












# The effect of linker conformation on performance and stability of a two-domain lytic polysaccharide monoxygenase

Received for publication, June 23, 2023, and in revised form, September 5, 2023. Published, Papers in Press, September 19, 2023.

<https://doi.org/10.1016/j.jbc.2023.105262>

Zarah Forsberg<sup>1,\*</sup> , Anton A. Stepnov<sup>1</sup> , Giulio Tesei<sup>2</sup> , Yong Wang<sup>2,3</sup> , Edith Buchinger<sup>4,5</sup>, Sandra K. Kristiansen<sup>5</sup>, Finn L. Aachmann<sup>5</sup> , Lise Arleth<sup>6</sup> , Vincent G. H. Eijsink<sup>1</sup> , Kresten Lindorff-Larsen<sup>2</sup> , and Gaston Courtade<sup>5,\*</sup> 

From the <sup>1</sup>Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences (NMBU), Ås, Norway; <sup>2</sup>Structural Biology and NMR Laboratory, Department of Biology, Linderstrøm-Lang Centre for Protein Science, University of Copenhagen, Copenhagen, Denmark; <sup>3</sup>College of Life Sciences, Zhejiang University, Hangzhou, China; <sup>4</sup>Vectron Biosolutions AS, Trondheim, Norway; <sup>5</sup>Department of Biotechnology and Food Science, NTNU Norwegian University of Science and Technology, Trondheim, Norway; <sup>6</sup>X-ray and Neutron Science, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

Reviewed by members of the JBC Editorial Board. Edited by Karen Fleming

A considerable number of lytic polysaccharide monoxygenases (LPMOs) and other carbohydrate-active enzymes are modular, with catalytic domains being tethered to additional domains, such as carbohydrate-binding modules, by flexible linkers. While such linkers may affect the structure, function, and stability of the enzyme, their roles remain largely enigmatic, as do the reasons for natural variation in length and sequence. Here, we have explored linker functionality using the two-domain cellulose-active ScLPMO10C from *Streptomyces coelicolor* as a model system. In addition to investigating the WT enzyme, we engineered three linker variants to address the impact of both length and sequence and characterized these using small-angle X-ray scattering, NMR, molecular dynamics simulations, and functional assays. The resulting data revealed that, in the case of ScLPMO10C, linker length is the main determinant of linker conformation and enzyme performance. Both the WT and a serine-rich variant, which have the same linker length, demonstrated better performance compared with those with either a shorter linker or a longer linker. A highlight of our findings was the substantial thermostability observed in the serine-rich variant. Importantly, the linker affects thermal unfolding behavior and enzyme stability. In particular, unfolding studies show that the two domains unfold independently when mixed, whereas the full-length enzyme shows one cooperative unfolding transition, meaning that the impact of linkers in biomass-processing enzymes is more complex than mere structural tethering.

Enzymatic depolymerization of polysaccharides in lignocellulosic biomass, in particular cellulose, is of great scientific and industrial interest to produce biofuels, biomaterials, and commodity chemicals (1). However, the recalcitrance of crystalline cellulose poses an obstacle for its efficient saccharification (2).

This obstacle may be overcome by applying lignocellulolytic enzyme cocktails comprising glycoside hydrolases (GHs, e.g., xylanases, cellulases) and copper-dependent redox enzymes known as lytic polysaccharide monoxygenases (LPMOs). Whilst GHs hydrolyze their substrates, LPMOs use an oxidative mechanism to cleave the  $\beta$ -1,4 glycosidic bonds (3, 4) in cellulose and various hemicellulosic  $\beta$ -glucans (5). Many carbohydrate-active enzymes are modular, consisting, for example, of a catalytic domain and a carbohydrate-binding module (CBM). These domains are connected by linker sequences of variable amino acid composition and length (6, 7). While there is extensive and in-depth understanding of structure–function relationships for the individual catalytic domains and CBMs, knowledge on the conformations and roles of linkers remain scant, despite pioneering work on fungal GH7 cellulases (8, 9).

Multidomain LPMOs with a family AA10 catalytic domain are often tethered by disordered linkers of different lengths to one or more catalytic or substrate-binding domains. For example, two chitin-binding modules (CBM5 and CBM73) in CjLPMO10A (10), a GH18 and a CBM5 in JdLPMO10A (11), two FnIII domains and a CBM5 in BtLPMO10A (12), or a CBM2 in ScLPMO10C (our model enzyme). It is well established that the CBM2 domain enhances binding affinity of the full-length enzyme to its cellulose substrate (13), which is important for LPMOs as localization close to the substrate protects the enzyme from self-inactivating off-pathway reactions (14). However, the roles of the linkers in these types of bacterial enzymes remain largely enigmatic.

Intrinsically disordered linkers have heterogeneous conformations, and characterizing the conformational ensembles of multidomain proteins is crucial to understand their function. Although they may be regarded as just keeping domains close, it is clear that linkers may be key functional regions that govern structurally and functionally important interactions between the folded domains (15). Linker functionality is affected by the amino acid composition, sequence patterns,

\* For correspondence: Zarah Forsberg, [zarah.forsberg@nmbu.no](mailto:zarah.forsberg@nmbu.no); Gaston Courtade, [gaston.courtade@ntnu.no](mailto:gaston.courtade@ntnu.no).

## Structural and functional roles of the linker in an LPMO

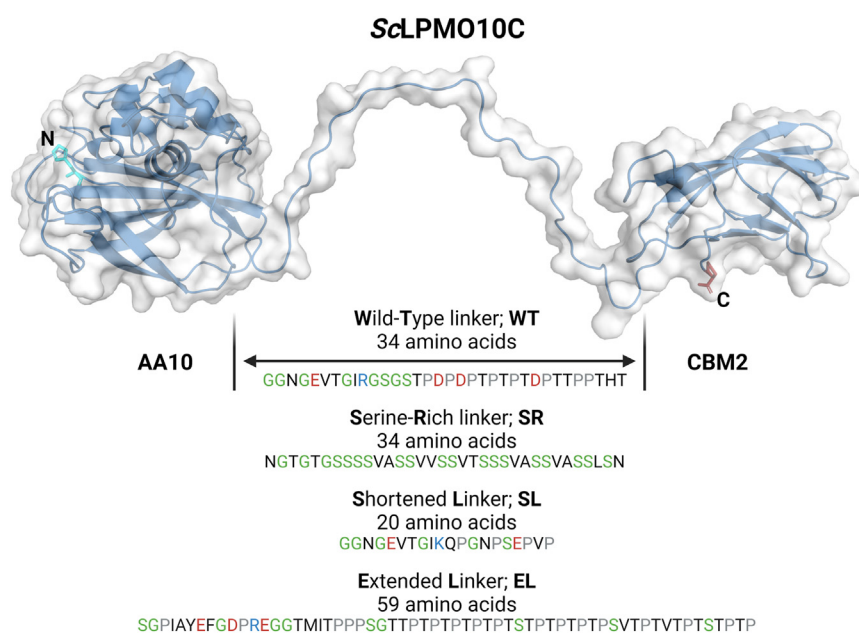
and length (16). For example, prolines and charged residues (aspartate, glutamate, arginine, and lysine) may promote extended conformations, and glycines promote flexibility and conformational freedom in linkers as well as intrinsically disordered proteins (IDPs) (15, 17, 18). As another example, serine-rich linkers (SRs) are thought to be more flexible than proline-rich linkers (19, 20). In the context of LPMOs, Tamburrini *et al.* (21) identified that numerous LPMOs have an intrinsically disordered region located at the C terminus. These C-terminal extensions are typically longer than the linkers found between domains and are unique to LPMOs and not observed in other carbohydrate-active enzymes or oxidoreductases. Most of these LPMO C-terminal extensions feature at least one putative binding site, underpinning the importance of these disordered regions in LPMO function.

Structural characterization of multidomain proteins is challenging because the flexibility of the linker allows for many individual conformations, including different interdomain distances and relative orientations. Therefore, elucidating the structure of multidomain proteins typically requires a combination of complementary biophysical techniques and computer simulations (22). One key technique in this respect is small-angle X-ray scattering (SAXS), which can reveal the average dimension and conformation of multidomain proteins in solution as well as provide information about the structural dispersion of flexible conformations (23, 24). Moreover, SAXS curves can be calculated from atomistic coordinates, allowing integration of experimental SAXS data with computer modeling or simulations (25). Molecular dynamics (MD) simulations allow modeling of the experimentally observed data and can yield models with predictive value. Accurate MD simulations hinge on realistic, sufficiently detailed, and computationally expensive models and the use of relevant simulation times, which may lead to prohibitive demands on

computational power. In practice, it is desirable to achieve a compromise between accuracy and computational efficiency, which can be attained by using coarse-grained (CG) models. CG models have been successfully used to model IDPs and flexible linkers (26).

Here, we use a combination of SAXS and MD simulations, as well as NMR spectroscopy, to investigate the conformation of a two-domain LPMO, namely ScLPMO10C (previously known as CelS2; 34.6 kDa). ScLPMO10C is a cellulose-oxidizing LPMO from *Streptomyces coelicolor* A3(2) (27) that is composed of an N-terminal catalytic domain belonging to family 10 of auxiliary activities in the carbohydrate-active enzyme database (ScAA10; Protein Data Bank [PDB] code: 4OY7; residues 35–228; 20.9 kDa; (28)), which is connected to a C-terminal family 2 CBM (ScCBM2; PDB code: 6F7E; residues 263–364; (13)) through a linker (34 residues; residues 229–262) that is rich in prolines, threonines, and charged residues.

To investigate the effect of linker composition and length, we engineered three variants of ScLPMO10C. These variants were designed by substituting the WT linker with linkers derived from other carbohydrate-active enzymes. The four linkers are illustrated in Figure 1, where we provide a graphical representation of the domain architecture of ScLPMO10C (AA10-linker-CBM2) and the primary structures of the linkers. Additional information about the linkers is presented in Table 1, and the complete protein sequences including domain boundaries are presented in Table S1. Our engineered variants comprise a shorter 20-residue linker of similar amino acid composition as the WT linker (shortened linker; SL), an extended 59-residue linker also of similar composition as the WT linker (extended linker; EL), and a linker rich in serine residues (serine-rich; SR) that has the same number of residues as the WT but is devoid of prolines or charged amino acids.



**Figure 1. Illustration of the domain architecture of ScLPMO10C with primary structures of the WT linker and the three engineered variants.** In the linkers, negatively and positively charged residues, as well as prolines, are colored red, blue, and gray, respectively. Glycines and serines are colored green.

**Table 1**  
Overview of the properties of four different linker variants of ScLPMO10C

Linker	<i>N</i>	NCPR	Fraction expanding	Fraction	Experimental	Experimental	Calculated
				Proline	$R_h$ (Å)	$R_g$ (Å)	$R_g$ (Å)
WT	34	-0.09	0.38	0.23	30.2 ± 0.1	35.3 ± 0.4	28.3 ± 1.1
SR	34	0.00	0.00	0.00	30.0 ± 0.3	ND	25.5 ± 0.5
SL	20	-0.05	0.35	0.20	29.5 ± 0.4	ND	25.4 ± 0.9
EL	59	-0.03	0.37	0.31	31.4 ± 0.6	ND	33.1 ± 1.3

The experimental hydrodynamic radius,  $R_h$ , was calculated from NMR diffusion experiments. The radius of gyration,  $R_g$ , was derived from SAXS data and calculated from uniformly weighted CG MD simulations. Errors are indicated as ±SD. *N*, number of amino acids in the linker region; NCPR, net charge per residue; fraction expanding, fraction of residues that are predicted to contribute to chain expansion (E/D/R/K/P). Abbreviation: ND, not determined.

Following the creation of these variants, we used SAXS, NMR, computational modeling, and functional assays to investigate how the structure, dynamics, stability, and catalytic performance of the enzyme is affected by variation in the linker. The results show that linker length may be more important than sequence for determining the conformation and activity of ScLPMO10C. Overall, our findings provide a basis for understanding the structural and functional roles of linkers in not only LPMOs but also multimodular proteins in general.

## Results

### NMR insights on the linker region of ScLPMO10C

Previously, we published the partial backbone assignment of ScLPMO10C (29) as well as preliminary  $T_1$  and  $T_2$  relaxation and  $^1\text{H}$ - $^{15}\text{N}$  NOE data (13). In our previous work, approximately 80% of the linker resonances were assigned because of difficulties arising from low signal dispersion of linker resonances combined with signal overlap and a high content of prolines. To enable more detailed studies of the linker region, we have here accomplished >97% assignment coverage of the  $\text{H}^N$ ,  $\text{N}$ ,  $\text{C}^\alpha$ ,  $\text{C}'$ , and  $\text{C}^\beta$  resonances in the linker region. To achieve a high assignment percentage of proline-rich regions in the linker, we used a  $^{13}\text{C}$ -detected CON-based approach (30) that correlates the amide N atom of an amino acid, *i*, with the carbonyl  $\text{C}'$  atom of the preceding amino acid, *i*-1. Using the peak intensities from the *i*-1 residue with respect to Pro, we observed that two prolines in the linker (P248 and P258) had a fraction of *cis*-Pro in the range of 6 to 7% (Fig. S1). This is consistent with the fraction of *cis*-Pro in the IDP  $\alpha$ -synuclein (31). The updated chemical shift assignment of full-length ScLPMO10C has been deposited in the Biological Magnetic Resonance Data Bank under accession number 27078.

The “average” secondary structure propensity for the linker region was estimated from the chemical shift assignment using TALOS-N, indicating an extended structure with dihedral angles typical of  $\beta$ -strands or polyproline II structures (Fig. 2A), in agreement with previous observations (13).

