

# The structural basis of fungal glucuronoyl esterase activity on natural substrates

Heidi A. Ernst<sup>1</sup>, Caroline Mosbech<sup>2</sup>, Annette E. Langkilde<sup>3</sup>, Peter Westh<sup>2</sup>, Anne S. Meyer<sup>2</sup>, Jane W. Agger<sup>2</sup> & Sine Larsen<sup>1</sup>

Structural and functional studies were conducted of the glucuronoyl esterase (GE) from *Cerrena unicolor* (CuGE), an enzyme catalyzing cleavage of lignin-carbohydrate ester bonds. CuGE is an  $\alpha/\beta$ -hydrolase belonging to carbohydrate esterase family 15 (CE15). The enzyme is modular, comprised of a catalytic and a carbohydrate-binding domain. SAXS data show CuGE as an elongated rigid molecule where the two domains are connected by a rigid linker. Detailed structural information of the catalytic domain in its apo- and inactivated form and complexes with aldouronic acids reveal well-defined binding of the 4-O-methyl- $\alpha$ -D-glucuronoyl moiety, not influenced by the nature of the attached xylo-oligosaccharide. Structural and sequence comparisons within CE15 enzymes reveal two distinct structural subgroups. CuGE belongs to the group of fungal CE15-B enzymes with an open and flat substrate-binding site. The interactions between CuGE and its natural substrates are explained and rationalized by the structural results, microscale thermophoresis and isothermal calorimetry.

<sup>1</sup>Department of Chemistry, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen Ø, Denmark. <sup>2</sup>Department of Biotechnology and Biomedicine, Technical University of Denmark, Søtofts Plads 221, 2800 Kongens Lyngby, Denmark. <sup>3</sup>Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, 2100 Copenhagen Ø, Denmark. ✉email: [jaag@dtu.dk](mailto:jaag@dtu.dk); [sine@chem.ku.dk](mailto:sine@chem.ku.dk)

Lignin-carbohydrate complexes (LCCs) play an integral role in maintaining plant cell wall integrity and recalcitrance of lignocellulosic biomass<sup>1,2</sup>. LCCs represent a major obstacle for the complete utilization of all plant cell wall components, which is a cornerstone in the future bioeconomy-based society because lignin is an important natural resource for replacing fossil-based products and materials with sustainable alternatives<sup>3</sup>. To fully exploit the potential of lignin, it is essential to recover lignin in a non-destructive manner and in high purity, free from residual carbohydrates. A common type of LCC in hardwood is the ester-LCC formed between  $\alpha$ -1,2-linked 4-*O*-methyl-*D*-glucuronoyl moieties of acetylated glucuronoxylan and  $\gamma$ -alcohols in lignin<sup>4</sup> (Fig. 1a). It was recently discovered that an insoluble hardwood lignin-preparation (LRP) is amenable to enzymatic hydrolysis by a glucuronoyl esterase (GE, EC 3.1.1.B11) releasing aldouronic acids<sup>5</sup>. This opens for the opportunity to use glucuronoyl esterases as a benign tool for processing of lignin in selective removal of residual carbohydrates in contrast to the harsh physio/chemical extractions otherwise used.

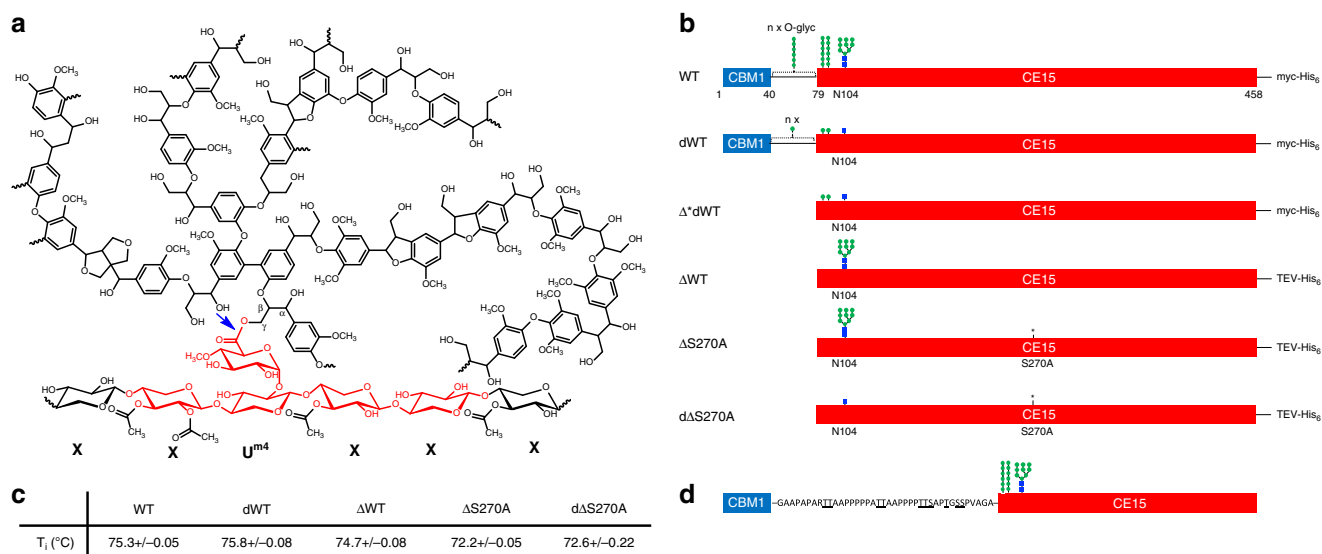
GEs are widespread in nature and found in fungal and bacterial lignocellulose degrading organisms<sup>6,7</sup>. The currently known GEs belong to the CE15 family<sup>8</sup> in the CAZy database. The first crystal structure of a CE15 GE, Cip2 from *Hypocrea jecorina*<sup>9</sup> and site-directed mutagenesis studies<sup>10</sup> established that the CE15 enzymes belong to the  $\alpha$ / $\beta$ -hydrolase superfamily with a classical catalytic nucleophile-acid-base (serine-glutamate-histidine) triad and their catalytic function conforms to other enzymes possessing an equivalent catalytic machinery<sup>11</sup>.

The CE15 family is diverse and structural studies of fungal<sup>9,12</sup> and bacterial<sup>13–15</sup> members have revealed significant differences within the family, including variations in catalytic machinery and active site topology. There is only one report of a CE15 complex,

the StGE2:monosaccharide complex<sup>12</sup>, unfortunately this study is not supported by any functional data.

