are debatable because the enzymes and linkers differ, the relative position of GHs within a scaffoldin has also been shown to be critical for the optimal degradation of microcrystalline cellulosic substrates [14]. More recently, systematic studies based on random or combinatorial approaches have sought the best enzyme positions along the scaffoldin [11,15,16]. At the scale of a DC, however, and given the flexibility of the assembly, colocalization is inherently blurred: the distance between two enzymes is not fixed but distributed over many conformations. It is therefore difficult to establish a quantitative relationship between topology and catalytic function, and the behavior of these assemblies at the molecular scale remains a black box. A large body of literature also describes ways of immobilizing enzymes to improve synergy [17,18], but almost none questions the relationship between distance, topology and activity.

To address this limitation, we propose to constrain enzyme topology and study it systematically, rather than average over all possible conformations. For this purpose, we used the Jo–In protein pair [19,20], which forms a covalent isopeptide bond spontaneously, to build rigid scaffolds in which fused enzymes are held in defined positions relative to one another [21]. Three cellulosomal GHs from *A. thermocellus* (endoglucanase AtCel8A [22], processive

endoglucanase AtCel9R [23] and xylanase AtXyn11A [24]) were assembled into a library of 12 Jo-In based complexes, whose activity was characterized on soluble and insoluble wheat coproducts (destarched bran and straw). Combining DNS assays, MALDI-TOF mass spectrometry and small-angle X-ray scattering (SAXS), we relate interdomain arrangement to catalytic activity and to the profile of released oligosaccharides. This systematic approach provides elements of a response to the putative role of spatial topology in multienzyme catalysis, together with a method that can be further refined. It also helps explain why flexible systems cannot resolve these effects, and thus serves to orient future engineering strategies.

Methods

Chemicals

2,5-Dihydroxybenzoic acid (DHB) was purchased from Sigma–Aldrich Co. (Saint Quentin Fallavier, France). *N*,*N*-dimethylaniline (DMA) was obtained from Fisher Scientific (Fisher Bioblock Scientific S. A, Illkirch, France). Acetonitrile (MeCN) and water (H2O) was of HPLC grade (Carlo-Erba Reagents, Val de Reuil, France). Pentaacetyl-chitopentaose > 95% purity was purchase from Megazyme (Bray, Ireland).

Glycosidic linking analysis

Linkage analysis of unfractionated dWB and WS were conducted following the protocols from [25]. Biomasses were ball milled into fine powder. Alcohol-insoluble residue (AIR) was prepared by extracting the samples three times with 80% (v/v) ethanol, with overnight agitation on a tube rotator, followed by centrifugation at 3,000 × *g* for 30 min. The residues were then washed once with acetone, and twice with ethanol with each wash performed for 20 min on the tube rotator. For all extraction and washing steps, solvents were added at a ratio of 1:10 (w/v). Following the final ethanol wash, residues were vacuum-dried using a Savant SPD131DDA

SpeedVac concentrator equipped with a Savant RVT5105 refrigerated vapor trap. Approximately 5 mg of each sample was converted to partially methylated alditol acetates (PMAAs) in triplicate, following [25], using myo-inositol as an internal standard. Briefly, AIR samples were subjected to one round of permethylation using dimethyl sulfoxide, sodium hydroxide, and methyl iodide, followed by liquid-liquid extraction. In a separate procedure, uronic acids were converted to 6,6'-dideuuterated neutral sugars via methanolysis and sodium borodeuteride ($NaBD_4$) reduction before permethylation. Samples were hydrolyzed using 4M trifluoroacetic acid at 100 °C for 4 hours, then reduced overnight with freshly prepared $NaBD_4$, and peracetylated with 0.5 mL of pure TFA and 2.5 mL of acetic anhydride at 60 °C for 1 hour. Following liquid-liquid extraction, the organic phase was passed through anhydrous Na_2SO_4, evaporated, and redissolved in ethyl acetate for GC-MS analysis.

PMAAs were analyzed using an Agilent 7890A-5977B GC-MS system equipped with a Supelco SP-2380 column and helium carrier gas (0.8 mL/min). Samples (1 µL) were injected at a 10:1 split ratio. Identification of partially methylated alditol acetates was based on their relative retention time in relation to the internal standard and fragmentation mass spectra. The results of linkage analysis are expressed as relative percentage of peak area of each identified structural unit in relation to their sum and accounting response factors in relation to myoinositol. Polysaccharide composition (Table 1) was estimated in accordance with Petrolino et al. [26].

Plasmid construction and protein expression

Expression plasmids were generated by one-pot Golden Gate assembly of pET28 backbone and amplified gene harboring *BsaI* restriction site with appropriate overhang (as described in Fig 1.B) using NEBridge Ligase Master Mix and *BsaI*-HFv2 (New England Biolabs). Assembly reactions were subjected to 20 cycles of 37 °C for 3 min and 16 °C for 2 min, followed by 60

°C for 5 min. *E. coli* XL1-Blue cells were transformed and correct assemblies were verified by restriction digestion and Sanger sequencing using T7 promoter and terminator primers.

All recombinant proteins used in this study were expressed in the *E. coli* BL21(DE3) strain. Six assemblies (XCC-A; XCC-B; XCC-E; XCC-F; XCC-K and XCC-L) along with the free enzyme counterparts were produced as described previously [21]. Cells were grown in LB (for *At*Cel8A and *At*Xyn11A) or ZYP-5052 medium (*At*Cel9R) supplemented with the appropriate antibiotic. For LB expression, cultures were grown at 37 °C to $OD_{600}\approx 0.6$, induced with 1 mM IPTG (final concentration) and incubated for 4 h. For ZYP-5052 expression, cultures were incubated at 25 °C for 24 h. Expression conditions were identical for all chimeric proteins, except for Cel9R-containing constructs, which were expressed in auto-inducible ZYP-5052 medium [27]. Cells were harvested by centrifugation and resuspended in 50 mM Tris-HCl pH 8.0, 300 mM NaCl, supplemented with protease inhibitors (Sigma FAST) to an OD_{600}equivalent of 40–80 before storage at −80 °C.

Complex assembly and protein purification

Cell-free extracts (CFE) were prepared by sonication (5 kJ delivered in 2.5 min) and centrifugation at 50,000 × *g* for 30 min. His_6-tagged free enzymes were purified by cobalt-based immobilized metal affinity chromatography (IMAC) using 50 mM Tris-HCl pH 8.0, 300 mM NaCl for equilibration and washing, and 100 mM imidazole for elution. Purified proteins were desalted into 50 mM Tris-HCl pH 8.0, 150 mM NaCl with a PD10 desalting column (Cytivia). Complexes were formed by mixing two different CFE of interest: (1) untagged In-containing chimeras, and (2) His_6-tagged Jo partners in at least a twofold molar excess (assessed using SDS-PAGE). Mixtures were incubated for 1.5 h at 20-23 °C with mild agitation, followed by 30 min at 50 °C. Soluble complexes were recovered by centrifugation and purified by IMAC as previously described. When required, complexes were polished by size-exclusion

chromatography on a Superdex 200 10/300 GL column (Cytivia) in 50 mM Tris-HCl pH 8.0, 150 mM NaCl at 1 $mL.min^{-1}$.

Protein purity was assessed by SDS-PAGE using Stain-Free precast gel (Biorad) and revealed by UV or Coomassie staining, depending on protein composition (Fig.Sup1) displays SDS-PAGE of all proteins and assemblies used in this study). Protein concentrations were determined by measuring absorbance at 280 nm using theoretical calculated extinction coefficients (ExPASy ProtParam, https://web.expasy.org/protparam/).

Enzymatic assays

All enzymatic assays were performed at 50 °C in triplicate. The substrates used in this study were carboxymethylcellulose (P-CMC4M, Megazyme), wheat arabinoxylan (WAX RS from Megazyme, at 1% w/v), Destarched wheat bran (dWB, at 2% w/v) and wheat straw (WS, at 2% w/v) [28]. dWB and WS were soaked in 10 mM sodium acetate buffer (pH 6.0) and preincubated overnight at 50 °C before assay. Released reducing sugars were quantified using the 3,5-Dinitrosalicylic acid (DNS) assay [29], using 50 µl of sample mixed with 50 µl of DNS. Enzymatic reactions were performed in 20 mM sodium acetate buffer (pH 6.0) at 50 °C using 25 nM enzyme for CMC and 10 nM for WAX for 15 and 10 min respectively, under constant agitation (Thermomix, Eppendorf) at 1000 RPM. dWB and WS were assessed with 250 nM and 500 nM of enzymes, respectively. Kinetics were monitored for 24 h (0 h, 1 h, 2 h, 4 h, 8 h and 24 h) under constant agitation. Glucose (0.5-11.1 mM) or xylose (0.67-13.3 mM) were used as calibration standards.

MALDI-TOF spectrometry and data normalization.

In parallel to DNS assay, 50 µl of samples from dWB and WS were collected (0 h, 1 h, 2 h, 4 h, 8 h and 24 h) and diluted 4 times in ultrapure water to reduce salt concentration prior to

analysis by MALDI-TOF mass spectrometry at the BIBS platform (BIA Nantes, France). dWB samples were once again diluted 10 times due to their high concentration of oligosaccharides. Each independent enzymatic experiment was analyzed in technical triplicate. Spectra were normalized using an internal standard added in aliquots (0,5 µg/mL acetylated chitosan pentooligosaccharide) according to:

$$I_{norm} = \frac{I_{peak}}{I_{IS}} * \overline{I_{IS}}$$

where I_{peak} is the oligosaccharide peak intensity, I_{IS} is the internal-standard intensity, and $\overline{I_{IS}}$ is the mean internal-standard intensity across all measurements. This normalization preserves the original order of magnitude while correcting for variations in internal-standard intensity. MALDI-TOF-MS analysis was performed on a Rapiflex TissueTyper mass spectrometer (Bruker Daltonics, Bremen, Germany), equipped with a Smartbeam II Laser (355 nm, 10 kHz) in reflector mode. Samples were directly mixed on a polished steel MALDI target plate with a solution of ionic liquid matrix DMA–DHB (2,5-dihydroxybenzoic acid 100 $mg.mL^{-1}$ in H_2O/ACN (50:50 vol/vol) with an addition of 0.2% of N,N-dimethylaniline) [30]. Spectra were recorded in reflectron mode in positive polarity (m/z range of 300–3250) at fixed laser intensity. Acquisitions were performed using FlexControl (Bruker Daltonics, Billerica, MA, USA) and processing using FlexAnalysis (Bruker Daltonics, Billerica, MA, USA).