In addition, we analyzed  $^{15}\text{N}$  relaxation data ( $R_1$ ,  $R_2$ , and  $^1\text{H}$ - $^{15}\text{N}$  NOE) to provide more complete and accurate insights into the dynamic features of ScLPMO10C and in particular its newly assigned linker region. Figure 2, B and C shows that amino acids in the linker stand out but display clear features of flexibility in the picosecond–nanosecond timescale, as can be

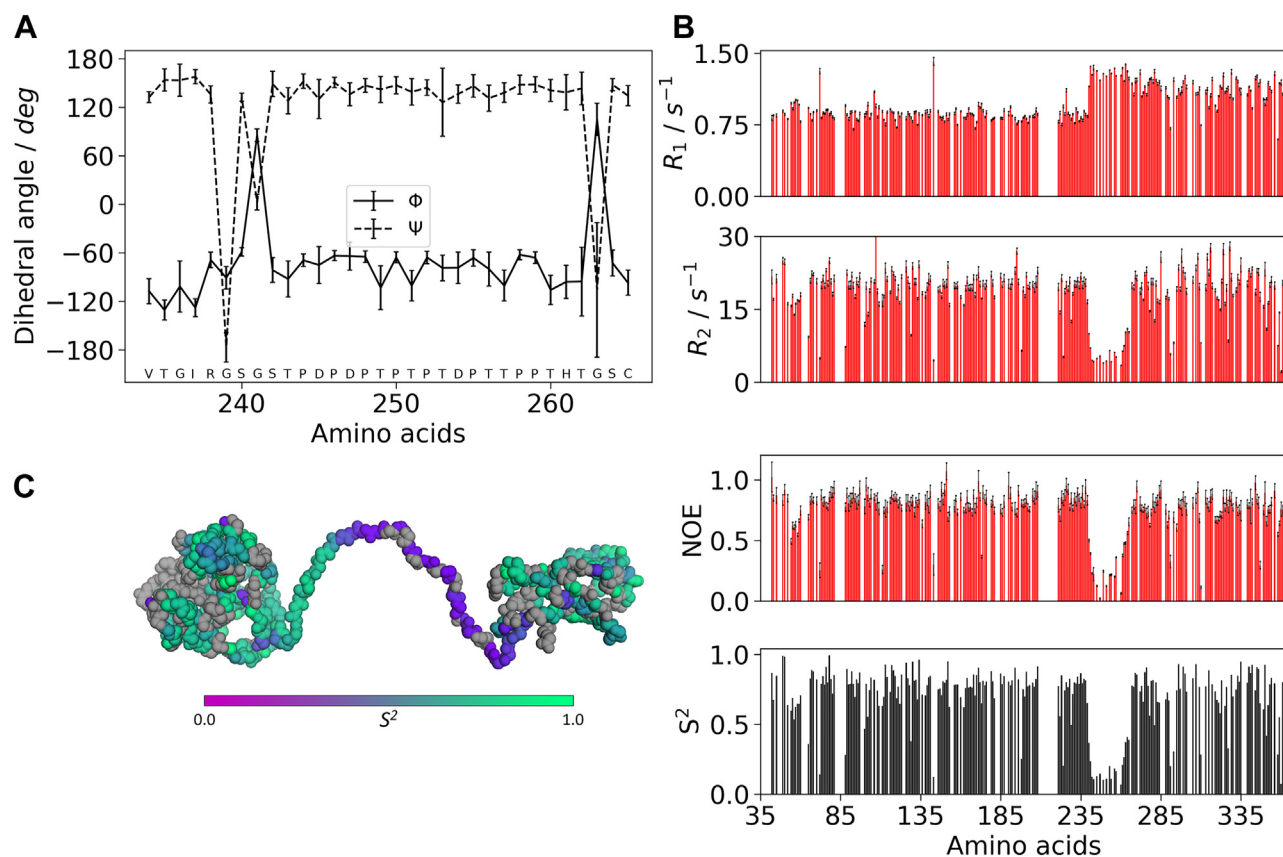
seen from increased  $R_1$  relaxation rates and decreases in  $R_2$ ,  $^1\text{H}$ - $^{15}\text{N}$  NOEs, and generalized order parameter,  $S^2$ . We note that it is difficult to interpret the  $S^2$  values as internal and overall motions cannot easily be separated for the linker residues.

### Overall dimensions and shape of ScLPMO10C

We studied full-length ScLPMO10C (Fig. 3, A–D) and its isolated catalytic domain (called ScAA10; Fig. 3, E and F) by SAXS to determine the effect of the linker on the overall conformation and shape of the LPMO. The average sizes of the proteins were determined by calculating the radius of gyration,  $R_g$ , from experimental SAXS profiles, using Guinier plots. Based on X-ray crystallography (28), SAXS data (Fig. 3, E–G), and NMR data showing that ScAA10 is rigid in solution (Fig. 2B), it can be assumed to have a roughly spherical shape. Therefore, the Guinier approximation  $qR_g \leq 1.3$  (32) was used to determine an  $R_g = 16.25 \pm 0.02$  Å (Fig. S2). On the other hand, ScLPMO10C is not spherical, and the flexible linker leads to an ensemble of conformations (13). Thus, the Guinier approximation was considered to be valid only up to  $qR_g \leq 1$ , an assumption previously reported to be valid for Cel45 from *Humicola insolens*, another two-domain carbohydrate-active enzyme (33). By considering this, the  $R_g$  of full-length ScLPMO10C was determined to be  $35.3 \pm 0.4$  Å (Table 1 and Fig. S2).

The distance distribution functions,  $p(r)$ , for ScAA10 and ScLPMO10C were calculated from their respective scattering intensities. The  $p(r)$  profile of ScAA10 exhibits two peaks at 17 Å and 24 Å, respectively, and a maximum distance,  $D_{max}$  around 45 Å (Fig. 3G). The  $p(r)$  profile of ScLPMO10C displays a biphasic pattern with an initial peak from zero to ~50 Å, a pronounced shoulder from ~50 to ~100 Å, and then a tail that flattens out to a  $D_{max}$  around 150 Å (Fig. 3G). This is consistent with the full-length enzyme having an elongated shape because of the dumbbell shape and the interconnecting flexible linker. The first peak, with a maximum at around  $r = 25$  Å represents intramolecular distances within the ScAA10 and ScCBM2 domains. The second part of the curve with the pronounced shoulder stems from (larger) interdomain distances between the ScAA10 and ScCBM2 domains, indicating an extended conformation of the protein (Fig. 3G). The SAXS profile calculated on the structure of ScAA10 (PDB code:

## Structural and functional roles of the linker in an LPMO



**Figure 2. Dihedral angles and  $^{15}\text{N}$ -relaxation measurements for the ScLPMO10C.** A, dihedral angles ( $\phi$  and  $\psi$ ) in the linker region and flanking residues of ScLPMO10C derived from the chemical shift assignment using TALOS-N. The values indicate that the linker is in an extended conformation. B, from top to bottom:  $^{15}\text{N}$ - $R_1$  rates,  $^{15}\text{N}$ - $R_2$  rates, heteronuclear  $\{^1\text{H}\}$ - $^{15}\text{N}$  NOEs, and calculated generalized order parameter,  $S^2$ . Amino acids displaying flexibility in the picosecond–nanosecond timescale show increased  $R_1$ , decreased  $R_2$ , heteronuclear NOEs, and  $S^2$ . C,  $S^2$  values colored on the structure of ScLPMO10C. Amino acids for which an  $S^2$  value could not be determined are colored gray.

4OY7) using CRYSOLO displays excellent agreement ( $\chi_r^2 = 1.45$ ) with the experimental data (Fig. 3E).

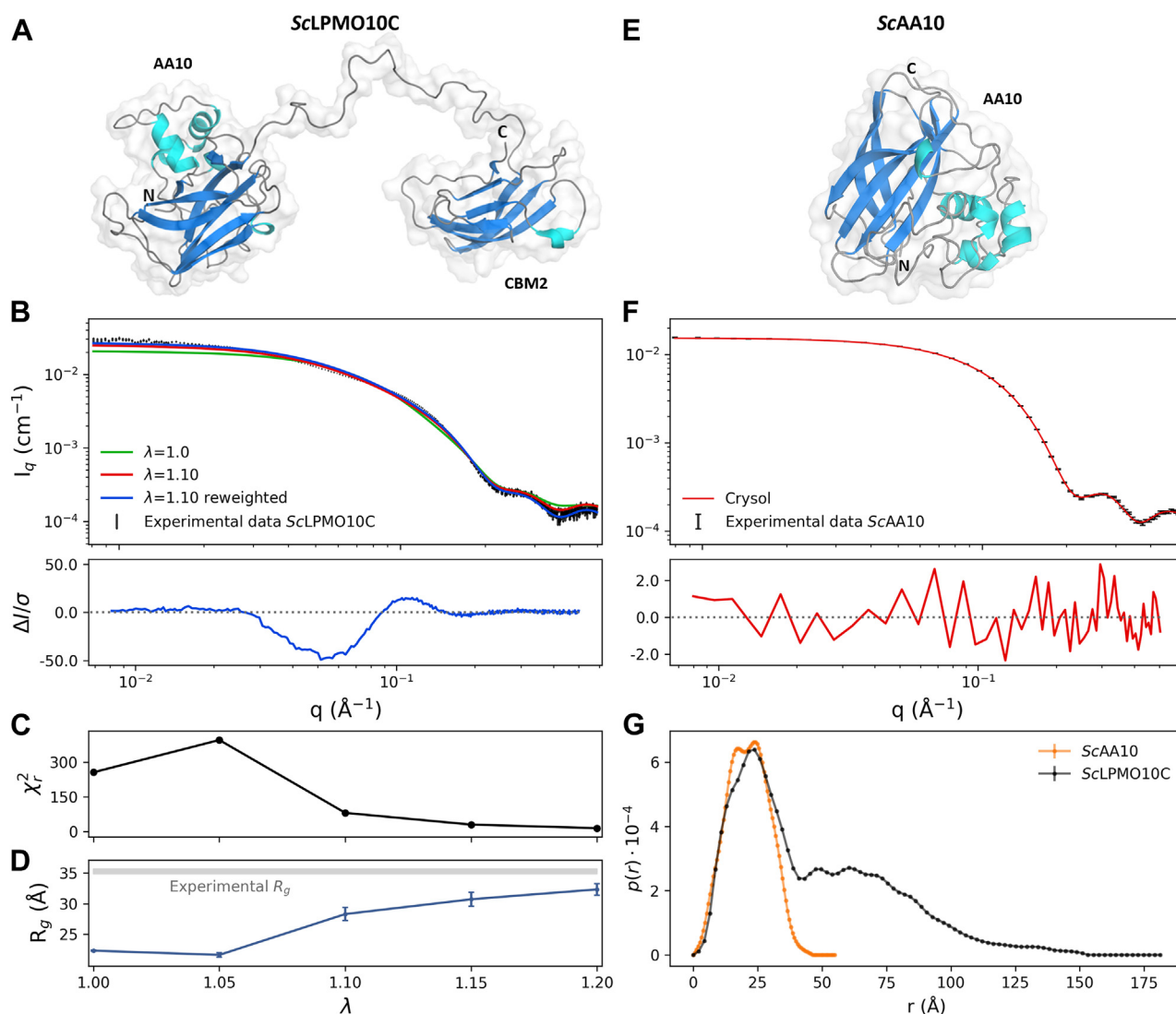
Despite our efforts, SAXS data for the three linker variants (SR, SL, and EL) were not obtained, so their size was instead determined by estimating the hydrodynamic radius,  $R_h$  (Table 1), from the translational diffusion coefficient,  $D_t$ , measured by pulsed-field gradient NMR diffusion experiments (Fig. S3). All proteins have  $R_h$  values in the range of 29 to 32 Å, and the uncertainties in the measurements are such that there is no discernible difference between the four enzyme variants. For the WT,  $R_g/R_h \approx 1.2$ , which is consistent with ratios found for expanded conformations in IDPs (34, 35). This is further evident that the linker in ScLPMO10C is predominantly extended.

### CG MD simulations describe SAXS data

To gain more insight into the conformation of the linkers in the ScLPMO10C variants, and to further investigate potential linker-dependent differences in the overall size of the protein, conformational ensembles of the WT and the three linker variants were generated using CG MD simulations with the Martini version 3.0.beta.4.17 force field (Fig. S4). We used the MD trajectories to calculate ensemble-averaged SAXS curves (Fig. 3B), where the contribution (weight) of each frame to the

scattering profile was optimized iteratively using Bayesian-maximum entropy (BME) (36) approach. Simulations of the WT enzyme resulted in a lower  $R_g$  value (22 Å; Fig. 3C) compared with the experimentally determined  $R_g$  value ( $35.3 \pm 0.4$  Å). Consequently, agreement between the experimentally determined and calculated SAXS profiles for the WT was poor ( $\chi_r^2 = 256.3$ ; Fig. 3D). Such unphysical compactness in MD simulations using the Martini force field has been observed previously (25, 37) and is attributed to protein “stickiness” that is a consequence of the force field that promotes too strong protein–protein interactions.