Studies on substrate analogs suggest that the 4-*O*-methylation on the glucuronoyl substitutions in xylan is important for substrate recognition<sup>16,17</sup>. While the ester linkage of the LCC itself may seem well-defined in terms of acid and alcohol donors (Fig. 1a), the structural variations of its surroundings, such as the degree and position of acetylations on the xylan backbone, the amount of glucuronoyl substitutions, the distance between two or more ester linkages, the composition of lignin and the average molecular size of both the carbohydrate and the lignin moieties add complexity and heterogeneity to LCCs as substrate. We have shown previously, that a two-domain glucuronoyl esterase from the basidiomycete *Cerrena unicolor* (*CuGE*) with high efficiency can catalyze the release of large soluble aldouronic acids from a lignin-rich birchwood fraction, most likely containing the already mentioned ester-linked LCCs<sup>5</sup>.

In the following, our results from extensive structural and functional studies of *CuGE* are presented. This includes structural investigations by small-angle X-ray scattering (SAXS) of the full-length enzyme, crystal structures of the apo form of the catalytic domain, as well as four relevant complexes with aldouronic acids, and binding affinity measurements on natural insoluble lignin substrate, containing LCCs and soluble substrate analogs. Collectively, these results give us important structural and functional knowledge clarifying the interactions between *CuGE* and the  $\alpha$ -1,2-linked 4-*O*-methyl-*D*-glucuronoyl moieties on xylo-oligomers and explaining how the enzyme accommodates large complex substrates. Furthermore, a comparative study with other members of the CE15 family reveals two distinct groups of enzymes, A and B of which *CuGE* belongs to group B. The two groups differ in the position of their catalytic acid and topology, which indicate



**Fig. 1** Proposed natural LCC substrate and *CuGE* variants. **a** Structural representation of a natural LCC substrate for glucuronoyl esterases. The ester-linked LCC (blue arrow) connects two polymeric structures; acetylated glucuronoxylan substituted with  $\alpha$ -1,2-linked 4-*O*-methyl-glucuronoyl moieties and lignin. The 4-*O*-methyl-glucuronoyls form ester bonds with  $\gamma$ -alcohols from lignin. The carbohydrate structure corresponding to the three different ligands in the *CuGE* crystallographic complexes are marked in red. Each xylosyl unit is named X and the 4-*O*-methyl-glucuronoyl substituted xylosyl is named  $U^{m4}$ . **b** *CuGE* variants used in this study. The full-length enzyme (WT) has a modular structure with a CBM1 domain appended to the catalytic CE15 domain via a linker. Positions that are post translationally modified in the recombinant enzyme with either N- or O-glycosylation are stylized (NAG blue squares; MAN green dots). Here, only the core structure of the glycans is shown, but especially the N-glycosylation is quite extensive and heterogeneous in the recombinant enzyme produced in *Pichia pastoris*. “ $\Delta$ ” symbolizes a truncated variant without the CBM1 and linker region. In the inactive  $\Delta$ S270A variant the catalytic nucleophile S270 has been changed to an alanine. “d” denotes variants that have been enzymatically deglycosylated. Notably, the cleaved variant  $\Delta^*dWT$  was spontaneously produced in crystallization experiments with dWT. **c** Relative stability of the *CuGE* variants.  $T_m$  is the apparent thermal unfolding temperature derived from changes in intrinsic fluorescence and stated with standard deviations. **d** Sequence of the proline-rich linker region in *CuGE*. Potential O-glycosylation sites (S or T) are underlined. Source data for panel **c** are provided as a Source Data file.

differences in substrate binding. This overall knowledge on the CE15 enzymes will pave the ground for the understanding of how lignin can be selectively separated from the lignocellulosic cell wall via enzymatic cleavage of ester-linked LCCs.

## Results

**CuGE variants.** WT *CuGE* has a modular architecture with a N-terminal carbohydrate-binding module 1 (CBM1) domain<sup>17,18</sup> connected to the catalytic domain by a proline-rich linker region. The structure of *CuGE* is potentially flexible and together with heterogenous N- and O-glycosylation<sup>17</sup> it could represent an obstacle for high-resolution structural studies of the full-length enzyme (WT). To enable structural and functional studies two different variants were constructed,  $\Delta$ WT, comprises residues 78–458 lacking the carbohydrate-binding domain (CBM1), and  $\Delta$ S270A an inactive variant of  $\Delta$ WT where the catalytic serine is replaced with an alanine (Fig. 1b–d). These *CuGE* variants were recombinantly produced in *Pichia pastoris* and subsequently purified. For the structural studies, WT and  $\Delta$ S270A were enzymatically deglycosylated to obtain homogeneous samples, named dWT and d $\Delta$ S270A, respectively (Fig. 1b–c). Comparative thermal unfolding measurements on the preparations show that neither the truncation nor the deglycosylation affects protein stability significantly (Fig. 1c). The inactive variant  $\Delta$ S270A is destabilized by ca 3 °C compared to variants with an intact catalytic triad, implying that some stabilizing interactions have been lost, but the overall conclusion is that *CuGE* is a highly stable enzyme with  $T_i$ 's in the range of 72.2–75.8 °C for all variants. Previous studies have shown that both WT and  $\Delta$ WT *CuGE* are fully active on both soluble and insoluble substrates<sup>18</sup>.

**Overall structure of *CuGE* by small-angle X-ray scattering.** In-house small-angle X-ray scattering measurements were performed on dWT and d $\Delta$ S270A (Fig. 2a). No concentration dependence was observed in the investigated concentration range (Supplementary Fig. 1), and the estimated molecular weights correspond well to monomeric species for both variants. Furthermore, the dimensionless Kratky plots shown in Fig. 2b for both samples, with their distinct bell-shaped curves reaching a plateau near zero at higher  $q$ -values, display the characteristics of rigid folded proteins. With a peak position in the dimensionless Kratky plot at  $(sR_g)^2 = 1.1$ , the data for the d $\Delta$ S270A variant indicate a compact and globular structure. The maximum interatomic distance ( $D_{\max}$ ) is 60 Å, derived from the pair distribution function  $P(r)$  (Fig. 2c).