Statistical analysis and data visualization

Specific activities on CMC and WAX were analyzed by one-way ANOVA followed by Dunnett's multiple-comparison test on prism. The free enzyme counterpart was used as reference and statistical significance was set at $p < 0.05$. Principal component analysis (PCA), was performed on normalized MALDI-TOF data after centering and scaling the data on RStudio (v. 3.6.0). Hierarchical clustering was performed using Euclidean distance and Ward.D2 linkage. Pairwise correlations between interdomain distances and normalized

oligosaccharides were calculated using Kendall's rank correlation coefficient (τ) using two-sided tests with $p < 0.05$ with no multiple-testing correction.

Data visualization was performed using GraphPad Prism and RStudio. Heatmaps, PCA and pairwise correlation plots were generated using the Factoextra, pheatmap, and corrplot R packages. R scripts were developed and debugging with the assistance of Claude (Anthropic, Claude Sonnet 4.5).

SAXS data collection and primary analysis

SAXS measurements were previously reported for six tri-enzymatic assemblies (XCC-A, -B, -E, -F, -K and -L) [21]. The six additional assemblies (XCC-C, -D, -G, -H, -I and -J) were characterized in the present study using the same experimental protocol. Prior to SAXS measurements, protein samples were purified by size-exclusion chromatography (Superdex 200 10/300 GL, Cytiva). A fraction collected near the apex of the elution peak was retained for SAXS analysis at low protein concentration, while the remaining peak fractions were pooled and concentrated using Vivaspin centrifugal filters to approximately 10–15 mg/mL. SAXS measurements were performed on a XEUSS 2.0 system (Xenocs, France) bench SAXS with a Genix3D copper source ($\lambda = 1.54$ Å, 8 keV). Proteins (10 - 15 mg/mL; concentrated using Vivaspin filters) were analyzed under static measurement at 18 °C. Data were collected on a Pilatus 1 M detector (1.216 m distance, $q = 0.005–0.5$ Å^{-1}), averaging 12 exposures of 600 s each. Buffer subtraction yielded scattering intensities I(q), and data reduction, including radial averaging and normalization, were performed using XSACT (Xenocs) following standard procedures to obtain the scattering intensity profile I(q). Scattering profiles from both injections (apex and concentrated fractions) were merged over a common q-range where both profiles showed consistent scattering behavior after scaling. The low-q region was preserved from the

dilute fraction to minimize potential interparticle interaction effects, while the concentrated sample was used to improve the signal-to-noise ratio at higher q, as previously described. Primary SAXS data reduction and analysis were carried out using PRIMUS from the ATSAS [31] and BioXTAS RAW 2 [32]. The radius of gyration (Rg) was determined from Guinier analysis, while the pair-distance distribution function P(r) and the maximum particle dimension (Dmax) were obtained by indirect Fourier transformation. Molecular weight (MW) estimates were derived from the forward scattering intensity I(0) following absolute intensity calibration using water as a secondary standard. All atom models were reconstructed using MODELLER [33] based on available structures and predicted models (PDB:1KWF, 7UNP, 1UXX, 5MKC and AF-O52780-F1 corresponding to AtCel8A, AtCel9R, AtCBM6, Jo-In and AtXyn11A, respectively. The relative spatial organization of the domains was subsequently optimized against the experimental SAXS profiles using the DADIMODO web server as previously (https://dadimodo.synchrotron-soleil.fr) [34]. Interdomain distances (catalytic amino acid) were calculated based on 5 independent models as reported previously [21].

Results

Design and construction of a library of Jo-In-based tri-enzymatic assemblies

A combined cellulase-xylanase system for the simultaneous depolymerization of the two main polysaccharides present in our samples was designed by selecting three enzymes from the *Acetivibrio thermocellus* cellulosomal arsenal: endoglucanase *At*Cel8A, processive endoglucanase *At*Cel9R, and xylanase *At*Xyn11A (Fig. 1A). These enzymes are highly abundant in the cellulosome during lignocellulose degradation, notably of wheat straw [35,36], acting on two intimately associated polysaccharides within PCW, amorphous and crystalline cellulose for *At*Cel8A and *At*Cel9R, respectively [37] and *At*Xyn11A on xylan [24,38]. Compared to their wild-type counterparts, the recombinant enzymes were designed to lack their

associated dockerin and linker sequences to minimize flexibility within the resulting assemblies but retained their carbohydrate-binding modules (CBMs), namely CBM3c in *At*Cel9R for its role in processivity [39] and CBM6 in *At*Xyn11A to preserve full catalytic activity [40]. All three enzymes were fitted with a His Tag at their C terminus.

To associate these three enzymes covalently, we relied on the Jo–In protein pair [19]. Jo and In are two small proteins (10 and 16 kDa) that recognize each other specifically and spontaneously form a covalent isopeptide bond. Since the N and C termini of both proteins are free, an enzyme can be fused at each of these four positions [20,28].Three enzymes can therefore be distributed over four fusion sites. As the Jo–In core is rigid [21], each combination should correspond to a distinct spatial topology, associating three different enzymes through the systematic construction of a series of plasmids, using a cloning platform based on a pET28 backbone (Fig. 1B). Inspired by the MoClo architecture [41], the platform assembles the enzyme-encoding and Jo/In-encoding sequences as interchangeable DNA modules by directional BsaI Golden Gate assembly, using four-nucleotide overhangs (Fig. 1B). Note that this strategy has been recently developed for DCs, known as the VersaTile approach [42]. This design thus provides a flexible framework for generating different chimeric proteins while maintaining a common genetic architecture across constructs. Using this platform, we designed a series of partial complexes: Jo-fusion proteins carrying two enzymes, one at each terminus, and In-fusion proteins carrying the third enzyme. Mixing the two partial complexes triggers the autocatalytic covalent association of Jo and In and yields pure tri-enzymatic assemblies, ready for biochemical and structural characterization (Fig. 1C). Because of the asymmetrical association of Jo-In, the theoretical design space comprised 24 enzyme arrangements, resulting from the different possible positions and orientations of the 3 enzymes within the assembly. Rather than generating all of them, we selected 12 architectures (XCC-A to XCC-L; Fig. 1C)

297 to systematically sample the spatial environments accessible around the approximately globular
298 Jo-In core.
299
300 **Fusion to Jo and In does not impair the activity of the enzymes on soluble polymers**
301 Once the 12 assemblies had been obtained, the resulting Jo-In complexes were assessed for
302 any potential alteration of enzymatic activity, as fusion to non-native modules can affect
303 enzyme function. [42,43]. Activity measurements on small, diffusible substrates were already
304 available for 6 of the 12 assemblies, and no major loss of activity had been reported [21]. To
305 evaluate all assemblies under more demanding conditions, activity was measured on soluble
306 polymers, i.e., carboxymethyl cellulose (CMC) used as a model cellulosic substrate and wheat
307 arabinoxylan (WAX) as a model hemicellulosic substrate (Fig. 2). Specific activities were
308 determined by measuring the accumulation of reducing sugar by the DNS assay.
309 On CMC, as expected, no activity was detected for *At*Xyn11A, whereas specific activities
310 of $8.5 \pm 0.8 \times 10^3$ and $2.0 \pm 0.4 \times 10^3$ U.µmol^{-1} were measured for *At*Cel8A and *At*Cel9R,
311 respectively. The higher activity of *At*Cel8A is consistent with previous observations on this
312 substrate [37]. When combined as free enzymes, an activity of $9.6 \pm 1.1 \times 10^3$ U.µmol^{-1} was
313 measured, indicating that no clear synergistic or antagonistic effect could be observed under
314 the tested conditions. No statistically significant differences were found between the free-
315 enzyme reference and any of the 12 assemblies.
316 On WAX, no activity was detected for either AtCel8A or AtCel9R, which is consistent with
317 their reported substrate specificities. For 6 of the 12 assemblies (XCC-A, B, -D, -J, -K and -L),
318 activity did not differ significantly from the free-enzyme reference, whereas reduced activity
319 was observed for the others (XCC-C, -E, -F, -G, -H and -I). Compared to the free-enzyme
320 reference, XCC-F and XCC-G exhibited decreases of approximately 17%, while XCC-H and
321 XCC-I showed the largest reductions, of approximately 24%. These differences did not follow

a clear relationship with the position of Xyn11A within the Jo-In core, and therefore could not be directly attributed to its position in the assembly.

Overall, enzyme activity on soluble polymers was largely preserved by the Jo-In scaffold, with no significant effect detected on CMC and only limited effects observed on WAX.

Assembly topology affects the degradation of insoluble lignocellulosic substrates

To determine whether different enzyme layouts affect the degradation of insoluble lignocellulosic substrates, the 12 assemblies were evaluated on destarched wheat bran (dWB) and wheat straw (WS), using the corresponding free enzymes as references. Glycosidic linkage analysis revealed distinct polysaccharide compositions between the two substrates (Table 1).

dWB was mainly composed of arabinoxylan (63.9 % ± 1.1) and cellulose (23.0 % ± 0.9), whereas WS contain more cellulose (49.5 % ± 1.7) and less arabinoxylan (41.3 % ± 0.9). In both substrates, xyloglucan was the second most abundant hemicellulose, and pectins represented a minor fraction. These differences provide distinct polysaccharide environments for enzymatic degradation. Substrate degradation was monitored by measuring the accumulation of reducing sugar by the 12 assemblies on dWB and WS compared with the free-enzyme control (Fig. 3; detailed kinetics are provided in Supplementary Fig. Sup2).

For both substrates, the initial degradation rate was higher with the free enzymes than with the assemblies, particularly during the first hours of incubation.