To improve agreement between SAXS data and simulations, we gradually increased the interaction strength between protein beads and water beads (Fig. 3D), as described in the Experimental procedures section. Using this approach,  $\lambda = 1.10$  (i.e., a 10% increase of the protein–water interaction strength) resulted in a slight improvement of the fit ( $\chi_r^2 = 80.9$ ; Fig. 3) between the experimental and computed SAXS curves and resulted in a calculated  $R_g = 28.3 \pm 1.1$  Å (Table 1). Although further incrementations of the protein–water interaction strength ( $\lambda = 1.15$  and  $\lambda = 1.20$ ) resulted in better agreement with SAXS data, we chose to continue with  $\lambda = 1.10$  as this was a compromise between improving the model and modifying the original force field as little as possible.



**Figure 3. Analysis of the fit to experimental SAXS data of WT ScLPMO10C and ScAA10.** *A*, representative conformation of ScLPMO10C, the AA10 and CBM2 domains, and the N and C termini are labeled. *B*, experimental SAXS data for ScLPMO10C (black) and calculated SAXS data from the ensemble with  $\lambda = 1.0$  (i.e., unmodified protein–water interactions; green), with  $\lambda = 1.10$  (red), and the reweighted ensemble with  $\lambda = 1.10$  (blue). *C*, Chi-squared ( $\chi^2_r$ ) values of the fit to experimental SAXS data of ensembles simulated at different values of  $\lambda$ . *D*, calculated radius of gyration ( $R_g$ ) of ensembles simulated at different values of  $\lambda$ ; the experimentally determined value is indicated by the gray horizontal line. *E*, representative conformation of the isolated catalytic domain, ScAA10 (Protein Data Bank code: 4OY7). *F*, experimental SAXS data for ScAA10 (black) and calculated SAXS data by crysol (red). *G*, distance distribution functions for ScAA10 and ScLPMO10C. Residual plots are shown below *B* and *F*, where  $\Delta I = I_{exp} - I_{fit}$  and  $\sigma$  is the experimental standard deviation. The difference between the experimental and calculated values, even when the simulations fit the data well, is because the experimental value represents the  $R_g$  of both the protein and the solvation layer, whereas the calculated value is of the protein only. The molecular mass of the ScLPMO10C and ScAA10 was estimated from the  $I(0)$  values to 34.6 kDa and 21.4 kDa, respectively. SAXS, small-angle X-ray scattering.

We used the simulations to interpret the experimental data by reweighting the trajectory (obtained with  $\lambda = 1.10$ ) using a BME protocol (36). In this approach, the weights of the simulation are optimized to fit the SAXS experiments. This reweighted ensemble had a calculated  $R_g = 32.3 \pm 1.1$  Å, and the SAXS curve calculated from the reweighted ensemble had a modest improvement of the fit ( $\chi^2_r = 5.8$ ) compared with the uniformly weighted ensemble. We note here that the difference between the experimental value (obtained from a Guinier fit to the SAXS experiments) and the calculated values is because the former includes a contribution from the solvent layer, whereas the latter represents the  $R_g$  value of the protein only.

Based on these results, we generated uniformly weighted ensembles of the three linker variants by running simulations using  $\lambda = 1.10$ . The results (Table 1 and Fig. S5) show that EL has the largest  $R_g$  value and the most extended conformations, whereas SR and SL have similar  $R_g$  values, of around 25 Å, and more compact conformations than EL and WT (Table 1). These results were unexpected because SR ( $N = 34$ ) is 14 amino acids longer than SL ( $N = 20$ ), and because both the WT and SR have linkers of equal length. The results suggest that differences in the calculated  $R_g$  values are not only caused by differences in the linker length (number of amino acids) but also caused by variation in the sequence composition of the linker. This is also supported by the lack of variance in  $R_h$

## Structural and functional roles of the linker in an LPMO

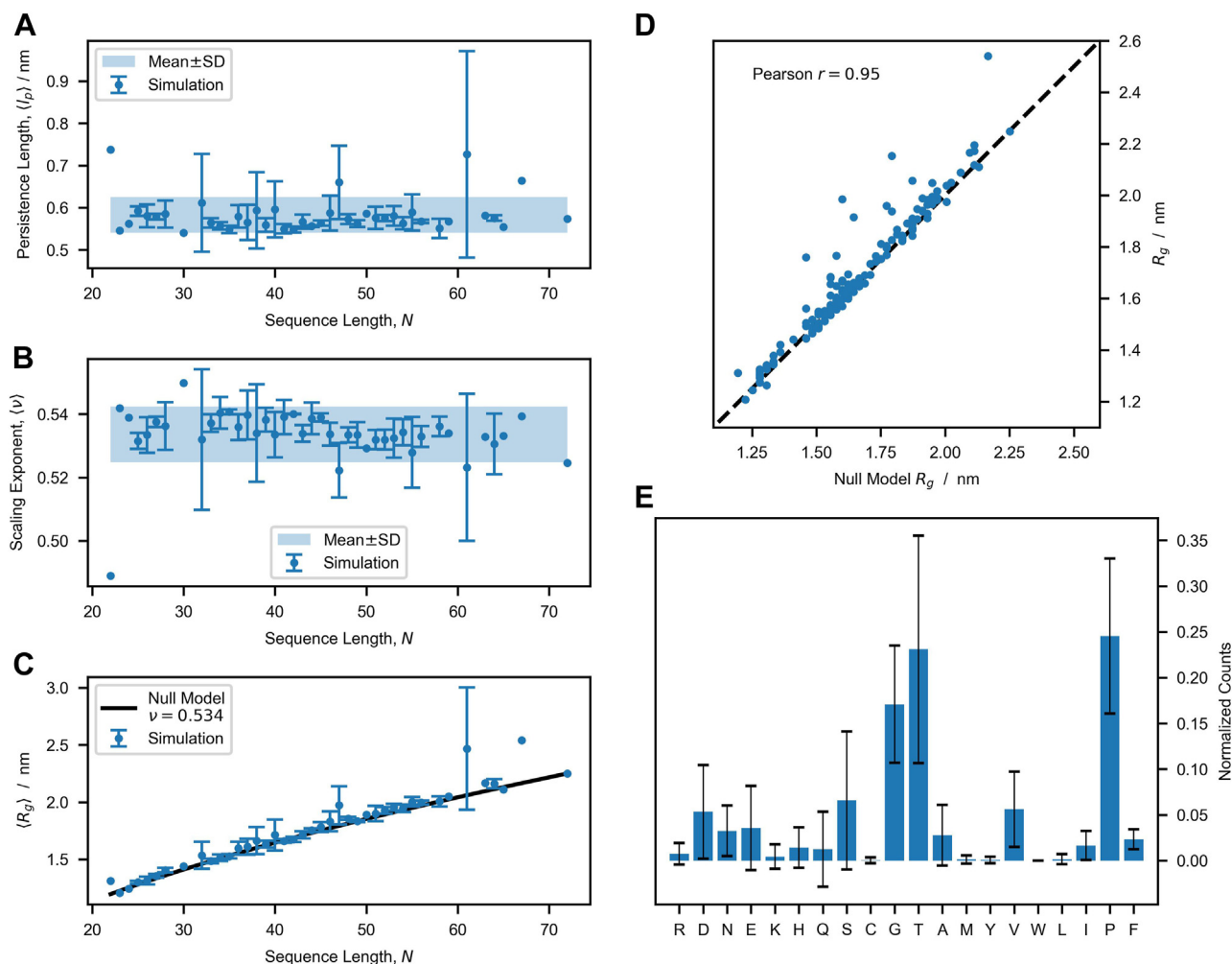
values (Table 1). To gain a better understanding of how linker sequences affect linker length, we conducted an *in silico* analysis to model a large number of linker sequences using recently developed techniques for the analysis of IDPs. The results of this analysis are presented later.

### Sequence length and not composition determines the size of isolated linkers

To obtain a larger coverage in terms of linker lengths and sequence compositions, we searched the UniProtKB Reference Proteomes (38) database for linker sequences in proteins with the same domain architecture as ScLPMO10C, that is, AA10-linker-CBM2. In total, we found 164 unique linkers (including the four discussed previously and depicted in Table S1), that we analyzed with MD simulations to investigate whether the  $R_g$  of the linker is influenced by its amino acid composition. In these simulations, linkers can be described as random polymer chains (39) with  $R_g \times kN^\nu$ , where  $N$  is the number of amino acids,  $k$  is a constant, and  $\nu$  is a scaling exponent that is sequence dependent (e.g.,  $\nu \sim 0.5$  for glycine-serine linkers,  $\nu \sim 0.57$  for proline-rich linkers, and  $\nu \sim 0.7$  for linkers with

charged residues (15)). Briefly explained,  $\nu = 0.5$  is the situation when protein-protein and protein-water interactions balance each other; when protein-water interactions are stronger  $\nu > 0.5$ , and when protein-protein interactions are stronger, the protein compacts and  $\nu < 0.5$ . Therefore, if the amino acid composition of the linkers influences the  $R_g$ , we would not expect a model where all the linkers are described by a single  $\nu$  to capture the  $R_g$  data.

We simulated the 164 linker sequences (average  $N = 40$ , median  $N = 38$ , minimum  $N = 22$ , maximum  $N = 72$ , SD = 11) using CALVADOS, a single-bead model trained with experimental data of IDPs (40). From the simulation trajectories of all the sequences, we calculated the  $R_g$  in three steps (see the Experimental procedures section for details). First, the persistence length,  $l_p$ , (Fig. 4A), which reflects the bending stiffness of the linkers, was calculated by fitting the autocorrelation function of bond vectors to a single exponential decay. Second, these  $l_p$  values were used as fixed parameters in the fit of the intramolecular pairwise distances to estimate the scaling exponent,  $\nu$  (Fig. 4B). We found that the distributions of  $l_p$  and  $\nu$  are both narrow ( $l_p = 0.58 \pm 0.04$  nm,  $\nu = 0.534 \pm 0.009$ ),



**Figure 4. Analysis of conformational ensembles of 164 LPMO linkers.** A and B, average persistence length,  $l_p$ , (A) and scaling exponent,  $\nu$ , (B) for linkers of equal sequence length,  $N$ . Error bars are SD, meaning that the graphs show mean  $\pm$  SD. C, comparison between the average  $R_g$  for linkers of  $N$  and prediction of the null model, that is,  $R_g = 0.23 \times N^{0.534}$  nm. D, correlation between the  $R_g$  values of the 164 linkers and the corresponding predictions of the null model. E, overall amino acid composition of the 164 sequences. The error bars show the SD of the distributions calculated for single linker sequences.

which means that all sequences are approximately equally disordered and flexible. The calculated value for the scaling exponent is in agreement with previous observations for IDPs ( $\nu = 0.51-0.60$ ) (18). Third, the resulting  $R_g$  data calculated solely on the average scaling exponent, that is,  $R_g = 0.23 \times N^{0.534}$ , was found to accurately capture the dependence of  $R_g$  on sequence length (Fig. 4, C and D). This means that for these 164 sequences, the amino acid composition does not influence linker conformation.

### The linker region affects enzyme stability under turnover conditions

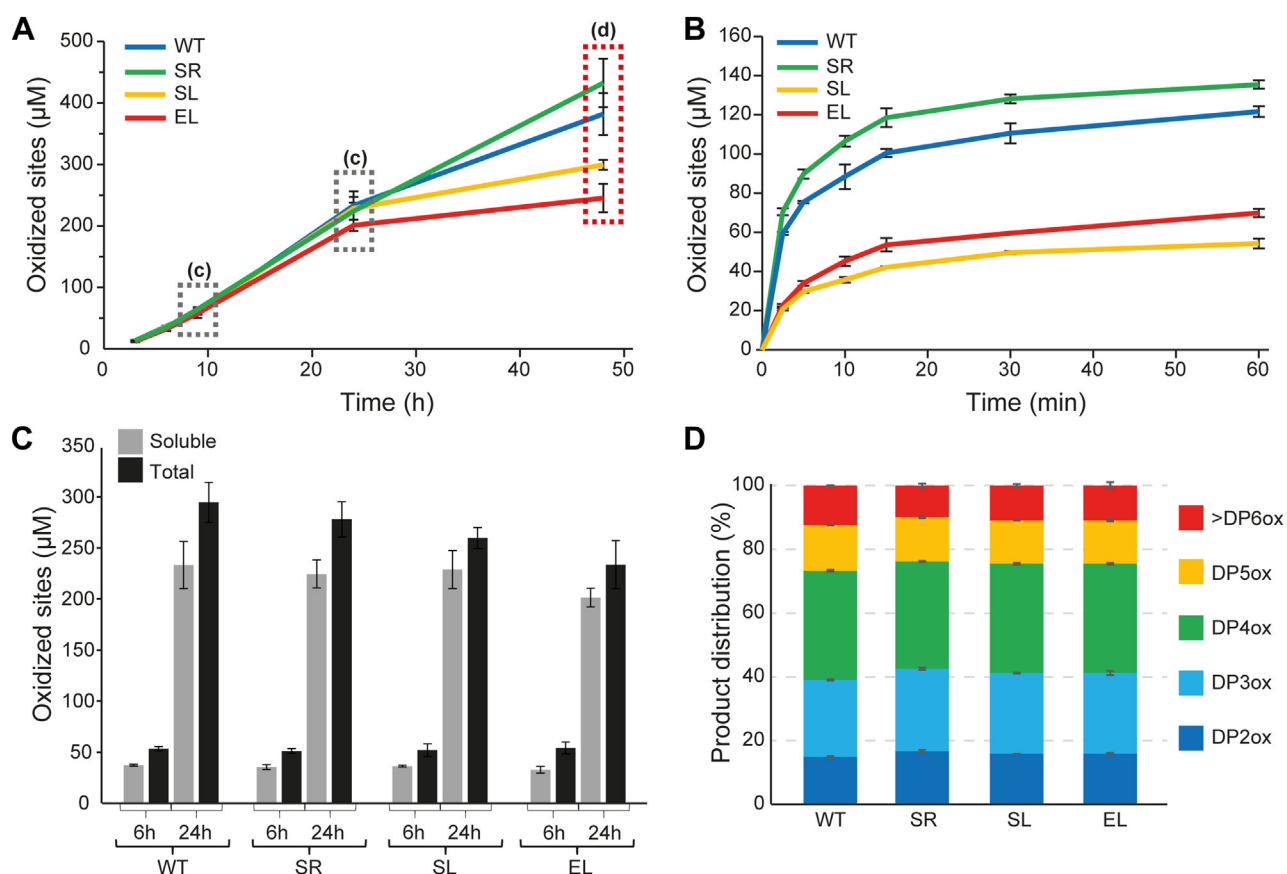
To probe the functionality of the linker variants, cellulose solubilization reactions were carried out in the absence or the presence of exogenously added hydrogen peroxide ( $H_2O_2$ ). In the experiment with no  $H_2O_2$  added (Fig. 5A), all LPMOs displayed similar substrate solubilization rates during the first 24 h of incubation. However, after 24 h, product release by SL and EL slowed down significantly compared with the other two enzymes, which is indicative of LPMO inactivation. The introduction of  $H_2O_2$  to LPMO reactions led to much (at least 10–100-fold) faster substrate oxidation rates (Fig. 5B), as expected (41–43). Interestingly, while the two variants with

equally long linkers, WT and SR, both converted essentially all added  $H_2O_2$  to oxidized products at similar speeds (*i.e.*, about 100  $\mu M$  of oxidized products after 10 min), SL and EL were much less effective. The latter is indicative of off-pathway reactions leading to enzyme inactivation that are promoted when an LPMO is exposed to  $H_2O_2$  while not optimally interacting with the substrate (13, 44, 45). The two LPMO variants with linkers of equal length, but very different sequences, are functionally highly similar.