The SAXS data of the full-length protein dWT, with an increased and right-shifted peak position in the dimensionless Kratky plot, display the characteristics of a multi domain elongated yet rigid particle. The pair distance distribution (Fig. 2c) includes both the profile of the catalytic domain and a second peak with relatively constant interdomain distance consistent with the presence of the CBM1 and a  $D_{\max}$  of 110 Å. Ab-initio models of dWT (Fig. 2d, e) reveal a relatively extended structure formed of two domains connected by the 40-residue long linker. The structure of full-length *CuGE* thus conforms to the emerging picture that linker regions in many CAZymes can be quite extended and rigid<sup>19–22</sup>. These properties appear to be associated with the high content of proline, threonine and serine in the linker region (Fig. 1d and Supplementary Material Fig. 2), which provides stiffness and potentially heavy O-glycosylation that may further reduce the flexibility of the linker<sup>19</sup>. Unfortunately it was not possible to unambiguously determine the conformation of the linker from the SAXS data.

The high-resolution structure of the catalytic domain (d $\Delta$ S270A) obtained from the crystal structure determinations

(described in detail later) matched the solution structure (Fig. 2a). This crystal structure and a homology model of *CuGE*-CBM1 (Supplementary Fig. 2) were placed in the refined DAMMIN envelope for reference (Fig. 2e). Owing to the globular nature of the individual domains the relative orientation of the two domains remains ambiguous.

**Structure of the catalytic domain of *CuGE*.** Crystals obtained from dWT turned out to contain the apo form of the catalytic domain, resulting from cleavage of the wild-type enzyme ( $\Delta^*$ dWT, Fig. 1b). This shows that the deglycosylated, unprotected linker region is susceptible to proteolytic cleavage. The structure of  $\Delta^*$ dWT was determined to 1.96 Å resolution. The structure of the inactive truncated variant, apo-d $\Delta$ S270A was determined to 1.46 Å resolution. Through soaking experiments with ligands mimicking parts of the natural substrate (Fig. 1a) it was possible to obtain high-resolution diffraction data for  $\Delta^*$ dWT and d $\Delta$ S270A with ligands bound (Supplementary Tables 1 and 2, Fig. 3a–c, and Fig. 4). All crystals of the catalytic domain of *CuGE* belong to the tetragonal space group  $P4_12_12$  with two molecules in the asymmetric unit (Supplementary Fig. 3). The structure of apo-d $\Delta$ S270A was determined by molecular replacement using Cip2 (3PIC)<sup>9</sup> as search model.

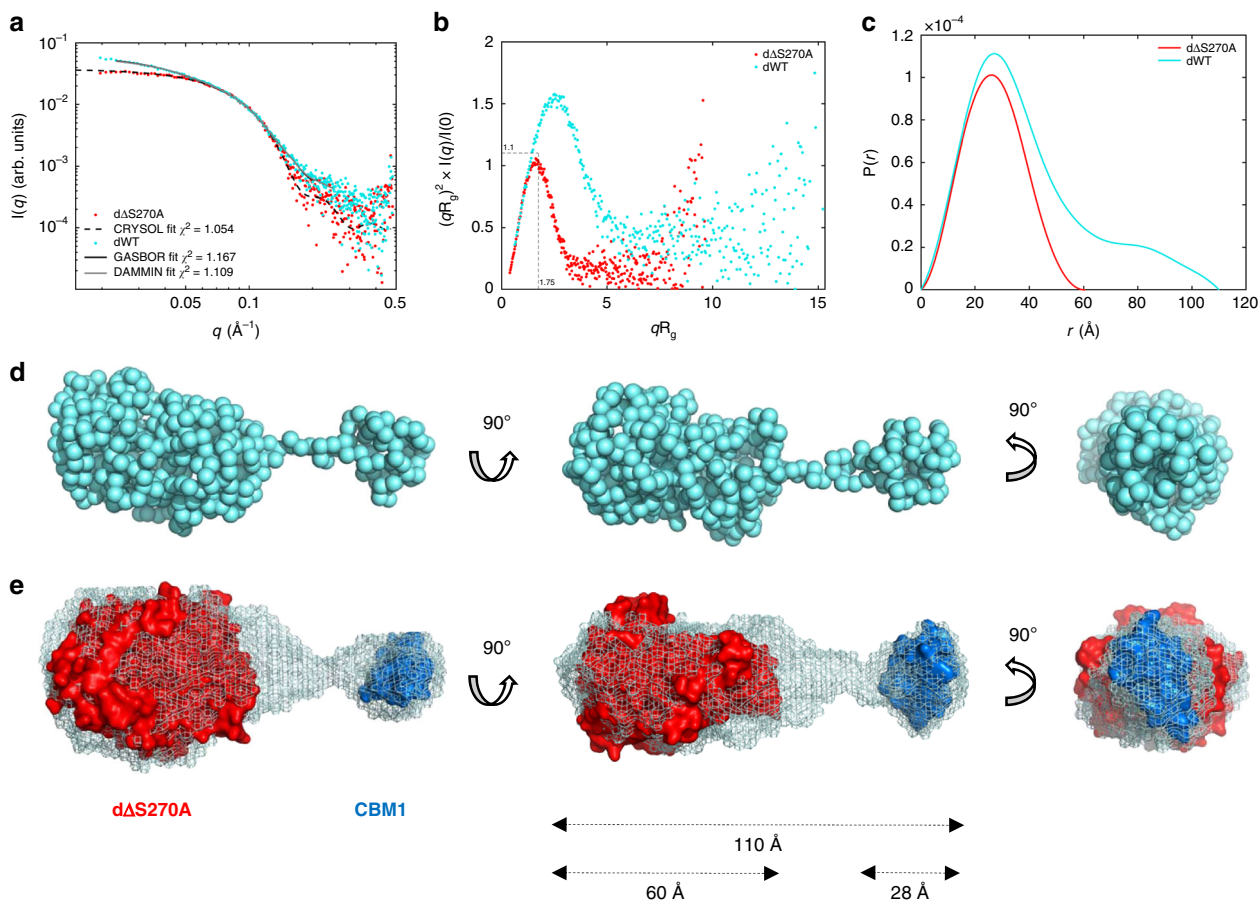
The overall fold of the catalytic domain of *CuGE* features a large ten-stranded twisted  $\beta$ -sheet flanked by  $\alpha$ -helical elements and a surface-exposed active site (Fig. 3a). The structure complies with the characteristics of a serine esterase belonging to the  $\alpha/\beta$ -hydrolase superfamily<sup>11,23,24</sup>. The electron density reveals a very well-defined domain comprising residues 81–458 (Figs. 1b and 3a–c). In all *CuGE* structures, a single N-acetyl glucosamine moiety is attached to N104 as a relic of the N-glycosylation. In addition, signs of O-glycosylation ( $\alpha$ -1-linked mannose) are observed in some of the structures (Fig. 3a).