On dWB, the reducing sugar release of all assemblies progressively reached (t = 4h) or exceeded that of the free-enzyme control (t = 8h), and five assemblies showed significantly higher values after 24 h. The highest final releases were obtained with XCC-G and XCC-H (27.1 mM ± 1.9 and 28.6 mM ± 1.5, respectively) (Fig. Sup2-A), corresponding to increases of approximately 43% and 51% relative to the free-enzyme control (19.0 mM ± 0.7). On WS, in contrast, the free-enzyme control reached 6.2 mM ± 0.7 after 24 h, while XCC-L, the highest-

producing assembly, reached 6.5 mM ± 0.4. Only three assemblies, XCC-D, XCC-F and XCC-L, reached final values comparable to the free-enzyme control (Fig. Sup2-B). No plateau was observed on WS, suggesting that degradation was still progressing after 24 h. Marked differences were also observed between assemblies on this substrate. Assembly XCC-A released 3.3 mM ± 0.0 compared with 6.5 ± 0.4 mM for XCC-L, a two-fold difference.

Activity on soluble substrates did not consistently align with the performance on insoluble biomass. For example, XCC-G and XCC-H, which showed the highest degradation of dWB were among the least active assemblies on CMC and WAX. Conversely, XCC-E and XCC-F, whose activities on soluble substrates were largely unaffected by assembly, were among the least efficient assemblies on dWB.

Thus, DNS analysis revealed distinct degradation kinetics between free enzymes and multienzyme assemblies, as well as marked differences between assembly topologies. Since quantification by DNS provides only a global measurement of hydrolytic events, the same samples were simultaneously subjected to mass spectrometry analyses of the soluble oligosaccharides to distinguish degradation product profiles.

Mass spectrometry reveals topology-dependent soluble oligosaccharide profiles on wheat straw

Mass spectrometry analysis was used to characterize the soluble oligosaccharides released during dWB and WS degradation. A total of 79 distinct oligosaccharides were identified in dWB and 60 in WS, and classified according to their polysaccharide origin (pentoses, hexaoses) and chemical modifications (acetylated or feruloylated) (Supplementary Table Sup1).

On dWB, the products detected derived exclusively from cellulose and arabinoxylan, consistent with the substrate composition and enzyme activities. Cellulose-derived

oligosaccharides reached DP12, whereas arabinoxylan-derived products reached DP24 and included acetylated and feruloylated species. In WS, oligosaccharides derived from cellulose, arabinoxylan, glucuronoxylan and xyloglucan were released with generally lower maximum DPs than in dWB (DP4 for cellulose and DP11 for xylan-derived products). Interestingly, xyloglucan-derived products were detected only in WS, although the xyloglucan content was similar in the two substrates (Table 1). Xylobiose and several substituted xylo-oligosaccharides, including acetylated and feruloylated species, were common to both substrates and may represent oligosaccharides that are not hydrolysable by *At*Xyn11A. Overall, the same enzyme set generated distinct oligosaccharide profiles depending on the biomass substrate.

The profiles obtained with the buffer control, the free enzymes and the multienzyme assemblies were therefore compared globally, grouping oligosaccharides into five and nine classes for dWB and WS, respectively (Fig. 4).

On dWB, signals were detected in the buffer control, particularly for the pentose and hexose families, probably reflecting pre-existing soluble oligosaccharides in this substrate. In contrast, on WS, the buffer control showed little to no signal in any of the oligosaccharide families. For both substrates, the free-enzyme control generally displayed higher normalized intensities than the assemblies in several product families. Nevertheless, differences were also observed between assemblies, suggesting that their spatial organization may influence the kinetics and/or distribution of product release. Particularly, XCC-C seems to distinguished on dWB with a different profile compare to XCC-G and XCC-H. Therefore, principal component analyses (PCA) were performed to reduce the dimensionality of the dataset and identify the major sources of variation among the assemblies.

For dWB, the PCA at t = 0 revealed a separation between two groups of assemblies. Since no enzymatic degradation had yet occurred, this separation may reflect differences in the solubilization of pre-existing oligosaccharides during the pre-equilibration period

(Supplementary Fig. Sup3). It was not maintained during the degradation, and no consistent clustering of the assemblies was observed at later time points. In contrast, the PCA of the WS (Supplementary Fig. Sup4) revealed a clearer separation between assemblies from t = 2 h onwards, with two distinct groups particularly apparent at t = 8 and 24 h (Fig. 5A). This indicates that the assembly configuration influenced the overall distribution of soluble degradation products during WS degradation.

To characterize these differences further, hierarchical clustering was performed on the oligosaccharide profiles at t = 8, based on the combined profile of xylan-, xyloglucan- and cellulose-derived products (Fig. 5B).

Two groups of assemblies were identified: XCC-F, -E, -G, -A, -B and -H on the one hand, and XCC-C, -K, -I, -J, -D and -L on the other. Qualitatively, the first group showed a higher release of cellulose-derived oligosaccharides ("cellulolytic" assemblies), and the second a higher release of xylan-derived oligosaccharides ("xylanolytic" assemblies). These analyses indicate that, on WS only, different assembly layouts can generate distinct degradation-product profiles. However, no clear relationship could be established between these profiles and the specific enzyme arrangements within the Jo-In scaffold.

SAXS-based models reveal distinct spatial organizations, but interdomain distances alone do not explain activity

Since the DNS and mass spectrometry analyses revealed differences in degradation behavior among the multienzyme assemblies without a clear link to the fusion arrangements, SAXS-constrained structural models were examined to investigate whether variations in interdomain distances could contribute to the observed degradation profiles.

SAXS data were obtained for all 12 multienzyme assemblies. Six assemblies (XCC-A, -B, -E, -F, -K and -L) had previously been characterized by SAXS and integrative structural

422 modelling [21]. The remaining six assemblies (XCC-C, -D, -G, -H, -I and -J) were
423 experimentally characterized by SAXS and modelled using DADIMODO in the present study.
424 Although the scattering curves displayed similar overall shapes (Fig. 6A), differences
425 were observed mainly in the $q \approx 0.05 - 0.15$ Å^{-1} region, as shown by the residuals in the inset,
426 suggesting differences in the organization of the assemblies. Nevertheless, the radius of
427 gyration (Rg 53.5 Å ± 0.9) and the maximum dimension (184 Å ± 8) (Fig. 6B) remained similar
428 across configurations.
429 The interdomain distances estimated from the SAXS-constrained models ranged from
430 approximately 30 to 150 Å, depending on the domain pair (Fig. 6C, Table Sup2). The XynCD-
431 XynCBM distance showed comparatively limited variation across the assemblies, providing an
432 internal reference for assessing the consistency of the structural models. Similar distances were
433 obtained for different assembly configurations, indicating that interdomain distances alone may
434 not uniquely define the spatial organization of an assembly. To investigate whether these
435 structural parameters were associated with enzymatic performance, Kendall correlations were
436 calculated between interdomain distances and soluble oligosaccharide profiles measured after
437 2, 4, 8 and 24 h of WS degradation (Fig. 6D). No significant correlations ($p > 0.05$) and no clear
438 pattern were detected between the estimated pairwise interdomain distances and
439 oligosaccharides produced.
440 Overall, SAXS analysis revealed differences in the spatial organization of the
441 assemblies, but the estimated pairwise interdomain distances alone were insufficient to explain
442 the observed differences in degradation behavior.
443
444 **Discussion**
445 Whether the spatial organization of enzymes governs the efficiency of multienzyme
446 complexes such as the cellulosome is difficult to establish, because the flexibility of these

assemblies blurs the relationship between enzyme position, distance and function [44,45]. In this study, three cellulosomal GHs were held in 12 topologies by the Jo–In pair, so that topology could be varied systematically while flexibility was largely reduced.

On soluble polysaccharides, the Jo-In assemblies largely retained the activity of their constituent enzymes. No significant effect was observed on CMC, and significant decrease on WAX were limited (up to 24%). This contrasts with previous reports in which dockerin or other fusion elements reduced the activity of individual enzymes (TfCel5A [42,43]. Differences observed on insoluble biomass are therefore unlikely to result from a general loss of activity after fusion. Soluble-substrate activity was also a poor predictor of biomass degradation. XCC-G and XCC-H were among the most efficient assemblies on dWB despite relatively low soluble activities on WAX, whereas XCC-E and XCC-F, whose soluble activities were largely preserved, were among the least efficient on dWB. The consequence of enzyme organization of these assemblies thus appears mainly under the accessibility constraints imposed by the insoluble substrates.

Free enzymes initially degraded both substrates faster than the assemblies. On dWB, however, the free enzymes reached a plateau, while all assemblies progressively reached or exceeded the free-enzyme control, some releasing 40–50% more reducing sugars after 24 h. On WS, only three assemblies reached the level of the free enzyme and none showed a clear plateau at 24 h. This behavior differs from DCs combining cellulases and xylanases, which released three to five times more soluble sugars from wheat straw than free enzymes within the first hours [9,13,46].

Several factors may contribute to the above results. First, our assemblies lacked the cellulose-binding CBM3a within the chimeric scaffoldin as usually found in DCs [47], which

anchor the multienzyme complex to insoluble cellulose and thereby increase the local enzyme concentration at the cellulose surface [48]. Second, the Jo-In assemblies are larger than the individual enzymes (approximately 175 kDa versus 38–72 kDa), which could slow diffusion through the heterogeneous and porous PCW matrix. Size alone, however, is unlikely to explain the differences, since larger DCs have been reported to show higher initial rates [9,10,13,46]. Third, the enzymatic composition may matter. Many DCs contain processive exoglucanases such as Cel48S from *A. thermocellus*, or Cel48F from *Clostridium cellulolyticum* [8], which exploit chain ends generated by endoglucanases [49] and are reported to move along cellulose microfibrils, with AFM studies suggesting that cellulosomes can progressively "dig into" cellulose, while free enzymes mainly remain at the substrate surface [49,50]. In our assemblies, *At*Cel9R is only moderately processive, and its fixed orientation in a locked topology could increase non-productive interactions with the heterogeneous surface. Finally, our approach combines three endo-acting enzymes, two of which target the same polymer, and endoglucanase–endoglucanase synergy is generally weak [51]. The weaker initial rates of the assemblies cannot be accounted for by this last factor alone, since the free control also contains both endoglucanases.