Earlier studies with WT ScLPMO10C and its CBM- and linker-free truncated variant have shown that, with the substrate concentrations used here, the presence of the CBM promotes localized substrate oxidation, leading to a higher fraction of soluble oxidized products (relative to insoluble products) that on average are shorter (13). Figure 5, C and D shows that there are no substantial differences between the four variants in terms of neither the fraction of soluble oxidized products nor the degree of polymerization of these products.

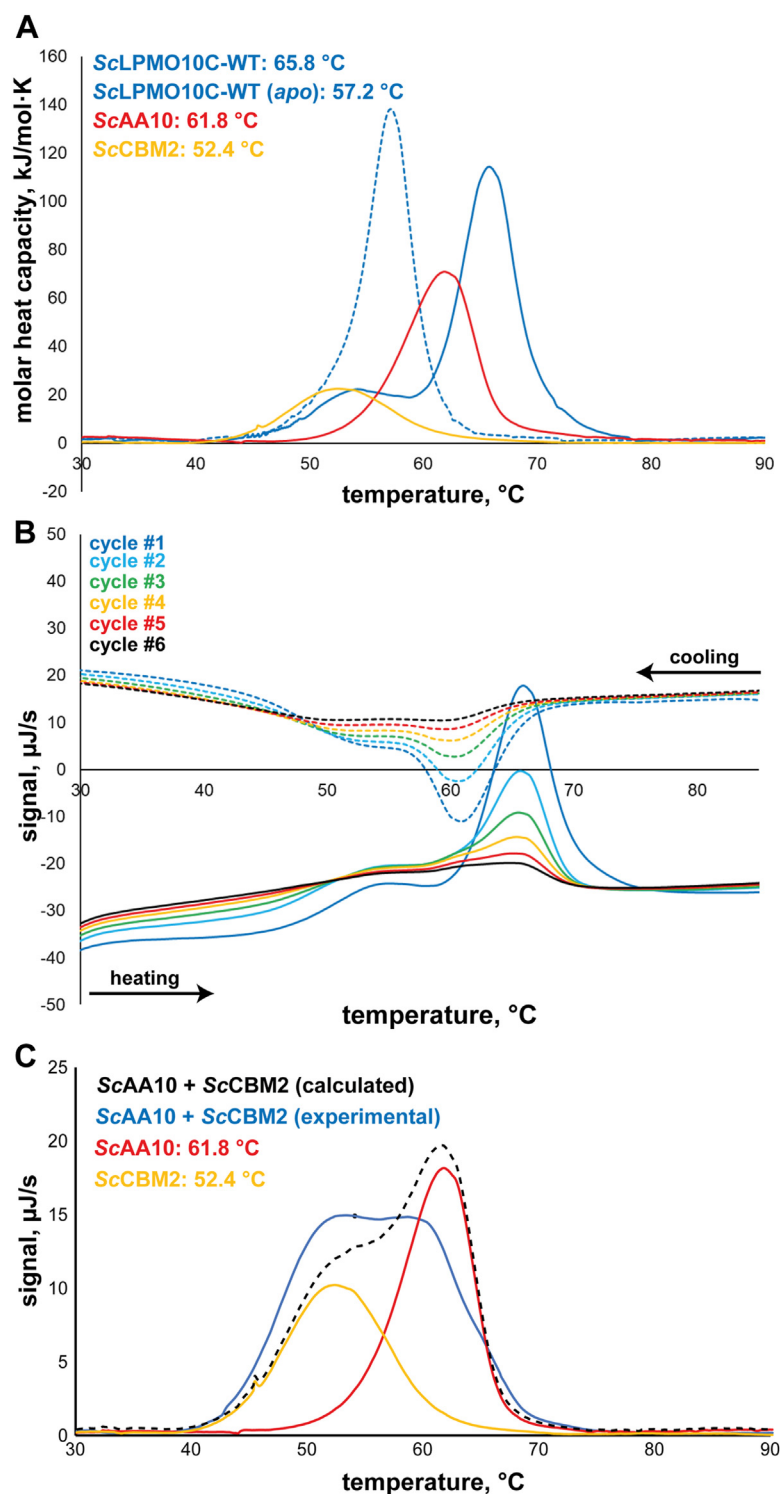
### The linker region affects thermal unfolding and thermostability

Another possible functional role of the linker relates to structural stability that could be affected by linker-mediated



**Figure 5. Catalytic activity of four LPMO variants in the absence or the presence of exogenously added  $H_2O_2$ .** A, progress curves for the formation of soluble oxidized products using 1 mM gallic acid to fuel the reactions. B, similar reaction as in A, but with exogenous  $H_2O_2$  (100  $\mu M$ ) added together with 1 mM ascorbic acid to initiate the reactions. Note the very different time scales in A and B. C, soluble oxidized products (gray bars) and total amount of oxidized products (soluble + insoluble fraction; black bars), determined for the reactions depicted in A at the time points labeled with dashed boxes in A. D, degree of polymerization of the soluble oxidized products (DPox) generated after 48 h in the reactions depicted in A. All reactions were carried out with 0.5  $\mu M$  LPMO and 10 g/l Avicel in 50 mM sodium phosphate buffer, pH 6.0. The error bars show  $\pm SD$  ( $n = 3$ ).  $H_2O_2$ , hydrogen peroxide; LPMO, lytic polysaccharide monoxygenase.

## Structural and functional roles of the linker in an LPMO



**Figure 6. DSC thermograms showing unfolding of ScLPMO10C.** A, thermal unfolding for the WT enzyme, the catalytic domain (ScAA10) and ScCBM2, and their respective calculated apparent  $T_m$  values. The molar heat capacity of the protein samples after buffer baseline subtraction is plotted against the temperature. A melting curve of *apo*-enzyme (i.e., copper free) was also recorded (dashed thermogram) showing an expected decrease in apparent  $T_m$ . The shoulder in the curve for *holo*-ScLPMO10C likely represents a small fraction of *apo*-enzyme. B, multiple unfolding and refolding cycles for ScLPMO10C showing some degree of reversibility and a consistent apparent  $T_m$ . Fig. S6 shows similar data for ScAA10 and ScCBM2. C, thermal unfolding of the individual domains, alone, or when combined; the *dashed curve* shows an unfolding curve calculated by summing the curves obtained for the individual domains. B, raw and unprocessed data in which buffer baseline was not subtracted from the signal. All experiments were conducted in sodium phosphate buffer, pH 6.0, at a protein concentration of 1 g/l (ScAA10 and ScCBM2) or 2.5 g/l (ScLPMO10C). The experiment with both ScAA10 and ScCBM2 in C contained 1 g/l of each protein. DSC, differential scanning calorimetry.



interdomain interactions. Therefore, we used differential scanning calorimetry (DSC) to investigate the conformational stability of ScLPMO10C and its isolated domains, that is, ScAA10 and ScCBM2. Figure 6 shows that the apparent melting temperature,  $T_{m(\text{app})}$ , of WT ( $T_{m(\text{app})} = 65.8\text{ }^{\circ}\text{C}$ ) is higher than for ScAA10 ( $T_{m(\text{app})} = 61.8\text{ }^{\circ}\text{C}$ ) and ScCBM2 ( $T_{m(\text{app})} = 52.4\text{ }^{\circ}\text{C}$ ). Importantly, the full-length enzyme showed a single transition (Fig. 4, A and B), whereas DSC scans of a mixture of ScAA10 and ScCBM2 (Fig. 6C) showed a double transition, at lower temperature compared with the full-length enzyme, with the signal amounting to the sum of the signals obtained for the individual domains. Together, these results clearly show that the linker affects structural stability and mediates domain interactions in the full-length enzyme.

Because of low protein yields, DSC could not be used for the engineered LPMO variants, which were therefore assessed using a thermoshift differential scanning fluorimetry (DSF) assay to determine  $T_{m(\text{app})}$  in experiments that included the WT full-length enzymes, ScAA10 and ScCBM2 (Fig. S7). The melting temperatures derived from DSC were slightly higher (+1–2  $^{\circ}\text{C}$ ) than the values obtained by DSF, but, generally, the  $T_{m(\text{app})}$  values followed the same trend (Table 2). While SR displayed a very similar  $T_{m(\text{app})}$  compared with WT, the two other variants (SL and EL) showed a somewhat lower  $T_{m(\text{app})}$ . In all cases, removal of the copper by adding EDTA reduced the  $T_{m(\text{app})}$  by  $\sim 9\text{ }^{\circ}\text{C}$ , except for ScCBM2, which has no copper site, and ScAA10 domain for which the difference in  $T_{m(\text{app})}$  was as large as  $\sim 15\text{ }^{\circ}\text{C}$ . The latter is remarkable and underpins the notion that the stability of this domain is affected by interactions with the linker and/or CBM.

Encouraged by the apparent ability of ScLPMO10C to withstand heat treatment, we then carried out a final test of how the linkers affect protein thermal stability. The various proteins were boiled for 0 to 120 min prior to starting LPMO reactions by adding Avicel and ascorbic acid (Figs. 7 and S8A) or carrying out a binding assay (for ScCBM2; Fig. S8B). Of note, the residual activity assays are not fully quantitative since LPMO catalysis in these standard assays is limited by the *in situ* generation of the cosubstrate,  $\text{H}_2\text{O}_2$ , and not linearly dependent on the concentration of active enzyme (45, 46). Figure 7 shows that for all linker variants, enzyme inactivation was not prominent after 60 min of preincubation, and three of the four variants behave similarly (Fig. 7B). In line with its higher  $T_{m(\text{app})}$  value, SR, containing a linker with a sequence

that is very different from all others, stands out by showing barely any signs of inactivation after 60 min (Fig. 7B).

Binding studies with heat-pretreated ScCBM2 showed that the unfolding of this domain is reversible, since the ability to bind cellulose was retained (Fig. S8B). Studies with ScAA10 showed that this domain is more sensitive to heat treatment compared with the full-length enzyme, with almost no activity retained after 60 min of incubation (Fig. S8A). This adds to the notion that both the linker and CBM2 increase the structural integrity and stability of ScLPMO10C.

## Discussion

Using a combination of SAXS, NMR, and MD simulations, we determined global structural features of full-length ScLPMO10C with its WT linker (Figs. 2, 3, S1–S3 and Table 1). The present results are a considerable improvement over the previous model of this enzyme (13), as we obtained a more accurate determination of the overall size in terms of  $R_g$  and  $R_h$  and a vastly improved coverage of the dynamic features of the linker region in the picosecond–nanosecond time scale, which were probed by  $^{15}\text{N}$  relaxation. The data show a linker that is flexible but, at the same time, favor an extended conformation.

After using SAXS data to optimize the Martini force field, we were able to generate conformational ensembles of ScLPMO10C with four different linkers using MD simulations. The force field modification consisted of a 10% increase of the protein–water interactions, which is in accordance with the 6 to 10% range found by Thomasen *et al.* (37) when optimizing the agreement between SAXS data and simulations of 12 IDPs and three multidomain proteins using the Martini force field. The results of the MD simulations (Table 1) neither did show a correlation between linker length and the calculated radius of gyration of the complete protein nor was there a good correlation between the calculated radius of gyration and the hydrodynamic radius determined by NMR. Next to revealing limitations of either the MD simulations or the NMR-based measurements of  $R_h$ , these observations suggest that variations in linker structure impact the spacing between the two domains.