The two crystallographically independent *CuGE* molecules differ in their crystal packing environments and accessibility to the active site (Supplementary Fig. 3). Nevertheless, molecule A and B superimpose well with an RMSD of 0.32 Å for 359 aligned  $C_\alpha$  atoms (apo-d $\Delta$ S270A). Structures of ligand-bound complexes showed that the ligands bind in the highly accessible and solvent exposed binding site of molecule B only. Overall, molecule B is less tightly packed and the higher  $B$ -values suggest that some surface-exposed loops are more flexible in this molecule (Fig. 3b, c). It is noteworthy that the binding of ligands hardly induce any structural changes of the residues in the active sites.

## The canonical $\alpha/\beta$ -hydrolase fold of two major CE15 groups.

*CuGE* is the first structurally characterized representative of a CE15 enzyme from basidiomycetes and displays an overall architecture similar to Cip2<sup>9</sup> and StGE2<sup>12</sup> from ascomycetes. The structural differences between *CuGE* and the two ascomycete's enzymes relate to minor variations in length of loops and differences in the structure of the N-terminal residues, regions distantly located from the active site.

On the contrary, a comparison of CE15 structures in general shows distinct differences between fungal and bacterial members of the CE15 family (Supplementary Fig. 4). While all CE15 structures belong to the  $\alpha/\beta$ -hydrolase superfamily<sup>11,23,24</sup> scrutiny of their sequences and structures, coupled to the generally conserved structural zones of the active site of  $\alpha/\beta$ -hydrolases<sup>25,26</sup>, shows that CE15 enzymes can be divided into two overall subgroups, CE15-A and CE15-B, corresponding to the catalytic acid configuration in the  $\alpha/\beta$ -hydrolase group A and B, respectively (Fig. 3d–f). In CE15-A, the acidic member of the catalytic triad is located after  $\beta$ 8 and is typically an aspartic acid (Fig. 3e). CE15-A comprises mostly bacterial enzymes, but also the



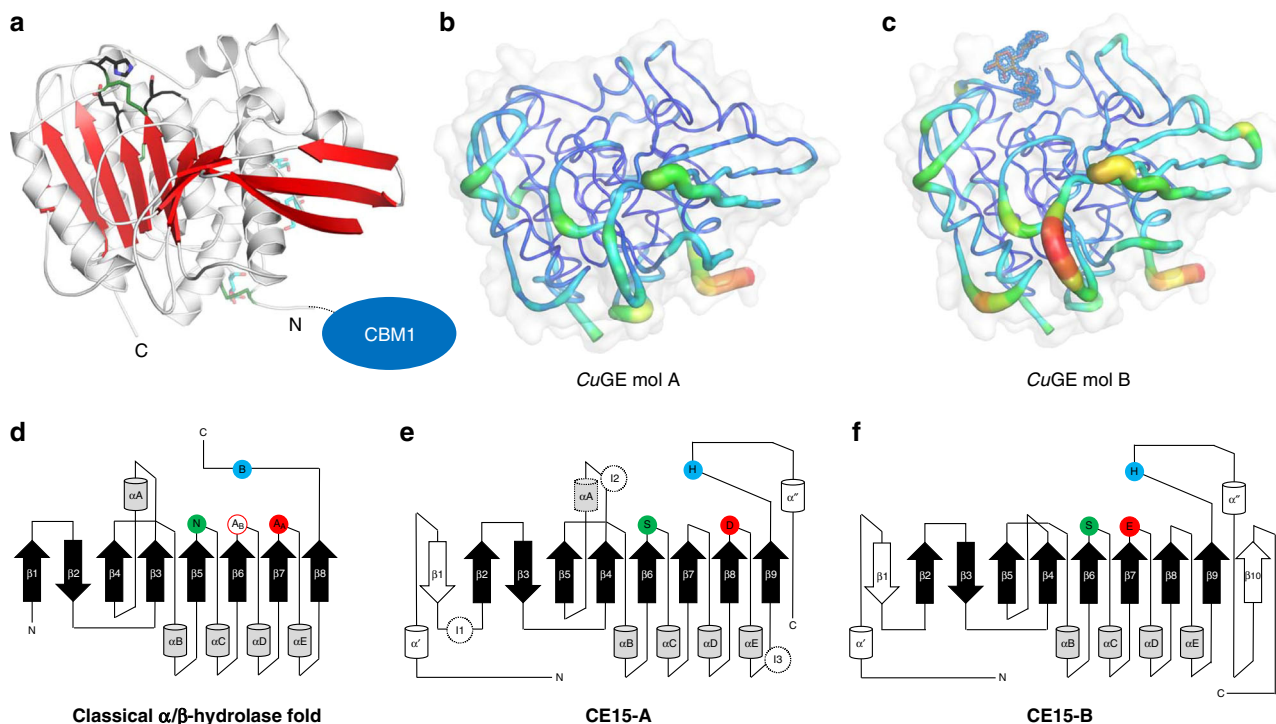
**Fig. 2** SAXS analyses and models of full-length CuGE. **a** SAXS scattering profiles for truncated, dΔS270A (3.6 mg mL<sup>-1</sup>), and full-length, dWT (3.7 mg mL<sup>-1</sup>), CuGE shown on double logarithmic scales. The CRYSOLO fit shows an excellent agreement between the solution structure of dΔS270A and the crystal structure. For dWT, the fits obtained by ab-initio methods, here GASBOR and DAMMIN, respectively, are depicted (black and gray lines, almost fully overlapping). **b** Dimensionless Kratky plots indicating folded and rigid structures. **c** Pair distance distribution functions,  $P(r)$ , with a  $D_{\max}$  of 60 Å for dΔS270A and 110 Å for dWT, respectively. **d** Ab-initio GASBOR model of dWT. **e** Ab-initio DAMMIN model of dWT with the crystal structure of the catalytic domain and a CuGE-CBM1 homology model (Supplementary Fig. 2) placed in the molecular envelope.

fungal subgroup previously referred to as PPR8<sup>6</sup> or G1<sup>27</sup>. In CE15-B, the catalytic acid is located after β7 and is always a glutamic acid. CE15-B comprises mostly fungal enzymes, including CuGE, CiP2, and StGE2, but also enzymes of bacterial origin. Variations in the identity and position of the catalytic acid among members of CE15 have been discussed previously<sup>6,7,13–15,18,27,28</sup>. However, with our direct reference to the α/β-hydrolase superfamily we have been able to obtain a clear division of CE15 enzymes into two subgroups with fundamentally different properties encoded in their structural arrangements.