Mass spectrometry showed that substrate architecture affects product profiles more than enzyme layout. The same enzyme set generated distinct oligosaccharide families from dWB and WS, and xyloglucan-derived products were detected in WS but not dWB, although both contain xyloglucan in the same proportion. This suggests that accessibility may contribute more than bulk composition to the order in which different polysaccharides are attacked, consistent with the heterogeneous organization of PCW polysaccharides, where xylan and other matrix polymers restrict access to cellulose [52]. Free enzymes and Jo-In assemblies produced largely similar product families, with differences mainly in kinetics. This differs from DC studies in

497 which enzyme proximity shifted toward shorter fragments [12] and from observations made
498 using Jo-In assemblies on soluble xylan [20]. Layout-dependent differences in products were
499 detected mainly on WS, where two groups of assemblies differed in the relative release of
500 cellulose- and xylan-derived oligosaccharides. Topology thus appears to influence the extent
501 and kinetics of degradation more than the identity of the products.

502 SAXS showed similar global dimensions across assemblies (Rg ≈ 53.5 Å, Dmax ≈ 184
503 Å), and no significant relationship was found between estimated interdomain distances and
504 degradation profiles. This absence of correlation does not mean that proximity is irrelevant,
505 since pairwise distances describe only part of the structure. Different topologies can give similar
506 interdomain distances while differing in enzyme orientation, solvent exposure, steric
507 accessibility, or overall compaction. SAXS also describes the assemblies in solution, not at the
508 solid–liquid interface, and previous AFM observations of bienzymatic Jo-In assemblies suggest
509 limited deformability on highly oriented pyrolytic graphite (HOPG) [21]. The Jo-In assemblies
510 are markedly more compact that a tri-enzymatic DCs with Cel8A and Cel9R (approximately
511 180 kDa, Rg ≈55 Å and Dmax ≈180 Å, *versus* 300 kDa, 91 Å, and 305 Å, respectively [53],
512 while cellulosomes and DCs can adopt a broad range of conformations, from compact to highly
513 extended [50,54–56], and flexibility has been proposed to favor productive engagement [7,57].
514 These observations lead us to hypothesize that flexibility and colocalization are two partly
515 separable parameters. Restricted conformational freedom could weaken initial degradation
516 rates, whereas colocalization could influence cumulative degradation and final yields as
517 accessible sites accumulate.

518

519 **Conclusions**

520 Fixing the position of the enzymes did not render their performance on insoluble
521 biomass predictable. Topology had little effect on activity on soluble substrates, but it affected

522 the kinetics of biomass degradation and, to a lesser extent, the distribution of released products.
523 This could enable the selection of layout depending on the distribution one wants to achieve at
524 a certain stage of the reaction. However, these effects were explained neither by soluble-
525 substrate activity nor by pairwise interdomain distances, and probably depend on factors such
526 as enzyme orientation, accessibility to the substrate and conformational freedom. We therefore
527 hypothesize that flexibility and colocalization are two partly separable parameters: restricted
528 flexibility could account for the lower initial rates, whereas colocalization could influence
529 cumulative degradation. These findings provide a framework and a method to dissect the
530 relationship between structure and function in multienzyme complexes, and highlight which
531 parameters should be considered to design programmable enzyme complexes for biomass
532 conversion.

533

534 **Abbreviations**

AX	Arabinoxylan
CBM	Carbohydrate-binding module
CMC	Carboxymethyl cellulose
DC	Designer cellulosome
Dmax	maximum particle size
DNS	3,5-dinitrosalicylic acid
DP	Degree of polymerization
dWB	Destarched wheat bran
GH	Glycoside hydrolase
PCA	Principal component analysis
PCW	Plant cell wall
Rg	Radius of gyration
SAXS	Small-angle X-ray scattering
WAX	Wheat arabinoxylan
WS	Wheat straw

535

536 **Supplementary information**

537 **Acknowledgments**:

We thank the PICT-ICEO facility dedicated to enzyme screening and discovery, part of the Integrated Screening Platform of Toulouse (PICT, IBiSA) for providing access to protein purification equipment. PICT-ICEO is a member of IBISBA-FR [58], the French node of the European research infrastructure, IBISBA (www.ibisba.eu). We thank the FERMAT platform (Federation of Research and Training in Materials and Processes, Toulouse, France) for access to SAXS instrumentation at LGC, Toulouse. We are deeply grateful to Dr. Sandrine Laguerre and Marvin Ramos-Hue (TBI INSA Toulouse 135 avenue de Rangueil 31077 Toulouse cedex 04, France) for their expertise in statistics and their invaluable help and advice in analyzing our data. We thank Dr. Xiaohui Xing for his advice on glycomics methods and analysis.

Authors' contributions

IPL wrote the original manuscript, designed and performed the experiments and analyzed the data. BA designed the mass experiments and performed the mass analysis and analyzed the data. MF analyzed the mass data, HR analyzed the mass data, DR designed the mass experiments and analyzed the mass data. DWA, BB, and CD conducted the glycosidic linkage analysis. DWA and CD acquired the funding for the glycosidic linkage analysis. SM, IM and EAB provide expertise in multienzymatic complexes. PR analyzed the SAXS data. CD supervised the work and analyzed the data. CM wrote the original manuscript, conceptualized the work, acquired the funding, analyzed the data and supervised the work. CD and CM received funding for an international mobility in Canada from INRAE and INSA Toulouse. All authors read, contribute in writing the manuscript and approved the final manuscript.

Funding:

This study has been funded by the National Research Agency under grant number ANR-21-E43-0024 CONCERTO and was supported by the ‘Investissements d’Avenir’ program ANR-

18-EURE-0021. Glycosidic linkage analysis was funded through the AAFC-INRAE project: J-003614.

Data availability

The data that support the findings of this study are openly available in SASBDB at https://www.sasbdb.org/. https://www.sasbdb.org/project/2840/ [21] and https://www.sasbdb.org/project/update/3011/ for the current study.

SAXS model used to calculate interdomains distances are available on demand.

MALDI-TOF spectrometry data are available on demand.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interest

The authors declare no competing interest

References

1. Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, et al. Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. Science. 2007;315:804–7. https://doi.org/10.1126/science.1137016

2. Artzi L, Bayer EA, Moraïs S. Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. Nat Rev Microbiol. 2017;15:83–95. https://doi.org/10.1038/nrmicro.2016.164

3. Fontes CMGA, Gilbert HJ. Cellulosomes: Highly Efficient Nanomachines Designed to Deconstruct Plant Cell Wall Complex Carbohydrates. Annu Rev Biochem. 2010;79:655–81. https://doi.org/10.1146/annurev-biochem-091208-085603

4. Artzi L, Morag E, Barak Y, Lamed R, Bayer EA. Clostridium clariflavum: Key Cellulosome Players Are Revealed by Proteomic Analysis. Lovley DR, editor. mBio. 2015;6:e00411-15. https://doi.org/10.1128/mBio.00411-15

5. Gal L, Pages S, Gaudin C, Belaich A, Reverbel-Leroy C, Tardif C, et al. Characterization of the cellulolytic complex (cellulosome) produced by Clostridium cellulolyticum. Appl Environ Microbiol. 1997;63:903–9. https://doi.org/10.1128/aem.63.3.903-909.1997

6. Mayer F, Coughlan MP, Mori Y, Ljungdahl LG. Macromolecular Organization of the Cellulolytic Enzyme Complex of *Clostridium thermocellum* as Revealed by Electron Microscopy. Appl Environ Microbiol. 1987;53:2785–92. https://doi.org/10.1128/aem.53.12.2785-2792.1987

7. Hammel M, Fierobe H-P, Czjzek M, Kurkal V, Smith JC, Bayer EA, et al. Structural Basis of Cellulosome Efficiency Explored by Small Angle X-ray Scattering. J Biol Chem. 2005;280:38562–8. https://doi.org/10.1074/jbc.M503168200

8. Fierobe H-P, Mechaly A, Tardif C, Belaich A, Lamed R, Shoham Y, et al. Design and Production of Active Cellulosome Chimeras. J Biol Chem. 2001;276:21257–61. https://doi.org/10.1074/jbc.M102082200

9. Fierobe H-P, Mingardon F, Mechaly A, Bélaïch A, Rincon MT, Pagès S, et al. Action of Designer Cellulosomes on Homogeneous Versus Complex Substrates. J Biol Chem. 2005;280:16325–34. https://doi.org/10.1074/jbc.M414449200

10. Moraïs S, Barak Y, Caspi J, Hadar Y, Lamed R, Shoham Y, et al. Cellulase-Xylanase Synergy in Designer Cellulosomes for Enhanced Degradation of a Complex Cellulosic Substrate. Ausubel FM, editor. mBio. 2010;1:e00285-10. https://doi.org/10.1128/mBio.00285-10

11. Lamote B, Vanderstraeten J, De Ridder N, Meert K, Boerjan W, Briers Y. Bio-inspired designer cellulosomes show strongest synergy on industrial substrates under natural-like conditions. J Ind Microbiol Biotechnol. 2026;53:kuag004. https://doi.org/10.1093/jimb/kuag004

12. Molinier A-L, Nouailler M, Valette O, Tardif C, Receveur-Bréchot V, Fierobe H-P. Synergy, Structure and Conformational Flexibility of Hybrid Cellulosomes Displaying Various Inter-cohesins Linkers. J Mol Biol. 2011;405:143–57. https://doi.org/10.1016/j.jmb.2010.10.013

13. Vazana Y, Barak Y, Unger T, Peleg Y, Shamshoum M, Ben-Yehezkel T, et al. A synthetic biology approach for evaluating the functional contribution of designer cellulosome components to deconstruction of cellulosic substrates. Biotechnol Biofuels. 2013;6:182. https://doi.org/10.1186/1754-6834-6-182

14. Stern J, Kahn A, Vazana Y, Shamshoum M, Moraïs S, Lamed R, et al. Significance of Relative Position of Cellulases in Designer Cellulosomes for Optimized Cellulolysis. Johnson EA, editor. PLOS ONE. 2015;10:e0127326. https://doi.org/10.1371/journal.pone.0127326