Sequence-dependent structural variation in the linker was assessed by simulations with a simpler single-bead model to analyze a large collection of natural linkers, which is an approach shown to be useful for assessing the conformation of flexible proteins (40). This analysis clearly showed that the  $R_g$  of the 164 natural linker sequences analyzed depends mostly on the number of amino acids and not on their identity, although the model encodes sequence specificity. This observation finds explanation in the low complexity of the linker sequences (Fig. 4E), which are predominantly rich in glycine, threonine, and proline. This finding contrasts recent work on the conformational buffering of disordered linkers in the adenovirus early gene 1A (E1A) protein (47). The end-to-end distances of E1A linkers appear to be under functional selection, and variations in sequence length and composition are compensated for to maintain an optimal end-to-end distance

**Table 2**  
Apparent melting temperatures measured by DSC and DSF

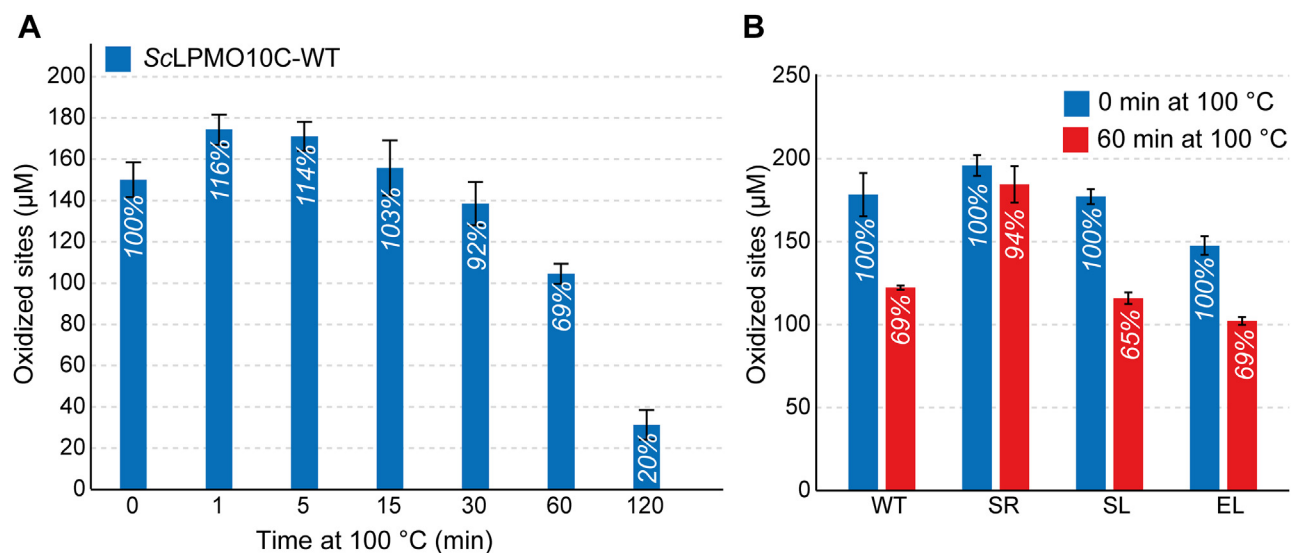
Enzyme variant	$T_{m(\text{app})}$ obtained by DSC <sup>a</sup> (holo/apo)	$T_{m(\text{app})}$ obtained by DSF <sup>b</sup> (holo/apo)
ScLPMO10C	65.8 $^{\circ}\text{C}$ /57.2 $^{\circ}\text{C}$	64.2 $\pm$ 0.2 $^{\circ}\text{C}$ /55.6 $\pm$ 0.2 $^{\circ}\text{C}$
ScAA10	61.8 $^{\circ}\text{C}$ /N/A	61.3 $\pm$ 0.5 $^{\circ}\text{C}$ /46.4 $\pm$ 0.6 $^{\circ}\text{C}$
ScCBM2	52.4 $^{\circ}\text{C}$ /N/A	51.3 $\pm$ 0.4 $^{\circ}\text{C}$ /52.9 $\pm$ 0.6 $^{\circ}\text{C}$
ScLPMO10C-SR	N/A/N/A	64.6 $\pm$ 0.1 $^{\circ}\text{C}$ /56.0 $\pm$ 0.2 $^{\circ}\text{C}$
ScLPMO10C-SL	N/A/N/A	62.6 $\pm$ 0.2 $^{\circ}\text{C}$ /54.5 $\pm$ 0.3 $^{\circ}\text{C}$
ScLPMO10C-EL	N/A/N/A	60.1 $\pm$ 0.4 $^{\circ}\text{C}$ /53.9 $\pm$ 0.1 $^{\circ}\text{C}$

Abbreviation: N/A, not available.

<sup>a</sup> Data derived from Figure 6A.

<sup>b</sup> Data derived from Fig. S7.

## Structural and functional roles of the linker in an LPMO



**Figure 7. Residual activity after boiling.** *A*, cellulose oxidation by ScLPMO10C that had been preexposed to boiling for various amounts of time. *B*, enzyme activity of the four linker types in reactions that were not heat exposed and in reactions where the enzymes were boiled for 60 min. The reactions were carried out in 50 mM sodium phosphate buffer (pH 6.0) in a thermomixer set to 30 °C and 800 rpm. The error bars show  $\pm$ SD ( $n = 3$ ).

that maximizes binding to the E1A interaction partner (the retinoblastoma protein) (47).

Importantly, this analysis of natural linkers leads to the conclusion that the aforementioned lack of correlation likely is not because of variation in linker structure and, thus, may be due to variation in interactions that involve the two protein domains. Such interactions also became apparent from the functional characterization of the enzyme variants, as discussed later.

We assessed multiple potential functional roles of the linker. Measurements of catalytic performance (Fig. 5, *A* and *B*) showed that the two variants containing equally long linkers with very different sequences have equal performance. The variants with a shorter linker or longer linker showed signs of increased enzyme inactivation, which, in the case of LPMOs, is indicative of less optimal enzyme–substrate interactions. It is conceivable that linker length affects the product distribution, since the linker restricts the area that the catalytic domain can reach when anchored to the substrate through the appended CBM. However, differences in product distributions were not detected (Fig. 5, *C* and *D*). Assessment of thermal stability by DSF showed that the linker variant with WT-like catalytic performance, SR, was also as equally stable as the WT, whereas the two poorer performing variants, EL and SL, demonstrated slightly reduced apparent  $T_m$  values. Importantly, all unfolding curves showed a single transition, which indicates that the AA10 and CBM2 domains interact in a manner that is mediated by the linker. The interactions between the domains and the role of the linker are supported by the DSC studies of the WT enzyme (Fig. 6), which show (1) that the full-length enzyme has a single unfolding transition with an apparent  $T_m$  that is higher than the  $T_m$  of the unfolding transitions observed for each of the individual domains and (2) that a mixture of the two domains has two folding transitions yielding an unfolding curve that represents the sum of the

unfolding curves for the individual domains (Fig. 6C). Of note, the final experiments, in which we subjected the LPMO to boiling prior to an activity assay (Figs. 7 and S7), showed that the isolated catalytic domain, ScAA10, is less stable than the full-length enzyme, adding to the notion that the presence of the linker and the CBM add to overall enzyme stability.

Our observations suggest that evolutionary pressure appears to shape linker length in LPMOs with two-domain architectures similar to ScLPMO10C, rather than influencing amino acid composition. In line with our observations, a study of linkers in cellulases also indicated a consistent linker length (34–35 amino acids) in bacterial GH6 enzymes with CBM2s independent of the location of the CBM (N- or C-terminal regions) (7).

Although our assessment of four linker variants indicates that linker length is most important and, in this study, a length of 34 residues was shown to be optimal for ScLPMO10C, our analysis of 164 linkers shows that there is no universally optimal functional linker for LPMOs with the domain organization of ScLPMO10C (*i.e.*, AA10-linker-CBM2). The linker in ScLPMO10C acts as a flexible spacer, giving the protein a dumbbell shape (Figs. 1 and 3A) where the individual domains are kept apart while preserving independent motions. At the same time, all other data show that linker-mediated domain interactions play a role. This is supported by a previous study on an LPMO produced by *Hypocrea jecorina* (HjLPMO9A), which showed that specific portions of the linker interact with the catalytic domain (48). An ideal functional linker would have a length and composition that favors stabilizing interactions either between the linker and domains or between the domains themselves. This is underscored by the observation of ScLPMO10C resilience, maintaining activity even after boiling (Fig. 7). Moreover, the variant with the SR linker with the same length as the WT exhibited even higher resistance to boiling, hinting at inherent stabilizing interactions for these linker lengths.

Linker variability may also relate to hitherto undetected variation in LPMO substrate specificity. Cellulosic substrates exist in multiple forms, and cellulose crystals have multiple faces with differing morphologies (49). Our activity data (Fig. 5, A and B) show that the linker length affects how well the LPMO interacts with its substrate. It is thus conceivable that natural variation in LPMO linker length reflects that these LPMOs have different optimal cellulose substrates.

This study found that, despite being extended and flexible, the linker affects interactions between the domains and with the substrate (as shown by the different stabilities under turnover conditions) and structural stability (as shown by the analysis of unfolding and thermal stability). So, while structural data seem to show that the linker is “just” an extended tether, the functional data show that the linker is more than that. These insights carry pivotal implications for biotechnological applications. Beyond traditional biomass degradation, where LPMOs already have recognized potential (50, 51), there is an emerging interest in the role of LPMOs in cellulose defibrillation processes for nanocellulose production. In this context, the presence of a CBM and linker has been shown to affect the carboxyl content in cellulose nanofibrils produced using LPMOs (52). Future endeavors in linker-oriented protein engineering could leverage these insights to develop or optimize enzymes tailored for specific applications.

## Experimental procedures

### Cloning, expression, and purification of ScLPMO10C and linker variants

WT ScLPMO10C and its truncated versions (ScAA10 and ScCBM2) were cloned and produced as previously described (13, 28) and loaded with copper according to the protocol described (53). The pRSET B expression vector harboring ScLPMO10C (UniProt ID: Q9RJY2) (28) was used as a starting point to replace the linker region (residues 229–262) with (i) a 34 amino acid polyserine linker from *Cellvibrio japonicus* LPMO10A (CjLPMO10A; UniProt ID: B3PJ79, residues 217–250) (10), (ii) a 20 amino acid linker from *Streptomyces scabies* LPMO10B (SscLPMO10B; UniProt ID: C9YVY4, residues 238–257) (54), and (iii) an EL of 59 amino acids from *Caldibacillus cellulovorans* LPMO10A (CcLPMO10A; UniProt ID: Q9RFX5, residues 226–284). The sequences of these linkers are provided in Figure 1 and Table S1. DNA encoding for the alternative linkers was derived from gene sequences that were codon optimized for expression in *Escherichia coli*. The pRSET B\_sclpmo10c-wt vector was amplified using a forward primer annealing to the start of the *cbm2* sequence (Forward-primer; 5'-GGTTCGTGTATGGCCGTCTATA-3') and a reverse primer that anneals to the end of the *aa10* sequence (Reverse primer; 5'-GTCGAAAACCACATCGGAGCAA-3'), thereby amplifying the entire 2.8 kb vector excluding the 102 nucleotides encoding the linker. The amplified vector was gel purified (1% agarose) and used for In-Fusion HD cloning (Clontech) with amplified and gel purified (2% agarose) linker inserts. The size of the inserts varied between 90 and 207 nucleotides. After fusing linker inserts to the

linearized vector, One Shot TOP10 chemically competent *E. coli* cells (Invitrogen) were transformed. Positive transformants were subsequently verified by Sanger sequencing (Eurofins GATC).

Expression of <sup>13</sup>C- and <sup>15</sup>N-labeled ScLPMO10C for chemical shift assignment and <sup>15</sup>N relaxation measurements was performed using the XylS/*Pm* expression system (55). The pJB\_SP\_sclpmo10c-wt vector was constructed as previously described (55), harboring ScLPMO10C downstream of a pelB signal peptide.

For regular protein production, BL21(DE3) T7 Express competent *E. coli* (New England Biolabs catalog number 2566) were transformed using a heat-shock protocol and grown at 37 °C on LB-agar plates. All media were supplemented with 100 µg/ml ampicillin. Precultures were made in shaking flasks by inoculating 5 ml LB medium (10 g/l tryptone, 5 g/l yeast extract, and 5 g/l NaCl) with recombinant cells, followed by incubation at 30 °C and 225 rpm overnight. Main cultures were made by inoculating 500 ml 2× LB medium (20 g/l tryptone, 10 g/l yeast extract, and 5 g/l NaCl) with 1% pre-culture, followed by incubation in a LEX-24 bioreactor (Epiphyte3) using compressed air (20 psi) for aeration and mixing. The SL, SR, and EL variants were incubated at 30 °C for 24 h, whereas the WT variant was incubated at 30 °C to an absorbance of ≈0.8 at 600 nm. The latter culture was cooled on ice for 5 min, induced with 0.1 mM *m*-toluic acid, and further incubated at 16 °C for 20 h.

For the production of isotopically labeled protein, main cultures were made by inoculating 500 ml M9 medium (6 g/l Na<sub>2</sub>HPO<sub>4</sub>, 3 g/l KH<sub>2</sub>PO<sub>4</sub>, and 0.5 g/l NaCl) supplemented with (<sup>15</sup>NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (98% <sup>15</sup>N), 4 g/l unlabeled or <sup>13</sup>C-labeled glucose (98% uniformly labeled with <sup>13</sup>C), 10 ml Bioexpress Cell Growth Media (Cambridge Isotope Laboratories), 5 ml Gibco MEM Vitamin Solution (100×), and trace metal solution (300 mg/ml MgSO<sub>4</sub>, 2 mg/l ZnSO<sub>4</sub>, 10 mg/l FeSO<sub>4</sub>, 2 mg/l CuSO<sub>4</sub>, and 20 mg/l CaCl<sub>2</sub>) with 1% preculture, followed by incubation at 30 °C to an absorbance at 600 nm ≈0.8. The culture was cooled on ice for 5 min, induced with 0.1 mM *m*-toluic acid, and further incubated at 16 °C for 20 h.