Besides the division in subgroups A and B based on the catalytic acid location and type, CE15 enzymes do not retain the canonical oxyanion zone observed in most α/β-hydrolases. The usual oxyanion hole conformation stabilizes the tetrahedral reaction intermediate by two main chain nitrogen atoms and involves a conserved H-G-x-X<sub>oxyII</sub> motif at the end of β3 (β4 in CE15)<sup>26</sup>. CE15 structures show variations in this region; bacterial CE15-A members typically have an extra insert before αA, referred to as “insert 2”<sup>13–15</sup> (Fig. 3e), whereas the fungal members of both CE15-A and CE15-B have a deletion and completely lack the canonical helix αA and the insert 2 (Fig. 3d–f and Supplementary Figs. 4–6). These differences have dramatic consequences for the overall shape of the substrate-binding site. A fundamental and unifying feature of both CE15 subgroups is the highly conserved CE15-signature motif GxSRxGK located in the

nucleophilic elbow between β6 and αC (Supplementary Fig. 4) harboring a conserved arginine that seemingly play a role in the creation of the oxyanion hole<sup>12,14</sup>.

**Active site architecture and ligand-induced oxyanion hole.** Consistent with the proven role in interfacial biocatalysis of derivatives of large LCC substrates<sup>5,18</sup>, the active site of CuGE is readily accessible at the surface of the protein. The crystal structures (resolution 1.39–1.73 Å) of CuGE complexes with aldouronic acids representing the carbohydrate portion of the natural substrate (XU<sup>m4</sup>X, Fig. 1a) significantly advance our knowledge on substrate recognition (Fig. 4a–i and Supplementary Tables 1 and 2). The active site appears remarkably robust with little variation between structures determined at different conditions (pH, active vs. inactive, apo vs. ligand-bound). Robustness of the active site region is a characteristic feature of α/β hydrolases<sup>25,26</sup>. In the fungal CE15 enzymes, this is reinforced by a conserved disulfide linkage that anchors two of the catalytic residues (H404 and S270 in CuGE; Fig. 4b, c). Another characteristic feature is the very short hydrogen bond (2.5 Å) between E320 and E324 (Fig. 4e) that contributes to the creation of a relatively rigid active site, in which conformational variations are only observed for two residues: Y366 and R271. In apo-dΔS270A Y366 adopts a unique conformation in molecule B (pH 7.5–8; Fig. 4c) that seems to be incompatible with substrate binding, but