15. Elias M, De Paepe B, Vanderstraeten J, Lamote B, De Mey M, Briers Y. A Scaffoldomics Platform for Modular In Vivo Enzyme Colocalisation and Its Application to Naringenin Biosynthesis. Microb Biotechnol. 2026;19:e70328. https://doi.org/10.1111/1751-7915.70328

16. Fonseca MJMD, Vanderstraeten J, De Coninck T, Vlyminck S, Briers Y. A funnel-type pipeline for accelerating discovery of tailored designer xylanosomes for efficient arabinoxylan valorization. Int J Biol Macromol. 2026;340:149852. https://doi.org/10.1016/j.ijbiomac.2025.149852

631 17. Artico M, Roux C, Peruch F, Mingotaud A-F, Montanier CY. Grafting of proteins onto polymeric
632 surfaces: A synthesis and characterization challenge. Biotechnol Adv. 2023;64:108106.
633 https://doi.org/10.1016/j.biotechadv.2023.108106

634 18. Wang S, Li R, Xie Y, Zhang H, Huang X, Chen S, et al. Recent Progress on Enzyme Immobilization:
635 Materials, Strategies, and Applications. Food Bioeng. 2026;5:220–43.
636 https://doi.org/10.1002/fbe2.70056

637 19. Bonnet J, Cartannaz J, Tourcier G, Contreras-Martel C, Kleman JP, Morlot C, et al. Autocatalytic
638 association of proteins by covalent bond formation: a Bio Molecular Welding toolbox derived from a
639 bacterial adhesin. Sci Rep. 2017;7:43564. https://doi.org/10.1038/srep43564

640 20. Enjalbert T, De La Mare M, Roblin P, Badruna L, Vernet T, Dumon C, et al. Characterisation of the
641 Effect of the Spatial Organisation of Hemicellulases on the Hydrolysis of Plant Biomass Polymer. Int J
642 Mol Sci. 2020;21:4360. https://doi.org/10.3390/ijms21124360

643 21. Pardo Larrabeiti I, Eibinger M, Esque J, Pradeau S, Fort S, Moraïs S, et al. Integrative SAXS and
644 AFM analysis of engineered carbohydrate-active enzyme assemblies with tunable spatial
645 organization. Protein Sci. 2026;35:e70649. https://doi.org/10.1002/pro.70649

646 22. Alzari PM, Ne Souchon H, Dominguez R. The crystal structure of endoglucanase CelA, a family 8
647 glycosyl hydrolase from Clostridium thermocellum. Structure. 1996;4:265–75.
648 https://doi.org/10.1016/S0969-2126(96)00031-7

649 23. Zverlov VV, Schantz N, Schwarz WH. A major new component in the cellulosome of *Clostridium*
650 *thermocellum* is a processive endo-Î²-1,4-glucanase producing cellotetraose. FEMS Microbiol Lett.
651 2005;249:353–8. https://doi.org/10.1016/j.femsle.2005.06.037

652 24. Fernandes AC, Fontes CM, Gilbert HJ, Hazlewood GP, Fernandes TH, Ferreira LM. Homologous
653 xylanases from Clostridium thermocellum: evidence for bi-functional activity, synergism between
654 xylanase catalytic modules and the presence of xylan-binding domains in enzyme complexes.
655 Biochem J. 1999;342 (Pt 1):105–10.

656 25. Bajwa B, Xing X, Serin SC, Hayes M, Terry SA, Gruninger RJ, et al. Characterization of
657 Unfractionated Polysaccharides in Brown Seaweed by Methylation-GC-MS-Based Linkage Analysis.
658 Mar Drugs. 2024;22:464. https://doi.org/10.3390/md22100464

659 26. Pettolino FA, Walsh C, Fincher GB, Bacic A. Determining the polysaccharide composition of plant
660 cell walls. Nat Protoc. 2012;7:1590–607. https://doi.org/10.1038/nprot.2012.081

661 27. Studier FW. Protein production by auto-induction in high-density shaking cultures. Protein Expr
662 Purif. 2005;41:207–34. https://doi.org/10.1016/j.pep.2005.01.016

663 28. Badruna L, Burlat V, Roblin P, Enjalbert T, Lippens G, Venditto I, et al. The Jo-In protein welding
664 system is a relevant tool to create CBM-containing plant cell wall degrading enzymes. New
665 Biotechnol. 2021;65:31–41. https://doi.org/10.1016/j.nbt.2021.07.004

666 29. Miller GL. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. Anal Chem.
667 1959;31:426–8. https://doi.org/10.1021/ac60147a030

668 30. Ropartz D, Bodet P, Przybylski C, Gonnet F, Daniel R, Fer M, et al. Performance evaluation on a
669 wide set of matrix-assisted laser desorption ionization matrices for the detection of oligosaccharides

in a high-throughput mass spectrometric screening of carbohydrate depolymerizing enzymes. Rapid Commun Mass Spectrom. 2011;25:2059–70. https://doi.org/10.1002/rcm.5060

31. Manalastas-Cantos K, Konarev PV, Hajizadeh NR, Kikhney AG, Petoukhov MV, Molodenskiy DS, et al. *ATSAS 3.0* : expanded functionality and new tools for small-angle scattering data analysis. J Appl Crystallogr. 2021;54:343–55. https://doi.org/10.1107/S1600576720013412

32. Hopkins JB. *BioXTAS RAW 2* : new developments for a free open-source program for small-angle scattering data reduction and analysis. J Appl Crystallogr. 2024;57:194–208. https://doi.org/10.1107/S1600576723011019

33. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinforma [Internet]. 2016 [cited 2026 Nov 4];54. https://doi.org/10.1002/cpbi.3

34. Rudenko O, Thureau A, Perez J. Evolutionary refinement of the 3D structure of multi-domain protein complexes from small angle X-ray scattering data. Proc Genet Evol Comput Conf Companion [Internet]. Prague Czech Republic: ACM; 2019 [cited 2026 Nov 4]. p. 401–2. https://doi.org/10.1145/3319619.3322002

35. Yoav S, Barak Y, Shamshoum M, Borovok I, Lamed R, Dassa B, et al. How does cellulosome composition influence deconstruction of lignocellulosic substrates in Clostridium (Ruminiclostridium) thermocellum DSM 1313? Biotechnol Biofuels. 2017;10:222. https://doi.org/10.1186/s13068-017-0909-7

36. Gold ND, Martin VJJ. Global View of the *Clostridium thermocellum* Cellulosome Revealed by Quantitative Proteomic Analysis. J Bacteriol. 2007;189:6787–95. https://doi.org/10.1128/JB.00882-07

37. Leis B, Held C, Bergkemper F, Dennemarck K, Steinbauer R, Reiter A, et al. Comparative characterization of all cellulosomal cellulases from Clostridium thermocellum reveals high diversity in endoglucanase product formation essential for complex activity. Biotechnol Biofuels. 2017;10:240. https://doi.org/10.1186/s13068-017-0928-4

38. Hayashi H, Takehara M, Hattori T, Kimura T, Karita S, Sakka K, et al. Nucleotide sequences of two contiguous and highly homologous xylanase genes xynA and xynB and characterization of XynA from Clostridium thermocellum. Appl Microbiol Biotechnol. 1999;51:348–57. https://doi.org/10.1007/s002530051401

39. Kuch NJ, Kutschke ME, Parker A, Bingman CA, Fox BG. Contribution of calcium ligands in substrate binding and product release in the Acetovibrio thermocellus glycoside hydrolase family 9 cellulase CelR. J Biol Chem. 2023;299:104655. https://doi.org/10.1016/j.jbc.2023.104655

40. Czjzek M, Bolam DN, Mosbah A, Allouch J, Fontes CMGA, Ferreira LMA, et al. The Location of the Ligand-binding Site of Carbohydrate-binding Modules That Have Evolved from a Common Sequence Is Not Conserved. J Biol Chem. 2001;276:48580–7. https://doi.org/10.1074/jbc.M109142200

41. Bird JE, Marles-Wright J, Giachino A. A User's Guide to Golden Gate Cloning Methods and Standards. ACS Synth Biol. 2022;11:3551–63. https://doi.org/10.1021/acssynbio.2c00355

42. Vanderstraeten J, Da Fonseca MJM, De Groote P, Grimon D, Gerstmans H, Kahn A, et al. Combinatorial assembly and optimisation of designer cellulosomes: a galactomannan case study. Biotechnol Biofuels Bioprod. 2022;15:60. https://doi.org/10.1186/s13068-022-02158-2

710 43. Caspi J, Barak Y, Haimovitz R, Irwin D, Lamed R, Wilson DB, et al. Effect of Linker Length and
711 Dockerin Position on Conversion of a *Thermobifida fusca* Endoglucanase to the Cellulosomal Mode.
712 Appl Environ Microbiol. 2009;75:7335–42. https://doi.org/10.1128/AEM.01241-09

713 44. Liao M, Wan Z, Wang L, Bi H, Gong C. Cellulosome engineering as biological macromolecular
714 assembly for lignocellulose deconstruction: Structure, interactions, and functional design. Int J Biol
715 Macromol. 2026;381:154163. https://doi.org/10.1016/j.ijbiomac.2026.154163

716 45. Smith SP, Bayer EA, Czjzek M. Continually emerging mechanistic complexity of the multi-enzyme
717 cellulosome complex. Curr Opin Struct Biol. 2017;44:151–60.
718 https://doi.org/10.1016/j.sbi.2017.03.009

719 46. Moraïs S, Barak Y, Hadar Y, Wilson DB, Shoham Y, Lamed R, et al. Assembly of Xylanases into
720 Designer Cellulosomes Promotes Efficient Hydrolysis of the Xylan Component of a Natural
721 Recalcitrant Cellulosic Substrate. Lovley DR, editor. mBio. 2011;2:e00233-11.
722 https://doi.org/10.1128/mBio.00233-11