Cells were harvested by centrifugation for 5 min at 5500g, and 4 °C, and periplasmic fractions were prepared by the osmotic shock method, as follows: the pellet was resuspended in 30 ml spheroplast buffer (pH 7.5, 100 mM Tris-HCl, 500 mM sucrose, and 0.5 mM EDTA) with a tablet of cComplete ULTRA protease inhibitor (Roche), followed by centrifugation for 10 min at 6150g and 4 °C. The pellet was incubated at room temperature for 10 min, prior to resuspension in 25 ml ice-cold water with another tablet of cComplete ULTRA protease inhibitor, followed by centrifugation for 30 min at 15,000g and 4 °C. The supernatant was filtered through a 0.2 µm pore size filter prior to protein purification. The proteins were purified by loading periplasmic extracts in buffer A (50 mM Tris-HCl, pH 8.5) onto a 5 ml HiTrap DEAE Sepharose FF anion exchanger (Cytiva) connected to an ÄKTA Pure FPLC system (Cytiva). LPMOs were eluted by applying a linear salt gradient toward 50 mM Tris-HCl, pH 8.5, 1 M NaCl (buffer B), at a flow rate of 5 ml/min. The LPMOs started to elute at 10 to

## Structural and functional roles of the linker in an LPMO

12% buffer B. The fractions containing LPMO were pooled and concentrated using ultrafiltration spin-tubes (10 kDa cutoff; Sartorius). The protein-containing fractions were analyzed by SDS-PAGE.

Before SAXS measurements, proteins were further purified by size-exclusion chromatography. Samples were loaded onto a HiLoad 16/600 Superdex 75 pg size-exclusion column (Cytiva), using 50 mM Tris-HCl, pH 7.5, 200 mM NaCl as running buffer and a flow rate of 1 ml/min.

### SAXS

ScLPMO10C was measured by SAXS at the P12 beamline at DESY (56). The scattering intensity,  $I(q)$ , was measured as a function of  $q = 4\pi \sin \theta/\lambda$ , where  $q$  is the scattering vector,  $2\theta$  is the scattering angle, and  $\lambda$  is the X-ray wavelength (1.24 Å at P12). Buffer backgrounds were measured before and after each sample, and all measurements were performed at 8 °C. The initial data reduction including azimuthal averaging, frame averaging, and background subtraction was performed using the PRIMUS program (57) that is part of the ATSAS software (EMBL Hamburg, <https://www.embl-hamburg.de/biosaxs/download.html>) (58). Data were converted into the usual units of scattering cross section per unit volume,  $\text{cm}^{-1}$ , using the well-known scattering cross sections of, respectively, water and bovine serum albumin as external standards (59, 60). Furthermore, the data were logarithmically rebinned with a bin factor of 1.02 using the WillItRebin program (61). Pair-wise distance distribution functions were calculated using the Bayesian Indirect Fourier Transformation algorithm implemented in BayesApp (Niels Bohr Institute, University of Copenhagen, <https://github.com/ehb54/GenApp-BayesApp>) (62).

### NMR

All NMR spectra were recorded in an NMR buffer (25 mM sodium phosphate and 10 mM NaCl, pH 5.5) containing 10% D<sub>2</sub>O (or 99.9% D<sub>2</sub>O for <sup>13</sup>C-detected experiments; see later) at 25 °C using a Bruker Ascend 800 MHz spectrometer with an Avance III HD (Bruker Biospin) console equipped with a 5 mm z-gradient CP-TCI (H/C/N) cryogenic probe, at the NV-NMR-Centre/Norwegian NMR Platform at the Norwegian University of Science and Technology. NMR data were processed and analyzed using Bruker TopSpin version 3.5, and Protein Dynamic Center software version 2.7.4 from Bruker BioSpin.

We performed a new chemical shift assignment of full-length ScLPMO10C using previously published (29) backbone assignments for CBM2 and AA10 as the starting point, together with newly acquired data from the following experiments. Resonances for nonproline amino acids, that is, possessing amide protons, were assigned by using <sup>15</sup>N-heteronuclear single quantum coherence, HNCA, HNCO, and CBCA(CO)NH. The <sup>13</sup>C-detected experiments, CON, CANCO, and CACO, were used to assign backbone resonances for prolines in the linker region (63).

Secondary structure elements in the linker region were analyzed using the web-based version of the TALOS-N

software (<http://spin.niddk.nih.gov/bax/software/TALOS-N/>) (64) using the <sup>13</sup>C and <sup>15</sup>N chemical shifts.

Nuclear spin relaxation rates  $R_1$  and  $R_2$  and heteronuclear <sup>1</sup>H-<sup>15</sup>N NOE measurements of amide <sup>15</sup>N were recorded as pseudo-3D spectra where two frequency dimensions correspond to the amide <sup>1</sup>H and <sup>15</sup>N chemical shifts, respectively, and the third dimension is made up of variable relaxation time delays. For  $R_1$ , the time points were 100, 200, 500, 1000, 1500, 2000, 3000, and 4000 ms. For  $R_2$ , the time points were 17, 34, 102, 170, 204, and 238 ms. For <sup>1</sup>H-<sup>15</sup>N NOE, two 2D planes were recorded, one with and one without presaturation. The generalized order parameter,  $S^2$ , was obtained using reduced spectral density mapping (65).

To calculate the translational diffusion coefficients,  $D_t$ , of WT, SR, SL, and EL, we used a pulsed-field gradient stimulated-echo sequence with 3-9-19 water suppression (stebpgp1s19) (66). A linear gradient ramp of 32 points in the range 2 to 95% of the gradient strength (where 100% corresponds to 53.5 G/cm) was used, and the gradient length and delay between gradients were set to  $\delta = 2$  ms and  $\Delta = 150$  ms, respectively. While it is in principle possible to calculate the absolute hydrodynamic radius,  $R_h$ , by using the Stokes-Einstein equation (67), this requires accurate knowledge of the viscosity of the sample solution, which is difficult to measure. An alternative approach for calculating  $R_h$  consists of using a reference molecule in the solution, with a known  $R_h$ . Here, 1% dioxane was used as a reference because it does not interact with proteins, and its  $R_h$  is known ( $R_{h,ref} = 2.12$  Å) (68). The diffusion coefficient ( $D_{t,ref}$ ) of dioxane was thus used to determine the  $R_h$  of the proteins:

$$R_h = \frac{D_{t,ref}}{D_t} R_{h,ref}$$

### CG simulations using the Martini force field

The CG Martini version 3.0.beta.4.17 force field was used in combination with GROMACS 2018.2 (<https://www.gromacs.org/about.html>) (69) to simulate ScLPMO10C with different linkers. The constructs were CG using the Martinize2 program (70). An elastic network model (71) was used to constrain the overall structure of the AA10 and CBM2 domains, but it was excluded from the linker region. The simulation box was filled with water beads, and the system was neutralized with beads corresponding to Na<sup>+</sup> and Cl<sup>-</sup> ions to an ionic strength of 0.15 M. The complex was energy-minimized using a steepest-descent algorithm (100 steps, 0.03 nm maximum step size) prior to being relaxed for 1 ns, with a time step of 5 fs, using the velocity-rescale thermostat (72), Parrinello-Rahman barostat (73), and Verlet cutoff scheme (74). Simulations were performed on the relaxed models with a time step of 20 fs, using the velocity-rescale thermostat, Parrinello-Rahman barostat, and Verlet cutoff scheme in the isothermal-isobaric (NPT) ensemble.

Proteins simulated using Martini force fields have been observed to aggregate, possible because of protein-protein

interaction strengths being too high relative to protein–water interactions (25, 37), leading to mismatches between experimental data and simulations. To circumvent this problem, we followed the approach described by Larsen *et al.* (25) to modify the Martini version 3.0.beta.4.17 forcefield to improve the fit between SAXS data and small-angle CG simulations. In summary, the interaction strength (*i.e.*,  $\epsilon$  parameter in the Lennard–Jones potential) between protein and water beads in the Martini force field was multiplied by a factor  $\lambda$ , ranging from 1.0 (unchanged) to 1.20 (20% increase of the protein–water interaction strength). Simulations for each value of  $\lambda$  were run for 10  $\mu$ s, and frames were written every 1 ns for each trajectory. To facilitate calculation of SAXS profiles, each of the 10,000 frames per CG simulations was used to reconstruct atomistic models by using the Backward program (75).

### Calculating SAXS profiles from conformational ensembles

The implicit solvent SAXS calculation program Pepsi-SAXS version 3.0 (76) was used to calculate SAXS profiles from the conformational ensemble (*i.e.*, the atomic coordinates of the 10,000 frames) of ScLPMO10C. Input for Pepsi-SAXS comprises atomic coordinates, experimental SAXS profiles, and four parameters that can either be provided directly or determined by the program while fitting between the calculated and experimental SAXS profiles: the scale of the profiles,  $I(0)$ , a constant background, *cst*, the effective atomic radius,  $r_0$ , and the contrast of the hydration layer,  $\delta_p$ . The approach used here was based on the methods previously described by Larsen *et al.* (25) and Pesce and Lindorff-Larsen (36). The parameter values were set to  $I(0) = 1.0$ , *cst* = 0.0,  $r_0 = 1.65$  Å, and  $\delta_p = 3.34$  e/nm<sup>3</sup>. Then, the iterative Bayesian maximum entropy method (36) was used to iteratively rescale and shift the calculated SAXS profiles, while reweighting the conformational ensemble to fit a global value of  $I(0)$  and *cst*.

The implicit solvent SAXS calculation program CRY SOL (77) was used to calculate a SAXS profile for the structure (PDB code: 4OY7) of ScAA10. The search limits for the fitting parameters *dro* (optimal hydration shell contrast), *Ra* (optimal atomic group radius), and *ExVol* (relative background) were set to [0.000–0.075], [1.40–1.80], and [23,744–27,594], respectively. The parameters for maximum order of harmonics were set to 15, the order of Fibonacci grid to 17, and the electron density of the solvent was set to 0.33 e/Å<sup>3</sup>.

### Linker simulations using a single-bead model

To find linker sequences similar to the linker in WT ScLPMO10C, we used *phmmer* in the HMMER webserver (78) to search the UniProt KB Reference Proteomes (38) database against the Pfam (79) profile of ScLPMO10C, that is, AA10-linker-CBM2. Using this approach, we found 160 unique sequences with the same domain architecture as ScLPMO10C. For each retrieved sequence, the AA10 and CBM2 domains were assigned using *hmmscan*, and the linker region was assigned as the amino acids in between the domains.

We added the four linker sequences of the WT enzyme, SR, SL, and EL to the retrieved sequences and simulated the 164

sequences at 298 K, I = 0.1 M, and pH = 6.5 using a model where each amino acid in the sequence is represented by a single bead placed at the C  $\alpha$  coordinates. The model has previously been optimized and validated against experimental SAXS and paramagnetic relaxation enhancement NMR data of 48 IDPs of sequence length ranging from 24 to 334 residues (40). In particular, for the amino acid–specific “stickiness” parameters, we used the M1 set proposed by Tesei *et al.* (40). With the aim of mimicking end-capped proteins, we did not modify the charges of the terminal residues, as opposed to other implementations of the model where the charge of the terminal carboxylate and amino groups is considered. Langevin simulations were performed using HOOMD-blue version 2.9.3 (80) using a time step of 5 fs, a friction coefficient of 0.01 ps<sup>−1</sup>, and a cutoff of 4 nm for both Debye–Hückel and nonionic nonbonded interactions. Each simulation started from the fully extended chain placed in a cubic box of 200 nm, and we extensively sampled chain conformations for a total simulation time of 14  $\mu$ s.

The persistence length was calculated by fitting the auto-correlation function of the bond vectors along the sequence to a single exponential decay:

$$\langle b_i \bullet b_{i+n} \rangle = \exp\left(-\frac{nl_b}{l_p}\right)$$

where  $b_i$  and  $b_{i+n}$  are bond vectors separated by  $n$  bonds,  $l_b = 0.38$  nm (the equilibrium bond length), and  $l_p$  is the persistence length.

The scaling exponent was calculated by fitting the long-distance region,  $|i - j| > 10$ , of the average intramolecular pairwise distances,  $R_{ij}$ , to the following equation (81)

$$R_{ij} = l_p |i - j|^\nu$$

The  $R_g$  values of the linkers were estimated from the sequence length *via* the following equation (39)

$$R_g = \sqrt{\frac{\gamma(\gamma+1)}{2(\gamma+2\nu)(\gamma+2\nu+1)}} b N^\nu$$

where  $\gamma = 1.1615$ ,  $b = 0.55$  nm, and  $\nu$  is the average scaling exponent for the 164 linker sequences.