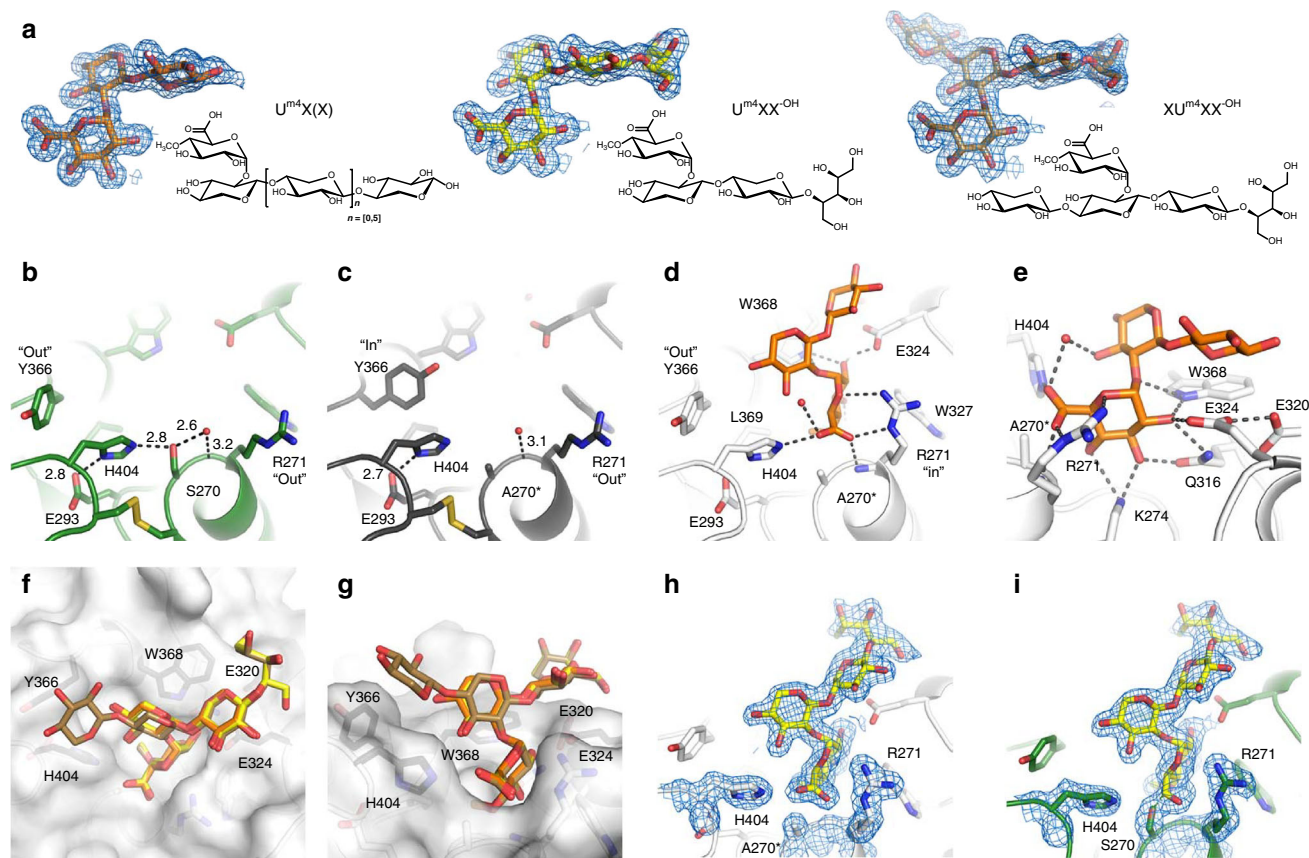


**Fig. 3 Overall CuGE structure, CE15 subgroups, and relation to the classical  $\alpha/\beta$ -hydrolase fold.** **a** Structure of the catalytic domain of CuGE (residues 79–458) with selected residues shown as sticks: the catalytic triad (S270, E293, and H404, black), disulfide bridges (C81–C116, C269–C405, C391–C377, green), and carbohydrates (cyan), indicating the location of N- and O-glycosylation sites (N104-NAG, S86-MAN, and T87-MAN). The break in strand  $\beta 3$  is caused by a proline residue (P169) in CuGE. **b, c** The two crystallographically independent molecules A and B from the  $d\Delta S270A:U^{m4}X$  complex colored by temperature factors to visualize flexibility. **b** Molecule A. **c** Ligand-bound molecule B. **d** Topology diagram for a classical  $\alpha/\beta$ -hydrolase fold. The illustration is based on information in Ollis et al.<sup>23</sup>, Carr et al.<sup>24</sup>, and Dimitrou et al.<sup>25</sup>. The location of the catalytic residues is shown (nucleophile serine in green, base histidine in blue, and acid glutamate in red). The  $\alpha/\beta$ -hydrolase group A and B are distinguished by different configurations of the catalytic triad with the catalytic acid located either in the turn after  $\beta 7$  ( $A_A$ ) or after  $\beta 6$  ( $A_B$ ), respectively<sup>25</sup>. **e** Topology diagram for CE15-A. I1–I3 indicate inserts observed in bacterial CE15-A structures<sup>13–15</sup>. **f** Topology diagram for CE15-B. For clarity, only the major secondary structure elements have been included in the topology diagrams in **e, f**; structures from both CE15 subgroups comprise additional small secondary structure elements, including multiple helical elements inserted after the strands  $\beta 5$  and  $\beta 7$  (Supplementary Fig. 4).

identifies Y366 as a structurally flexible element. We observe that binding of aldouronic acids is coupled to a conformational change of R271 (Fig. 4c, d) that allows the side chain to engage in substrate recognition and complete the formation of the oxyanion hole. Thus, the oxyanion hole is created solely by R271 with both the backbone amide and guanidinium side chain nitrogen atoms forming close contacts with the O6 carbonyl of the 4-O-methyl-glucuronoyl moiety (3.2 and 2.8 Å, respectively), mimicking the interactions in a tetrahedral reaction intermediate (Fig. 4d–e). The unliganded structures of bacterial CE15 enzymes all display a pre-organized oxyanion hole with the arginine in an orientation very similar to the conformation observed in the ligand-bound structures of CuGE. However, a corresponding conformational change of the conserved arginine was not observed in the StGE2 complex with the substrate analog methyl-4-O-methyl-glucuronic acid<sup>12</sup>.

**Recognition of the  $XU^{m4}X$ -fragment.** The aldouronic acid ligands dock onto the surface of CuGE in all the complexes:  $d\Delta S270A:U^{m4}X$ ,  $d\Delta S270A:U^{m4}XX-OH$  and  $d\Delta S270A:XU^{m4}XX-OH$  without any major conformational changes of the slightly twisted, branched oligosaccharide. The  $\alpha$ -1,2-linked 4-O-methyl-D-glucuronoyl moiety is nested in the small substrate-binding pocket, that provides distinct recognition of all hydroxyl groups and a hydrophobic environment for accommodation of the 4-O-methyl-D-glucuronoyl substituent, while the xylan-backbone runs

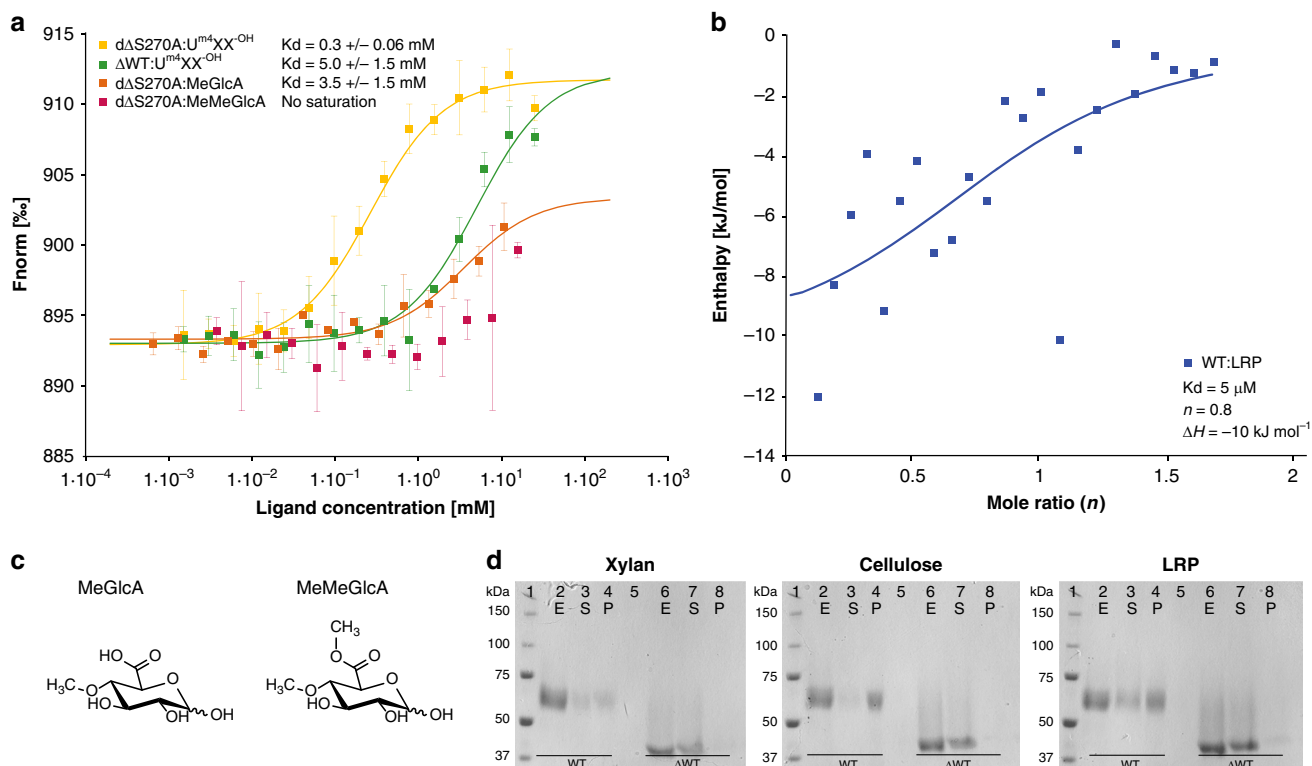
along the surface and displays only hydrophobic, van der Waals and water-mediated interactions with the protein (Fig. 4d–g). The notion that no direct hydrogen bonds are involved in the recognition of the xylan-backbone is consistent with the variable structure of a natural LCC substrate with potential decorations at the C2- and C3-hydroxyl positions (Fig. 1a). Spatially, there is room for such decorations, except at the O3-position of the 4-O-methyl-glucuronoyl substituted xylosyl, where decoration could be unfavorable due to steric effects (clash with H404). In the complex formed at pH 7.5–8 ( $d\Delta S270A:U^{m4}X$ ), Y366 moves “out” upon substrate binding (Fig. 4c, d), and no significant differences are observed in substrate recognition interactions compared to the complexes formed at pH 5.5 ( $d\Delta S270A:U^{m4}XX-OH$  and  $d\Delta S270A:XU^{m4}XX-OH$ ) (Fig. 4f, g). All hydroxyl groups in the 4-O-methyl-D-glucuronoyl moiety are engaged in hydrogen bonds and the free carboxylate group displays polar interactions with both the catalytic H404 and R271 (Fig. 4d, e). R271 furthermore interacts with the O5 atom in the carbohydrate ring (Fig. 4d, e). Despite these extensive interactions with the 4-O-methyl-D-glucuronoyl moiety, soaks with the methyl ester of 4-O-methyl-glucuronic acid in concentrations up to 0.5 M did not result in a complex, indicating a low binding affinity for the ester and a clear preference for the larger aldouronic acid ligands. Indeed, the structure of  $U^{m4}XX-OH$  in complex with the active variant  $\Delta^*dWT$  shows continuous electron density, suggesting that a covalent acyl-enzyme intermediate has formed (Fig. 4i).