723 47. Yaniv O, Morag E, Borovok I, Bayer EA, Lamed R, Frolow F, et al. Structure of a family 3a
724 carbohydrate-binding module from the cellulosomal scaffoldin CipA of *Clostridium thermocellum*
725 with flanking linkers: implications for cellulosome structure. Acta Crystallograph Sect F Struct Biol
726 Cryst Commun. 2013;69:733–7. https://doi.org/10.1107/S174430911301614X

727 48. Yarbrough JM, Hengge NN, Xu Q, Ziegler SJ, Chung D, Huang S, et al. Deconstruction by *C.*
728 *thermocellum* —from microbe mediated to dynamic redistribution of cellulosomes. Life Sci Alliance.
729 2026;9:e202503239. https://doi.org/10.26508/lsa.202503239

730 49. Zajki-Zechmeister K, Eibinger M, Nidetzky B. Enzyme Synergy in Transient Clusters of Endo- and
731 Exocellulase Enables a Multilayer Mode of Processive Depolymerization of Cellulose. ACS Catal.
732 2022;12:10984–94. https://doi.org/10.1021/acscatal.2c02377

733 50. Eibinger M, Ganner T, Plank H, Nidetzky B. A Biological Nanomachine at Work: Watching the
734 Cellulosome Degrade Crystalline Cellulose. ACS Cent Sci. 2020;6:739–46.
735 https://doi.org/10.1021/acscentsci.0c00050

736 51. David H, Vita N, Pagès S, Štravs P, Berlec A, Fierobe H-P, et al. Synergies between cellulosomal
737 and non-cellulosomal bacterial cellulases complexed in designer cellulosomes. Int J Biol Macromol.
738 2025;309:142819. https://doi.org/10.1016/j.ijbiomac.2025.142819

739 52. Tryfona T, Bourdon M, Delgado Marques R, Busse-Wicher M, Vilaplana F, Stott K, et al. Grass
740 xylan structural variation suggests functional specialization and distinctive interaction with cellulose
741 and lignin. Plant J. 2023;113:1004–20. https://doi.org/10.1111/tpj.16096

742 53. Dorival J, Moraïs S, Labourel A, Rozycki B, Cazade P-A, Dabin J, et al. Mapping the deformability of
743 natural and designed cellulosomes in solution. Biotechnol Biofuels Bioprod. 2022;15:68.
744 https://doi.org/10.1186/s13068-022-02165-3

745 54. Barth A, Hendrix J, Fried D, Barak Y, Bayer EA, Lamb DC. Dynamic interactions of type I cohesin
746 modules fine-tune the structure of the cellulosome of *Clostridium thermocellum*. Proc Natl Acad Sci
747 [Internet]. 2018 [cited 2026 Nov 4];115. https://doi.org/10.1073/pnas.1809283115

748 55. Bomble YJ, Beckham GT, Matthews JF, Nimlos MR, Himmel ME, Crowley MF. Modeling the Self-
749 assembly of the Cellulosome Enzyme Complex. J Biol Chem. 2011;286:5614–23.
750 https://doi.org/10.1074/jbc.M110.186031

56. García-Alvarez B, Melero R, Dias FMV, Prates JAM, Fontes CMGA, Smith SP, et al. Molecular Architecture and Structural Transitions of a Clostridium thermocellum Mini-Cellulosome. J Mol Biol. 2011;407:571–80. https://doi.org/10.1016/j.jmb.2011.01.060

57. Czjzek M, Fierobe H-P, Receveur-Bréchot V. Small-Angle X-ray Scattering and Crystallography. Cellulases [Internet]. Elsevier; 2012 [cited 2026 Nov 4]. p. 183–210. https://doi.org/10.1016/B978-0-12-415931-0.00010-0

58. IBISBA FR. Industrial Biotechnology Innovation and Synthetic Biology Accelerator [Internet]. INRAE; 2021 [cited 2026 Nov 6]. https://doi.org/10.15454/08BX-VJ91

Tables

Table 1: Polysaccharide's composition of WS and dWB used in this study

%	Cellulose	Hemicelluloses			Pectins				Other
	-	AX	XG	HM	ARA	RG	AG1	AG2	U.A
dWB	23,0 ± 0,9	63,9 ± 1,1	2,0 ± 0,1	1,2 ± 0,0	2,4 ± 0,5	1,4 ± 0,3	0,7 ± 0,1	0,5 ± 0,1	4,5 ± 0,3
WS	49,5 ± 1,7	41,3 ± 0,9	2,8 ± 0,3	1,2 ± 0,2	0,9 ± 0,1	0,5 ± 0,0	0,2 ± 0,0	0,5 ± 0,1	2,95 ± 0,5

Abbreviations: AX: arabinoxylan; XG: xyloglucan; HM: heteromannan; Ara: arabinan; AG1: arabinogalactan type I; AG2: arabinogalactan type II; RG: rhamnogalacturonan; U.A.: Unassigned linkages.

Figure legend

Figure 1: Schematic representation of the multienzymatic assemblies used in this work. A: Truncated forms of the GHs and CBMs used to generate the assemblies. Numbers indicate the corresponding GH (GH8, GH9, and GH11) or CBM (CBM3c and CBM6) family, with His Tags attached to the C-terminal end. Jo and In modules are shown in grey and yellow, respectively. B: Modular cloning platform adapted to Jo–In for the generation of the chimeric proteins used in this study. The nucleotide sequences of the *BsaI* overhangs used for modular assembly are indicated. C. Schematic representation of the 12 tri-enzymatic XCC assemblies (XCC-A to XCC-L). The constructs combine the 3 enzymes through the Jo and In modules.

Figure 2: Specific activities of the multienzyme assemblies and their free counterparts on soluble cellulose and xylan substrates. Specific activities were measured using (A) carboxymethyl cellulose and (B) arabinoxylan and are expressed as µmol reducing sugar equivalents (min^{-1}.µmol $enzyme^{-1}$) to enable comparison between free enzymes and the corresponding multienzyme assemblies. Statistical significance was assessed by one-way ANOVA followed by Dunnett's multiple-comparisons test, with each multienzyme assembly compared with the

Free enzymes reference. Significance levels are indicated as follows: $p \geq 0.05$, non-significant (ns); $p < 0.05$, *; $p < 0.01$, **; $p < 0.001$, ***; $p < 0.0001$, ****. N.D., not detected; N.A., not assessed.

Figure 3: Destarched wheat bran (dWB) and wheat straw (WS) degradation by free enzymes and multienzymatic assemblies. Total soluble sugar released from (A) destarched wheat bran and (B) wheat straw, expressed as mM reducing sugar equivalents (DNS assay) for n = 3, with T = 0h00 values subtracted. For clarity, mean kinetics of all 12 assemblies are shown as dashed lines, with their range (min-max) highlighted in grey. The best and the worst performer at t = 24h00 are highlighted in green and red respectively.

Figure 4: Oligosaccharide profiles generated during dWB and WS degradation by MALDI-TOF MS. Heatmaps showing the normalized intensity (Log I_norm, a.u.) of oligosaccharide families released during enzymatic degradation of wheat bran (A) and wheat straw (B), at different incubation time points (T0, T1, T2, T4, T8, and T24 hours). Each column corresponds to a tested condition: BU (buffer), FE (free enzymes) and enzymatic assemblies A to L. Rows represent the different oligosaccharide families identified by MALDI-TOF (pentoses, hexoses, feruloylated pentoses, acetylated and feruloylated pentoses, acetylated pentoses in panel A, and for panel B, additional categories: mixed pentoses-hexoses, hexuronic acid-containing pentoses, p-coumaroylated pentoses and acetylated mixed pentoses-hexoses). Signal intensity is represented on a color scale ranging from blue (< 4) to red (8), indicating the relative abundance of each oligosaccharide family.

Figure 5: Principal component analysis (PCA) and hierarchical clustering of WS degradation profiles link assemblies to cellulolytic or xylanolytic oligosaccharide release patterns. (A) PCA of oligosaccharide release profiles at different times. Each assembly (A to K; n = 9) is shown with its 95% confidence ellipse. The red and blue ellipses enclose the cellulolytic and xylanolytic assemblies, respectively. XCC-C and XCC-L were excluded from the PCA because of their high variability, which allows a clearer visualization of the cluster distribution. The percentage of variance explained by each principal component is indicated on the axes. B. Hierarchical clustering of the 12 enzymatic assemblies, based on the normalized intensity (Log_{10} I_{norm}, a.u.) of three degraded plant cell wall polysaccharides (xylans, xyloglucans and cellulose) after 8 h of incubation. Terms “Cellulolytic assemblies” and “Xylanolytic assemblies” stand for a higher release of cellulose-derived oligosaccharides and a higher release of xylan-derived oligosaccharides assemblies, respectively.

Figure 6: Structural characterization of the 12 XCC complexes by SAXS and correlation with oligosaccharide degradation profiles. (A) SAXS scattering profiles of the 12 XCC complexes, vertically offset and scaled by powers of 10 for clarity from A to L from top to bottom. Colors were kept identical. The lower panel shows the normalized residuals (Δ/σ) between XCC-A and all other assemblies to magnify low q differences. (B) Radius of gyration (Rg) and maximum particle dimension (Dmax) determined from the SAXS data. (C) Inter-domain distances determined for the five pairs of domains. (D) Kendall’s τ correlation matrix between inter-domain distances and xylo- and cello-oligosaccharide profiles measured at 1, 2, 4, 8, and 24 h from WS degradation. No statistically significant correlations were observed.