### LPMO activity assays

Enzyme activity assays were set up using 10 g/l Avicel, 0.5  $\mu$ M LPMO, 1 mM ascorbic acid, or 1 mM gallic acid (in 100% dimethyl sulfoxide) in 50 mM sodium phosphate buffer, pH 6.0, in the absence or the presence of 100  $\mu$ M H<sub>2</sub>O<sub>2</sub>, and were incubated for up to 48 h in a thermomixer set to 30 °C. Aliquots were taken at various time points and vacuum filtered through a 0.45  $\mu$ m membrane to separate the solid fraction. The supernatants (*i.e.*, the soluble fractions) were further treated with *Tf*Cel6A (E2; (82)) to convert longer oxidized cello-oligosaccharides to oxidized dimers and trimers, which can be quantified by high-performance anion-exchange chromatography with pulsed

## Structural and functional roles of the linker in an LPMO

amperometric detection (HPAEC-PAD), using in-house made standards (83). To determine the total concentration of oxidized products (*i.e.*, soluble and insoluble), the LPMOs were inactivated by boiling the reaction mixture for 15 min at 100 °C. Subsequently, the reaction mixtures were cooled on ice and diluted in buffer to a final concentration of 2 g/l Avicel including 5  $\mu$ M *Tj*Cel6A and 2  $\mu$ M *Tj*Cel5A (84), followed by 48 h incubation at 50 °C in an Eppendorf Thermomixer with shaking at 800 rpm. As a result of this procedure, all oxidized sites are solubilized as oxidized dimers and trimers, which were quantified by HPAEC-PAD (see later).

Product distribution profiles were assessed by HPAEC-PAD using samples of LPMO-generated soluble products that had not been treated with cellulases. The peak areas (nCxmin) of oxidized LPMO products were used to calculate apparent ratios of oxidized oligosaccharides with different degrees of polymerization (DP2-DP6).

### Residual activity and cellulose binding of heat-treated proteins

To test whether heat-treated and potentially refolded LPMOs retained their functional properties, solutions of WT and *ScAA10* at a concentration of 2  $\mu$ M in 50 mM sodium phosphate buffer, pH 6.0, were boiled for 0 to 120 min and then rapidly cooled on ice prior to starting the LPMO reaction by addition of reductant (1 mM ascorbic acid) and substrate (10 g/l Avicel). After incubation of the reactions for 24 h at 30 °C, *Tj*Cel6A was added to filtrated supernatants, and the oxidized dimers and trimers were quantified using HPAEC-PAD to determine residual LPMO activity.

The absorbance at 280 nm of a solution with 0.2 g/l CBM2 in 50 mM sodium phosphate buffer, pH 6.0, was measured prior to boiling this solution for 0 to 120 min. After cooling the samples on ice, the proteins were filtrated using a 96-well filter plate (Millipore) to remove precipitated protein, and the absorbance at 280 nm was remeasured before starting the binding assay. Importantly, the loss in protein concentration because of precipitation was barely noticeable. For the binding assay, reactions were set up containing 0.1 g/l of heat-pretreated protein and 10 g/l Avicel in 50 mM sodium phosphate buffer, pH 6.0, and incubated at 800 rpm and 22 °C. After 1 h of incubation, the samples were filtrated and absorbance at 280 nm was measured again to estimate the concentration of unbound CBM2.

Control reactions containing (i) 2 mM ascorbic acid in buffer or (ii) 2  $\mu$ M *ScLPMO10C* and 2 mM ascorbic acid in buffer were boiled for 15 min prior to starting reactions by adding *ScLPMO10C* and Avicel (i) or Avicel only (ii). In the end, all reactions contained 10 g/l Avicel, 1 mM ascorbic acid, 1  $\mu$ M *ScLPMO10C* in 50 mM sodium phosphate buffer, pH 6.0. The reactions were incubated overnight followed by filtration, treatment with *Tj*Cel6A, and quantification of oxidized products by HPAEC-PAD.

### HPAEC-PAD

HPAEC-PAD analysis of LPMO-generated products was carried out using an ICS-5000 system from Dionex equipped

with a disposable electrochemical gold electrode as previously described (85). Samples of 5  $\mu$ l were injected onto a CarboPac PA200 (3  $\times$  250 mm) column operated with 0.1 M NaOH (eluent A) at a flow rate of 0.5 ml/min and a column temperature of 30 °C. Elution was achieved using a stepwise gradient with increasing amounts of eluent B (0.1 M NaOH + 1 M NaOAc), as follows: 0 to 5.5% B over 3 min; 5.5 to 15% B over 6 min; 15 to 100% B over 11 min; 100 to 0% B over 0.1 min; and 0% B (reconditioning) for 5.9 min. Chromatograms were recorded and analyzed by peak integration using Chromeleon 7.0 software (Thermo Fisher Scientific).

### Apparent melting temperatures

A Nano-Differential Scanning Calorimeter III (Calorimetry Sciences Corporation) was used to determine  $T_{m(app)}$  of *ScLPMO10C* and its individual domains. Solutions containing 1 mg/ml (*ScAA10* and *ScCBM2*) or 2.5 mg/ml (full-length) protein in 50 mM sodium phosphate buffer, pH 6.0 (filtered and degassed), were heated from 25 to 90 °C at 1 °C/min followed by cooling from 90 to 25 °C at the same rate. At least three cycles (*i.e.*, three heating and three cooling steps) were recorded for each run. Buffer baselines were recorded and subtracted from the protein scans unless stated otherwise. The melting curve for the *apo*-form of *ScLPMO10C* was obtained in the same manner by introducing 5 mM EDTA to both the protein sample and the control sample (*i.e.*, sample lacking the enzyme). The data were analyzed with NanoAnalyze software (<https://www.tainstruments.com>).

For the engineered enzyme variants that did not efficiently express, a DSF assay based on the use of SYPRO Orange (Thermoshift assay kit; Thermo Fisher Scientific) was used to minimize protein consumption (86). The quantum yield of the SYPRO Orange dye is significantly increased upon binding to hydrophobic regions of the protein that become accessible as the protein unfolds. The fluorescence emission was monitored using a StepOnePlus real-time PCR machine (Thermo Fisher Scientific). The apparent  $T_m$  was calculated as the temperature corresponding to the minimum value of the derivative plot ( $-d[\text{relative fluorescence unit}]/dT$  versus  $T$ ; Fig. S7). Solutions containing 0.1 g/l protein and SYPRO Orange (1 $\times$ ) in 50 mM sodium phosphate buffer, pH 6.0, were heated in a 96-well plate from 25 to 95 °C, over 50 min. For each protein, the experiment was carried out in quadruplicates (*i.e.*,  $n = 4$ ).

### Data availability

NMR chemical shift assignments have been deposited in the BioMagnetic Resonance Databank under the ID 27078. All the data, input files, and code required to reproduce the simulation results reported in this article, including the protein ensembles, are available online at <https://github.com/gcourtade/papers/tree/master/2023/ScLPMO10C-linkers>.

*Supporting information*—This article contains supporting information.

**Acknowledgments**—Part of the computations were performed on resources provided by Sigma2—the National Infrastructure for High Performance Computing and Data Storage in Norway.

**Author contributions**—Z. F., A. A. S., G. T., Y. W., F. L. A., L. A., V. G. H. E., K. L.-L., and G. C. conceptualization; Z. F., A. A. S., G. T., Y. W., E. B., S. K. K., and G. C. methodology; G. T. and Y. W. software; Z. F., G. T., and G. C. formal analysis; Z. F., A. A. S., G. T., E. B., S. K. K., and G. C. investigation; F. L. A., L. A., V. G. H. E., and K. L.-L. resources; Z. F., A. A. S., G. T., Y. W., E. B., S. K. K., F. L. A., V. G. H. E., K. L.-L., and G. C. writing—original draft; Z. F., A. A. S., G. T., Y. W., E. B., S. K. K., F. L. A., L. A., V. G. H. E., K. L.-L., and G. C. writing—review & editing; Z. F., G. T., and G. C. visualization; Z. F., F. L. A., L. A., V. G. H. E., K. L.-L., and G. C. supervision; Z. F. and G. C. project administration; Z. F., F. L. A., V. G. H. E., K. L.-L., and G. C. funding acquisition.

**Funding and additional information**—This work was funded by The Novo Nordisk Foundation, project numbers NNF18OC0055736 (to Z. F.) and NNF18OC0032242 (to G. C.), and by the Research Council of Norway through the Norwegian NMR Platform infrastructure grant 226244, and research grants 269408 and 262853 (to V. G. H. E.). G. T., Y. W., and K. L.-L. were supported by the BRAINSTRUC structural biology initiative from the Lundbeck Foundation.

**Conflict of interest**—The authors declare that they have no conflicts of interest with the contents of this article.

**Abbreviations**—The abbreviations used are: BME, Bayesian-maximum entropy; CBM, carbohydrate-binding module; CG, coarse-grained; DSC, differential scanning calorimetry; DSF, differential scanning fluorimetry; E1A, early gene 1A; EL, extended linker; GH, glycoside hydrolase; H<sub>2</sub>O<sub>2</sub>, hydrogen peroxide; HPAEC-PAD, high-performance anion-exchange chromatography with pulsed amperometric detection; IDP, intrinsically disordered protein; LPMO, lytic polysaccharide monoxygenase; MD, molecular dynamics; PDB, Protein Data Bank; SAXS, small-angle X-ray scattering; SL, shortened linker.

## References

- Sheldon, R. A. (2016) Green chemistry, catalysis and valorization of waste biomass. *J. Mol. Catal. A Chem.* **422**, 3–12
- Rubin, E. M., Himmel, M. E., Ding, S., Johnson, D. K., and Adney, W. S. (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Nature* **454**, 804–807
- Horn, S. J., Vaaje-Kolstad, G., Westereng, B., and Eijsink, V. G. H. (2012) Novel enzymes for the degradation of cellulose. *Biotechnol. Biofuels* **5**, 45
- Vaaje-Kolstad, G., Westereng, B., Horn, S. J., Liu, Z., Zhai, H., Sørli, M., et al. (2010) An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science* **330**, 219–222
- Agger, J. W., Isaksen, T., Várnai, A., Vidal-Melgosa, S., Willats, W. G. T., Ludwig, R., et al. (2014) Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 6287–6292
- Gilbert, J., and Hazlewood, P. (1993) Review article bacterial cellulases and xylanases. *J. Gen. Microbiol.* **139**, 187–194
- Sammond, D. W., Payne, C. M., Brunecky, R., Himmel, M. E., Crowley, M. F., and Beckham, G. T. (2012) Cellulase linkers are optimized based on domain type and function: insights from sequence analysis, biophysical measurements, and molecular simulation. *PLoS One* **7**, e48615
- Kończakowski, B. M., Schaller, K. S., Sørensen, T. H., Peters, G. H. J., Jensen, K., Krogh, K. B. R. M., et al. (2020) Removal of N-linked glycans in cellobiohydrolase Cel7A from *Trichoderma reesei* reveals higher activity and binding affinity on crystalline cellulose. *Biotechnol. Biofuels* **13**, 136
- Payne, C. M., Resch, M. G., Chen, L., Crowley, M. F., Himmel, M. E., Taylor, L. E., et al. (2013) Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 14646–14651
- Forsberg, Z., Nelson, C. E., Dalhus, B., Mekasha, S., Loose, J. S. M., Crouch, L. I., et al. (2016) Structural and functional analysis of a lytic polysaccharide monoxygenase important for efficient utilization of chitin in *Cellvibrio japonicus*. *J. Biol. Chem.* **291**, 7300–7312
- Mekasha, S., Tuveng, T. R., Askarian, F., Choudhary, S., Schmidt-Dannert, C., Niebisch, A., et al. (2020) A tri-modular bacterial enzyme combining hydrolytic activity with oxidative glycosidic bond cleavage efficiently degrades chitin. *J. Biol. Chem.* **295**, 9134–9146
- Manjeet, K., Madhuprakash, J., Mormann, M., Moerschbacher, B. M., and Podile, A. R. (2019) A carbohydrate binding module-5 is essential for oxidative cleavage of chitin by a multi-modular lytic polysaccharide monoxygenase from *Bacillus thuringiensis* serovar kurstaki. *Int. J. Biol. Macromol.* **127**, 649–656
- Courtade, G., Forsberg, Z., Heggset, E. B., Eijsink, V. G. H., and Aachmann, F. L. (2018) The carbohydrate-binding module and linker of a modular lytic polysaccharide monoxygenase promote localized cellulose oxidation. *J. Biol. Chem.* **293**, 13006–13015
- Forsberg, Z., and Courtade, G. (2023) On the impact of carbohydrate-binding modules (CBMs) in lytic polysaccharide monoxygenases (LPMOs). *Essays Biochem.* **67**, 561–574
- Sørensen, C. S., and Kjaergaard, M. (2019) Effective concentrations enforced by intrinsically disordered linkers are governed by polymer physics. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 23124–23131
- Das, R. K., Ruff, K. M., and Pappu, R. V. (2015) Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **32**, 102–112
- Das, R. K., and Pappu, R. V. (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 13392–13397
- Marsh, J. A., and Forman-Kay, J. D. (2010) Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.* **98**, 2383–2390
- Howard, M. B., Ekborg, N. A., Taylor, L. E., Hutcheson, S. W., and Weiner, R. M. (2004) Identification and analysis of polyserine linker domains in prokaryotic proteins with emphasis on the marine bacterium *Microbulbifer* degradans. *Protein Sci.* **13**, 1422–1425
- Shen, H., Schmuck, M., Pilz, I., Gilkes, N. R., Kilburn, D. G., Miller, R. C., et al. (1991) Deletions of the linker connecting the catalytic and cellulose-binding domains of endoglucanase A (CenA) of *Cellulomonas fimi* alters its conformation and catalytic activity. *J. Biol. Chem.* **266**, 11335–11340
- Tamburrini, K. C., Terrapon, N., Lombard, V., Bissaro, B., Longhi, S., and Berrin, J.-G. (2021) Bioinformatic analysis of lytic polysaccharide monoxygenases reveals the pan-families occurrence of intrinsically disordered C-terminal extensions. *Biomolecules* **11**, 1632
- Thomassen, F. E., and Lindorff-Larsen, K. (2022) Conformational ensembles of intrinsically disordered proteins and flexible multidomain proteins. *Biochem. Soc. Trans.* **50**, 541–554
- Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**, 5656–5664
- Karlsen, M. L., Thorsen, T. S., Johner, N., Ammendrup-Johnsen, I., Erlendsson, S., Tian, X., et al. (2015) Structure of dimeric and tetrameric complexes of the BAR domain protein PICK1 determined by small-angle X-ray scattering. *Structure* **23**, 1258–1270
- Larsen, A. H., Wang, Y., Bottaro, S., Grudin, S., Arleth, L., and Lindorff-Larsen, K. (2020) Combining molecular dynamics simulations with small-angle X-ray and neutron scattering data to study multi-domain proteins in solution. *PLoS Comput. Biol.* **16**, e1007870
- Latham, A. P., and Zhang, B. (2022) Unifying coarse-grained force fields for folded and disordered proteins. *Curr. Opin. Struct. Biol.* **72**, 63–70