**Fig. 4** Active site, crystallographic complexes, and  $XU^{m4}X$ -recognition. **a** Aldouronic acid ligands in the complexes  $d\Delta S270A:U^{m4}X$ ,  $d\Delta S270A:U^{m4}XX-OH$ , and  $d\Delta S270A:XU^{m4}XX-OH$ .  $2F_o-F_c$  electron density maps for the refined ligands are shown at the  $1.0 \sigma$  level. **b** Apo- $\Delta^*$ dWT with an intact active site (pH 5.5). Hydrogen bonds between residues in the catalytic triad are indicated. **c** Apo- $d\Delta S270A$  (pH ca. 7.5) with Y366 in a different conformation (molecule B only). Loss of interactions to S270 might explain the slight destabilization observed for the inactive variant (Fig. 1d). **d** Ligand-bound structure,  $d\Delta S270A:U^{m4}X(X)$  (pH ca. 7.5), shown in the same orientation as in **b**, **c** for a direct comparison with the unliganded structures. Note that two movements are observed upon substrate binding: Y366 moves “out”, whereas R271 moves “in” and participates in substrate recognition and formation of the oxanion hole. **e**  $d\Delta S270A:U^{m4}X(X)$  complex shown in a different perspective to visualize the extensive network of hydrogen bonds in recognition of the 4-O-methyl- $\alpha$ -D-glucuronoyl moiety. **f**, **g** Two different views on aldouronic acid ligand recognition mode focusing on interactions with the xylo-oligosaccharide backbone. The superposition of the  $d\Delta S270A$  complexes with  $U^{m4}X$ ,  $U^{m4}XX-OH$ , and  $XU^{m4}XX-OH$ , based on the protein  $C_\alpha$  atoms, illustrates little variation in the position of the ligand. Residues within 5 Å from the xylosyl moieties are labeled and shown in black. **h**  $d\Delta S270A:U^{m4}XX-OH$ , aldotetrauronic acid in complex with the inactive variant. **i**  $\Delta^*$ dWT: $U^{m4}XX-OH$ , aldotetrauronic acid in complex with catalytically active CuGE. Continuous electron density between the catalytic nucleophile S270 and the ligand suggests the formation of a covalent reaction intermediate.

**CuGE affinity for natural substrate and soluble fragments.** Binding affinity measured by microscale thermophoresis (MST) and isothermal titration calorimetry (ITC; Fig. 5a, b) shows differences in binding between soluble and insoluble ligands with a high affinity towards the insoluble substrate compared to soluble ligands and substrate analogs. MST measurements of  $\Delta WT$  and  $d\Delta S270A$  respectively on aldotetrauronic acid ( $U^{m4}XX-OH$ ) indicate equilibrium dissociation constants in the mM range with a moderately stronger binding (one order of magnitude) between  $d\Delta S270A$  and  $U^{m4}XX-OH$  compared to the active variant. The strong binding in  $d\Delta S270A:U^{m4}XX-OH$  is in accordance with the well-defined binding observed in the complexes of CuGE. The binding observed between  $\Delta WT$  and  $U^{m4}XX-OH$  ( $K_d$  5 mM) can be interpreted as a measure of the binding affinity of the enzyme to the carbohydrate moiety of the reaction product. Binding between  $d\Delta S270A$  and the 4-O-methyl-glucuronoyl moiety alone (structure shown in Fig. 5c) has  $\sim 10$  times higher dissociation constant compared to  $d\Delta S270A:U^{m4}XX-OH$  and no saturation was obtained with the ester-mimicking model substrate methyl-4-O-methyl-D-glucuronic acid (structure shown in Fig. 5c), indicating a very low affinity ( $K_d > 10$  mM). Besides the obvious lack

of xylo-oligosaccharide backbone in the latter two ligands, neither of these two possesses a fixed  $\alpha$ -anomeric configuration on C1, which may presumably bind stronger compared to the  $\beta$ -anomer. In nature, only the  $\alpha$ -anomer have been found in glucuronoxylan and hence the binding may be optimized for this configuration. ITC measurements and model fitting of data points for binding between WT CuGE and the insoluble lignin-rich precipitate from birchwood (LRP) rationalizes binding close to a 1:1 ratio with a relatively low enthalpy ( $\sim 10$  kJ mol $^{-1}$ ) and a dissociation constant in the low  $\mu$ M range (Fig. 5b). The ITC titration data are corrected for background measurements from experiments without enzyme, and these controls reveal relatively high background signals, possibly related to use of an insoluble, particulate substrate. These background signals result in a low signal-to-noise ratio in the actual binding experiment, but the estimated  $K_d$  value (5  $\mu$ M) imply a strong interaction between the enzyme and this insoluble substrate in comparison to those observed for soluble ligands. Previous kinetic studies with WT CuGE on the same substrate<sup>18</sup> revealed a  $K_M$  of  $\sim 220$   $\mu$ M and if  $K_M$  is taken as a measure of substrate affinity, the higher value compared to  $K_d$  measured by ITC suggests non-productive binding for instance