Figures

Figure 1

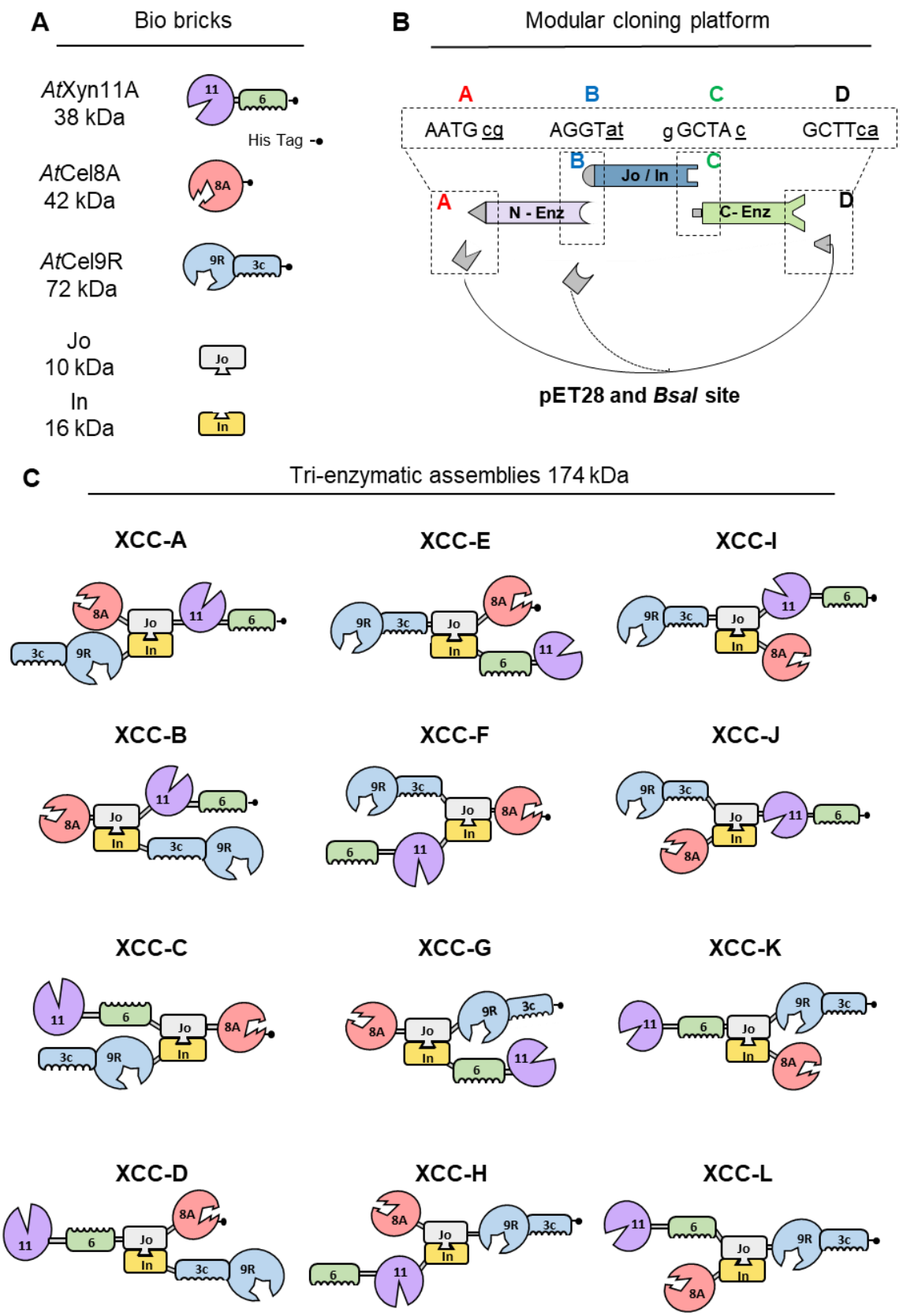


Figure 2

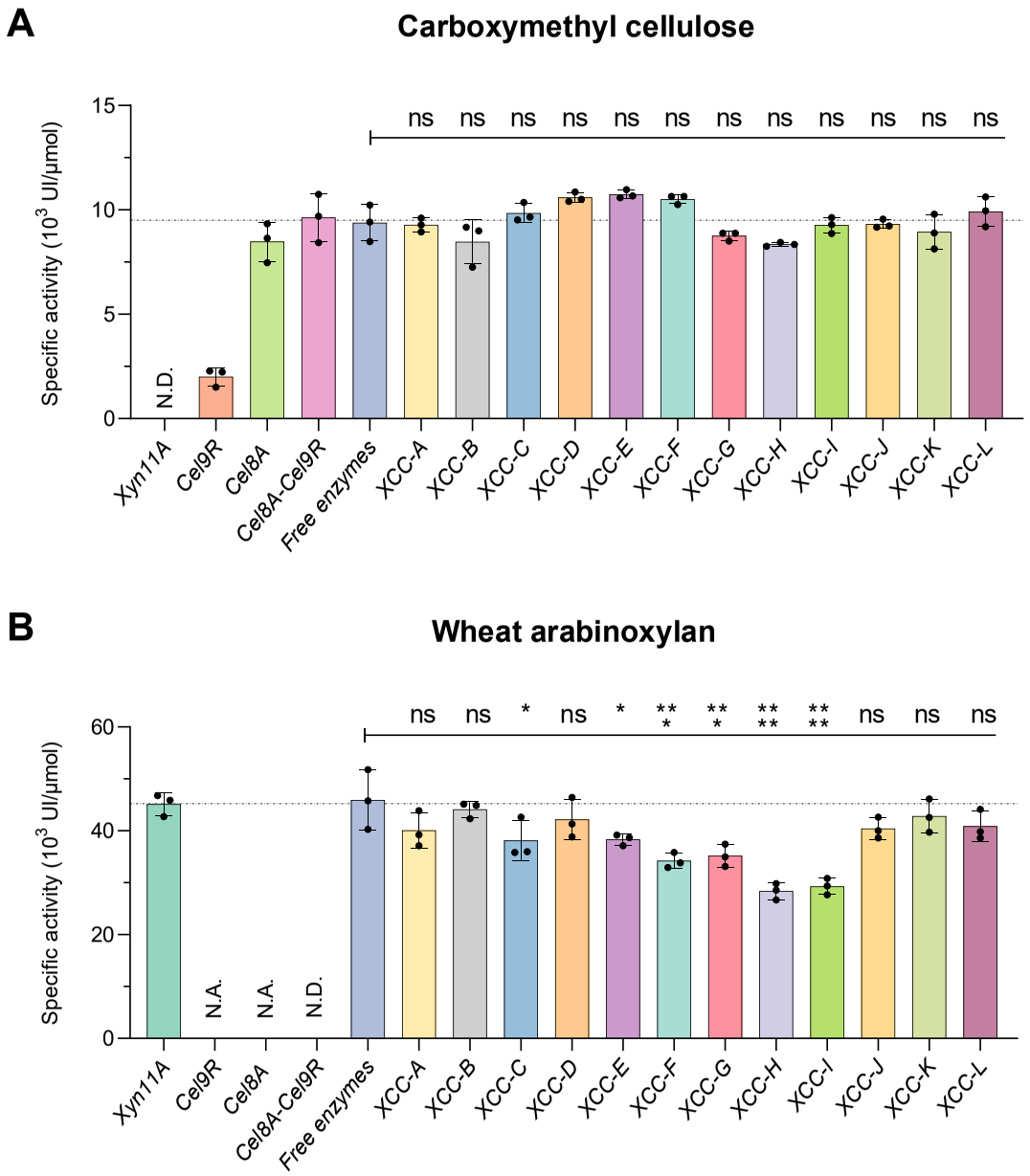

A
Carboxymethyl cellulose
Specific activity (10^3 UI/µmol)
15
10
5
0
ns ns ns ns ns ns ns ns ns ns ns ns
N.D.
Xyn11A
Cel9R
Cel8A
Cel8A-Cel9R
Free enzymes
XCC-A
XCC-B
XCC-C
XCC-D
XCC-E
XCC-F
XCC-G
XCC-H
XCC-I
XCC-J
XCC-K
XCC-L
B
Wheat arabinoxylan
Specific activity (10^3 UI/µmol)
60
40
20
0
ns ns * ns * ** ** ** ** ns ns ns
N.A.
N.A.
N.D.
Xyn11A
Cel9R
Cel8A
Cel8A-Cel9R
Free enzymes
XCC-A
XCC-B
XCC-C
XCC-D
XCC-E
XCC-F
XCC-G
XCC-H
XCC-I
XCC-J
XCC-K
XCC-L