## Structural and functional roles of the linker in an LPMO

27. Forsberg, Z., Vaaje-Kolstad, G., Westereng, B., Bunæs, A. C., Stenström, Y., MacKenzie, A., *et al.* (2011) Cleavage of cellulose by a CBM33 protein. *Protein Sci.* **20**, 1479–1483
28. Forsberg, Z., Mackenzie, A. K., Sørli, M., Røhr, Å. K., Helland, R., Arvai, A. S., *et al.* (2014) Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monoxygenases. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8446–8451
29. Courtade, G., Forsberg, Z., Vaaje-Kolstad, G., Eijsink, V. G. H., and Aachmann, F. L. (2017a) Chemical shift assignments for the *apo*-form of the catalytic domain, the linker region, and the carbohydrate-binding domain of the cellulose-active lytic polysaccharide monoxygenase ScLPMO10C. *Biomol. NMR Assign.* **11**, 257–264
30. Jiménez, M. A., Ponte, L., Chaves-Arquero, B., Pantoja-Uceda, D., Roque, A., Suau, P., *et al.* (2018) A CON-based NMR assignment strategy for pro-rich intrinsically disordered proteins with low signal dispersion: the C-terminal domain of histone H1.0 as a case study. *J. Biomol. NMR* **72**, 139–148
31. Alderson, T. R., Lee, J. H., Charlier, C., Ying, J., and Bax, A. (2018) Propensity for cis-proline formation in unfolded proteins. *Chembiochem* **19**, 37–42
32. Putnam, C. D., Hammel, M., Hura, G. L., and Tainer, J. A. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40**, 191–285
33. Receveur, V., Czjzek, M., Schüle, M., Panine, P., and Henrissat, B. (2002) Dimension, shape, and conformational flexibility of a two domain fungal cellulase in solution probed by small angle X-ray scattering. *J. Biol. Chem.* **277**, 40887–40892
34. Nygaard, M., Kragelund, B. B., Papaleo, E., and Lindorff-Larsen, K. (2017) An efficient method for estimating the hydrodynamic radius of disordered protein conformations. *Biophys. J.* **113**, 550–557
35. Pesce, F., Newcombe, E. A., Seiffert, P., Tranchant, E. E., Olsen, J. G., Grace, C. R., *et al.* (2023) Assessment of models for calculating the hydrodynamic radius of intrinsically disordered proteins. *Biophys. J.* **122**, 310–321
36. Pesce, F., and Lindorff-Larsen, K. (2021) Refining conformational ensembles of flexible proteins against small-angle X-ray scattering data. *Biophys. J.* **120**, 5124–5135
37. Thomasen, F. E., Pesce, F., Roesgaard, M. A., Tesei, G., and Lindorff-Larsen, K. (2022) Improving Martini 3 for disordered and multidomain proteins. *J. Chem. Theory Comput.* **18**, 2033–2041
38. Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Apweiler, R., Alpi, E., *et al.* (2015) UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212
39. Zheng, W., Dignon, G., Brown, M., Kim, Y. C., and Mittal, J. (2020) Hydropathy patterning complements charge patterning to describe conformational preferences of disordered proteins. *J. Phys. Chem. Lett.* **11**, 3408–3415
40. Tesei, G., Schulze, T. K., Crehuet, R., and Lindorff-Larsen, K. (2021) Accurate model of liquid-liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2111696118
41. Bissaro, B., Røhr, Å. K., Müller, G., Chylenski, P., Skaugen, M., Forsberg, Z., *et al.* (2017) Oxidative cleavage of polysaccharides by monocopper enzymes depends on H<sub>2</sub>O<sub>2</sub>. *Nat. Chem. Biol.* **13**, 1123–1128
42. Chang, H., Gacias Amengual, N., Botz, A., Schwaiger, L., Kracher, D., Scheiblbrandner, S., *et al.* (2022) Investigating lytic polysaccharide monoxygenase-assisted wood cell wall degradation with microsensors. *Nat. Commun.* **13**, 6258
43. Kont, R., Bissaro, B., Eijsink, V. G. H., and Våljamäe, P. (2020) Kinetic insights into the peroxygenase activity of cellulose-active lytic polysaccharide monoxygenases (LPMOs). *Nat. Commun.* **11**, 5786
44. Loose, J. S. M., Arntzen, M. Ø., Bissaro, B., Ludwig, R., Eijsink, V. G. H., and Vaaje-Kolstad, G. (2018) Multipoint precision binding of substrate protects lytic polysaccharide monoxygenases from self-destructive off-pathway processes. *Biochemistry* **57**, 4114–4124
45. Stepnov, A. A., Eijsink, V. G. H., and Forsberg, Z. (2022b) Enhanced *in situ* H<sub>2</sub>O<sub>2</sub> production explains synergy between an LPMO with a cellulose-binding domain and a single-domain LPMO. *Sci. Rep.* **12**, 6129
46. Filandr, F., Man, P., Halada, P., Chang, H., Ludwig, R., and Kracher, D. (2020) The H<sub>2</sub>O<sub>2</sub>-dependent activity of a fungal lytic polysaccharide monoxygenase investigated with a turbidimetric assay. *Biotechnol. Biofuels* **13**, 37
47. González-Foutel, N. S., Glavina, J., Borchers, W. M., Safranchik, M., Barrera-Vilarmou, S., Sagar, A., *et al.* (2022) Conformational buffering underlies functional selection in intrinsically disordered protein regions. *Nat. Struct. Mol. Biol.* **29**, 781–790
48. Hansson, H., Karkehabadi, S., Mikkelsen, N., Douglas, N. R., Kim, S., Lam, A., *et al.* (2017) High-resolution structure of a lytic polysaccharide monoxygenase from *Hypocrea jecorina* reveals a predicted linker as an integral part of the catalytic domain. *J. Biol. Chem.* **292**, 19099–19109
49. Zugenmaier, P. (2001) Conformation and packing of various crystalline cellulose fibers. *Prog. Polym. Sci.* **26**, 1341–1417
50. Harris, P. V., Xu, F., Kreel, N. E., Kang, C., and Fukuyama, S. (2014) New enzyme insights drive advances in commercial ethanol production. *Curr. Opin. Chem. Biol.* **19**, 162–170
51. Johansen, K. S. (2016) Lytic polysaccharide monoxygenases: the microbial power tool for lignocellulose degradation. *Trends Plant Sci.* **21**, 926–936
52. Koskela, S., Wang, S., Xu, D., Yang, X., Li, K., Berglund, L. A., *et al.* (2019) Lytic polysaccharide monoxygenase (LPMO) mediated production of ultra-fine cellulose nanofibres from delignified softwood fibres. *Green. Chem.* **21**, 5924–5933
53. Loose, J. S. M., Forsberg, Z., Fraaije, M. W., Eijsink, V. G. H., and Vaaje-Kolstad, G. (2014) A rapid quantitative activity assay shows that the *Vibrio cholerae* colonization factor GbpA is an active lytic polysaccharide monoxygenase. *FEBS Lett.* **588**, 3435–3440
54. Stepnov, A. A., Christensen, I. A., Forsberg, Z., Aachmann, F. L., Courtade, G., and Eijsink, V. G. H. (2022a) The impact of reductants on the catalytic efficiency of a lytic polysaccharide monoxygenase and the special role of dehydroascorbic acid. *FEBS Lett.* **596**, 53–70
55. Courtade, G., Le, S. B., Sætrom, G. I., Brautaset, T., and Aachmann, F. L. (2017b) A novel expression system for lytic polysaccharide monoxygenases. *Carbohydr. Res.* **448**, 212–219
56. Blanchet, C. E., Spilotros, A., Schwemmer, F., Graewert, M. A., Kikhney, A., Jeffries, C. M., *et al.* (2015) Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J. Appl. Crystallogr.* **48**, 431–443
57. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282
58. Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., *et al.* (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225
59. Mylonas, E., and Svergun, D. I. (2007) Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *J. Appl. Crystallogr.* **40**, 245–249
60. Orthaber, D., Bergmann, A., and Glatter, O. (2000) SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. *J. Appl. Crystallogr.* **33**, 218–225
61. Pedersen, M. C., Arleth, L., and Mortensen, K. (2013) WillItFit: a framework for fitting of constrained models to small-angle scattering data. *J. Appl. Crystallogr.* **46**, 1894–1898
62. Hansen, S. (2012) BayesApp: a web site for indirect transformation of small-angle scattering data. *J. Appl. Crystallogr.* **45**, 566–567
63. Bermel, W., Bertini, I., Duma, L., Felli, I. C., Emsley, L., Pierattelli, R., *et al.* (2005) Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* **44**, 3089–3092



64. Shen, Y., and Bax, A. (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241
65. Fushman, D. (2002) Determination of protein dynamics using 15N relaxation measurements. In *BioNMR in Drug Research*, Wiley-VCH Verlag, GmbH & Co KGaA, Weinheim, Germany: 283–308
66. Price, W. S., Elwinger, F., Vigouroux, C., and Stilbs, P. (2002) PGSE-WATERGATE, a new tool for NMR diffusion-based studies of ligand-macromolecule binding. *Magn. Reson. Chem.* **40**, 391–395
67. Edward, J. T. (1970) Molecular volumes and the Stokes-Einstein equation. *J. Chem. Educ.* **47**, 261–270
68. Wilkins, D. K., Grimshaw, S. B., Dobson, C. M., Jones, J. A., Smith, L. J., Receveur, V., *et al.* (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **38**, 16424–16431
69. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., *et al.* (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19–25
70. de Jong, D. H., Singh, G., Bennett, W. F. D., Arnarez, C., Wassenaar, T. A., Schäfer, L. V., *et al.* (2013) Improved parameters for the Martini coarse-grained protein force field. *J. Chem. Theory Comput.* **9**, 687–697
71. Periolo, X., Cavalli, M., Marrink, S.-J., and Ceruso, M. A. (2009) Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. *J. Chem. Theory Comput.* **5**, 2531–2543
72. Bussi, G., Donadio, D., and Parrinello, M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101
73. Parrinello, M., and Rahman, A. (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190
74. Páll, S., and Hess, B. (2013) A flexible algorithm for calculating pair interactions on SIMD architectures. *Comput. Phys. Commun.* **184**, 2641–2650
75. Wassenaar, T. A., Pluhackova, K., Böckmann, R. A., Marrink, S. J., and Tieleman, D. P. (2014) Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models. *J. Chem. Theory Comput.* **10**, 676–690
76. Grudinin, S., Garkavenko, M., and Kazennov, A. (2017) Pepsi-SAXS: an adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles. *Acta Crystallogr. D Struct. Biol.* **73**, 449–464
77. Svergun, D., Barberato, C., and Koch, M. H. J. (1995) Crysol - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773
78. Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., and Finn, R. D. (2018) HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200–W204
79. Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., *et al.* (2014) Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–D230
80. Anderson, J. A., Glaser, J., and Glotzer, S. C. (2020) HOOMD-blue: a Python package for high-performance molecular dynamics and hard particle Monte Carlo simulations. *Comput. Mater. Sci.* **173**, 109363
81. Shrestha, U. R., Smith, J. C., and Petridis, L. (2021) Full structural ensembles of intrinsically disordered proteins from unbiased molecular dynamics simulations. *Commun. Biol.* **4**, 243
82. Spezio, M., Wilson, D. B., and Karplus, P. A. (1993) Crystal structure of the catalytic domain of a thermophilic endocellulase. *Biochemistry* **32**, 9906–9916
83. Bissaro, B., Forsberg, Z., Ni, Y., Hollmann, F., Vaaje-Kolstad, G., and Eijsink, V. G. H. (2016) Fueling biomass-degrading oxidative enzymes by light-driven water oxidation. *Green. Chem.* **18**, 5357–5366
84. Irwin, D. C., Spezio, M., Walker, L. P., and Wilson, D. B. (1993) Activity studies of eight purified cellulases: specificity, synergism, and binding domain effects. *Biotechnol. Bioeng.* **42**, 1002–1013
85. Westereng, B., Agger, J. W., Horn, S. J., Vaaje-Kolstad, G., Aachmann, F. L., Stenström, Y. H., *et al.* (2013) Efficient separation of oxidized cello-oligosaccharides generated by cellulose degrading lytic polysaccharide monoxygenases. *J. Chromatogr. A* **1271**, 144–152
86. Huynh, K., and Partch, C. L. (2015) Analysis of protein stability and ligand interactions by thermal shift assay. *Curr. Protoc. Protein Sci.* **79**, 28.9.1–28.9.14