Figure 3

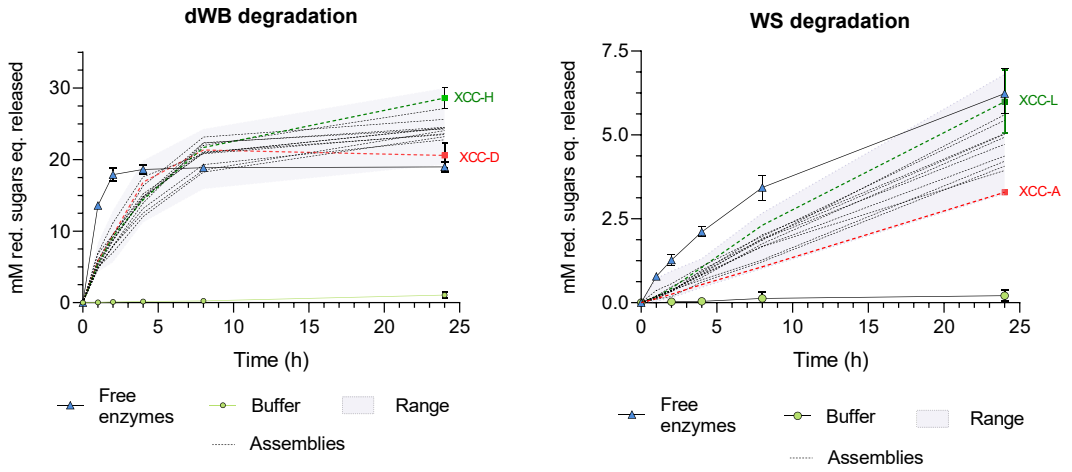


Figure 4

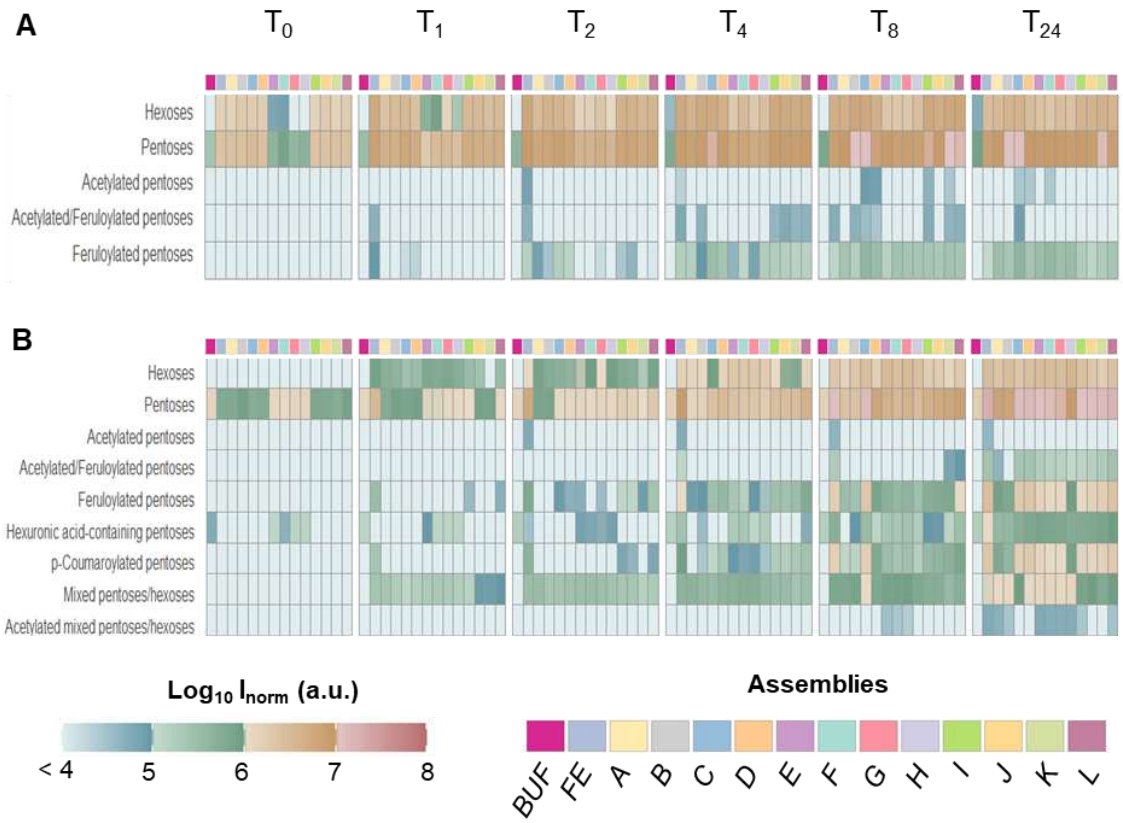
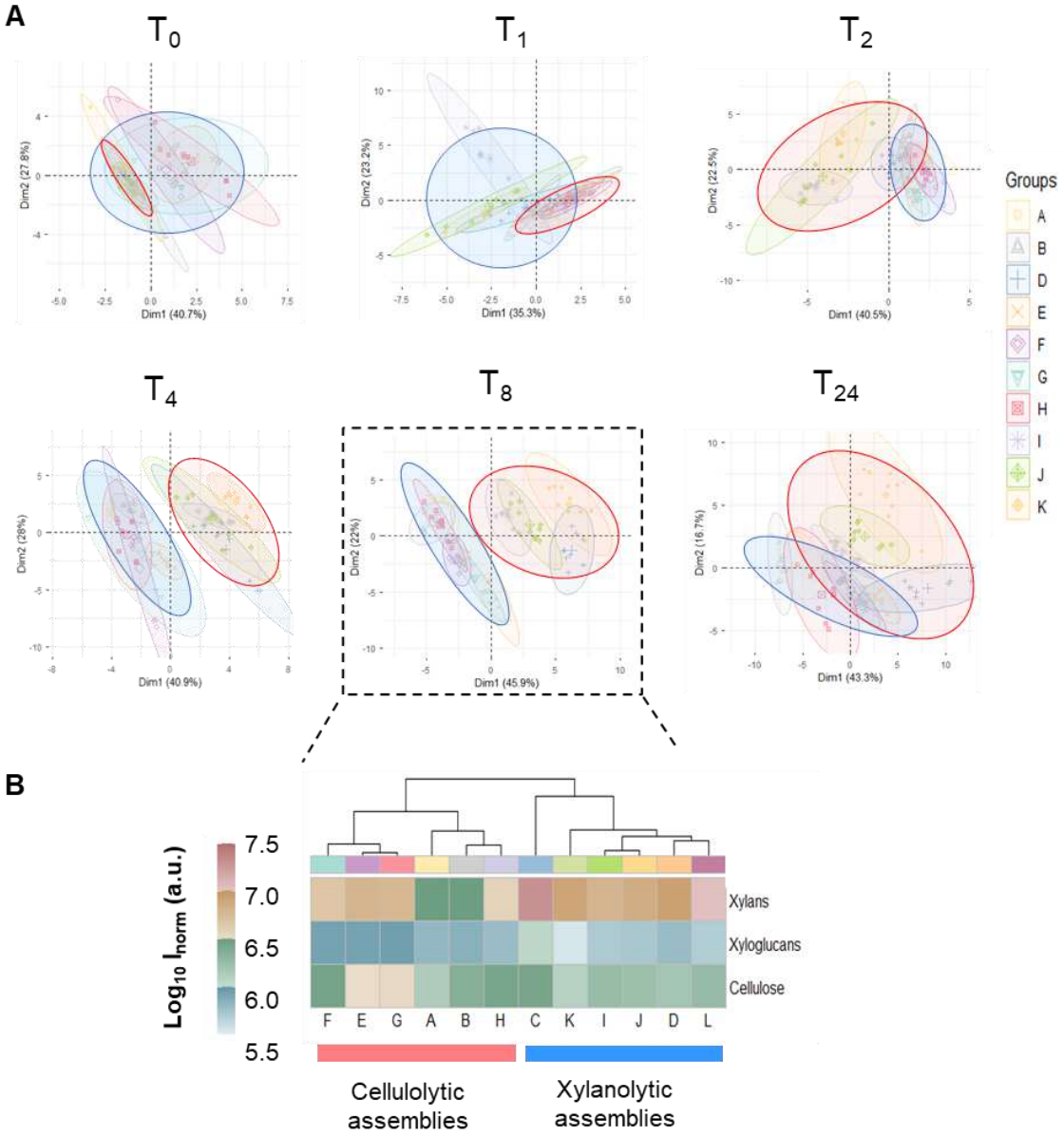
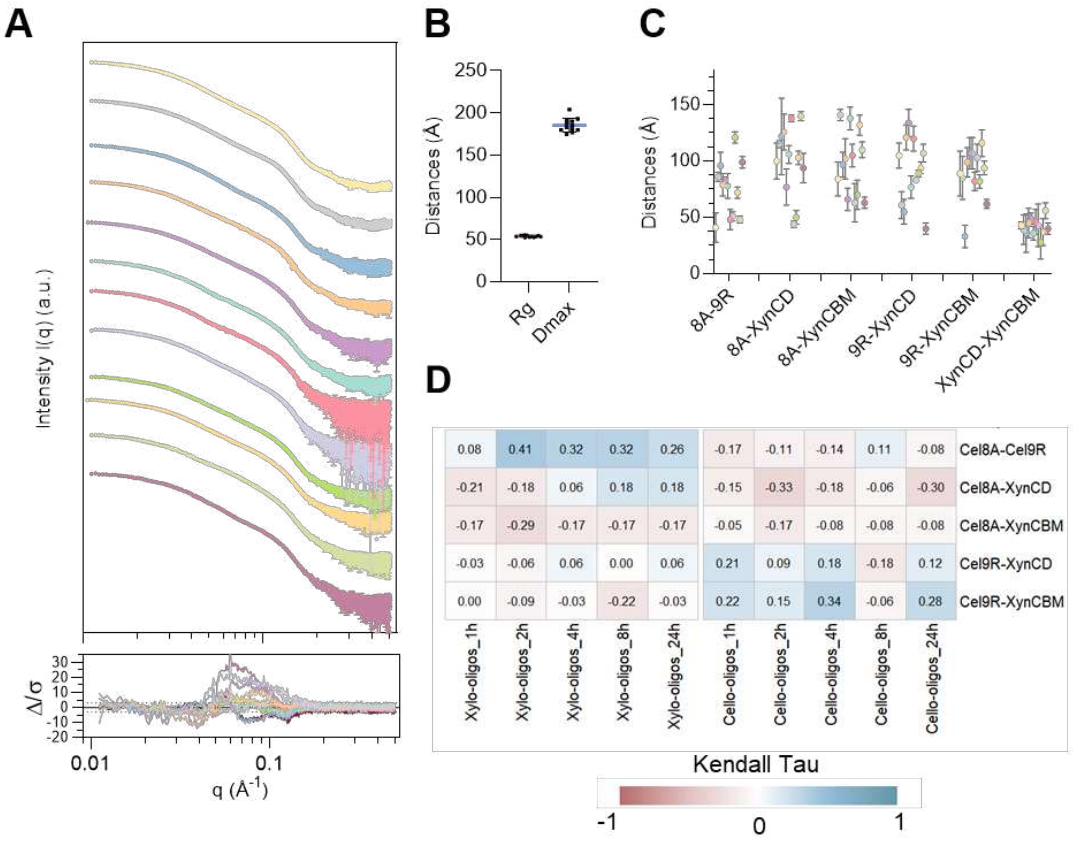


Figure 5


A
T0
T1
T2
T4
T8
T24
Dim1 (40.7%)
Dim2 (27.8%)
Dim1 (35.3%)
Dim2 (23.2%)
Dim1 (40.5%)
Dim2 (22.5%)
Dim1 (40.9%)
Dim2 (28%)
Dim1 (45.9%)
Dim2 (22%)
Dim1 (43.3%)
Dim2 (16.7%)
Groups
A
B
D
E
F
G
H
I
J
K
B
Log10 Inorm (a.u.)
7.5
7.0
6.5
6.0
5.5
Xylans
Xyloglucans
Cellulose
F E G A B H C K I J D L
Cellulolytic assemblies
Xylanolytic assemblies

865 Figure 6



866

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SILarrabeiti.docx