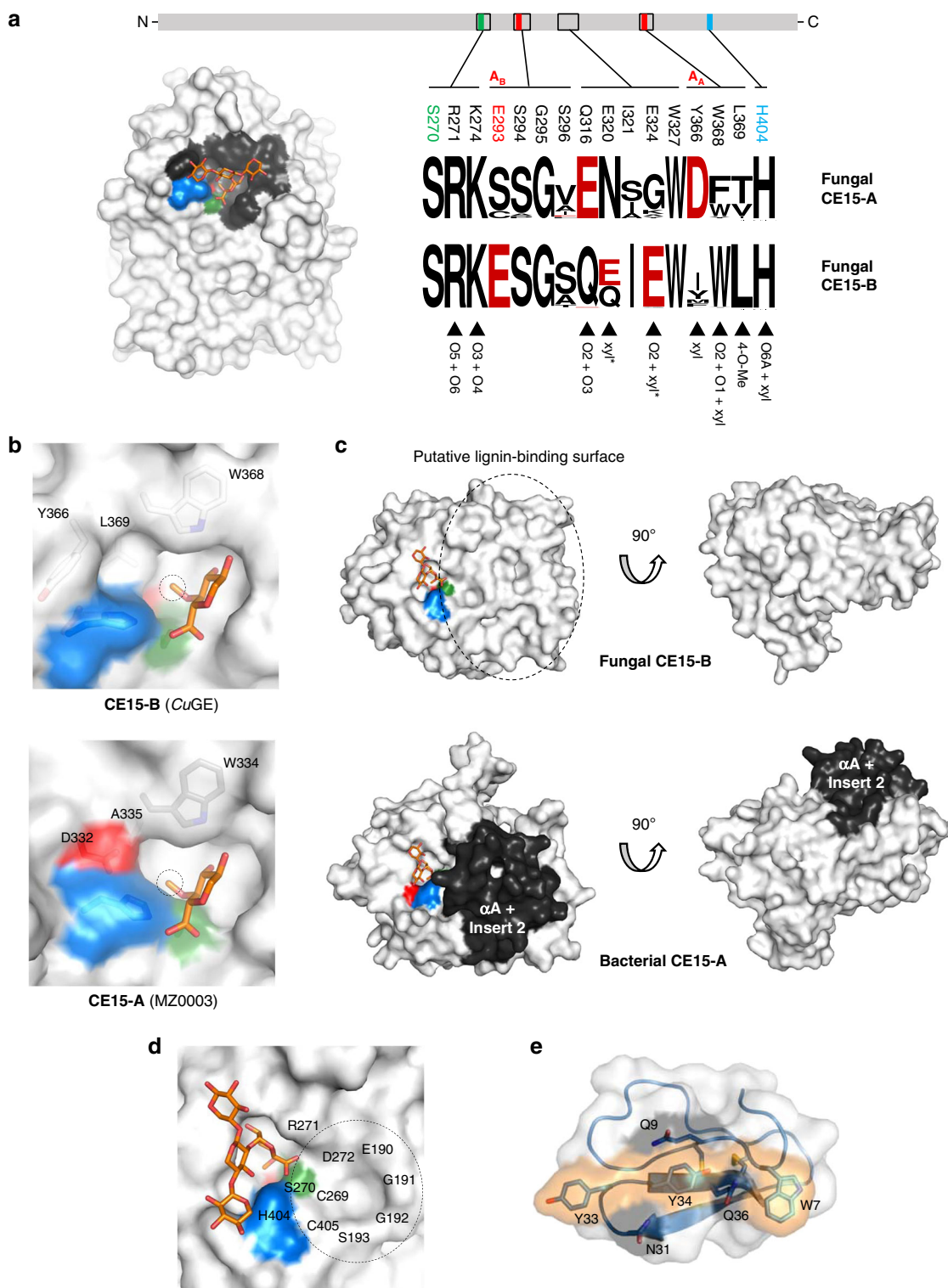
**Fig. 5** Affinity for natural substrates or derived soluble fragments. **a** Ligand-binding affinities of soluble ligands to various *CuGE* variants as determined by LabelFree microscale thermophoresis (MST). Legend details and estimated  $K_d$ -values are listed in the panel and error bars represents standard deviation. **b** ITC enthalpogram of the interaction between WT *CuGE* and the enzyme's natural insoluble lignin-rich substrate (LRP) at 5°C. The titration curve is depicted as a function of the molar ratio ( $n$ ) between enzyme and the calculated concentration of 4-*O*-methyl-glucuronoyl in the assay. The latter is an estimation of the concentration of recognition sites in the substrate. **c** Chemical structures of small ligands 4-*O*-methyl- $\beta$ -D-glucuronic acid (MeGlcA) and the methyl ester of 4-*O*-methyl- $\beta$ -D-glucuronic acid (MeMeGlcA) used in MST measurements of binding affinity in **a**. The structures are depicted to illustrate that equilibrium between the  $\alpha$ - and  $\beta$ -anomers exists in an aqueous solution. **d** Qualitative SDS-PAGE-based “pull-down” assay illustrating the effect of CBM1 on the binding to the insoluble biopolymers xylan, cellulose and LRP, respectively. 1: Molecular size ladder, 2: WT without substrate, 3: Supernatant after binding between WT *CuGE* and insoluble substrate, 4: Pellet after binding between WT *CuGE* and insoluble substrate, 5: Insoluble substrate without enzyme, 6:  $\Delta$ WT *CuGE* without substrate, 7: Supernatant after binding between  $\Delta$ WT *CuGE* and insoluble substrate, 8: Pellet after binding between  $\Delta$ WT *CuGE* and insoluble substrate. The description of lane content applies to all three gel pictures. Source data for panel **a**, **b**, and **d** are provided as a Source Data file.

by adsorption mediated by the CBM1 domain. Binding between *CuGE* WT or  $\Delta$ WT and the soluble ligands is too weak to be measured by ITC. Pull-down assays where WT and  $\Delta$ WT *CuGE* were incubated with insoluble xylan, cellulose or LRP (Fig. 5d) show that the full-length enzyme binds well to all three insoluble biomasses, whereas the truncated enzyme without its CBM1 binds less. In addition,  $K_M$  for  $\Delta$ WT *CuGE* on LRP is approx. twice as high compared to the full-length enzyme and  $k_{cat}$  is reduced by >30%<sup>18</sup>. Non-productive binding to lignin may occur in the presence of the CBM1 and could slow down the reaction speed. However, substrate recognition, binding and possibly also positioning of the enzyme to close-proximity cellulose or xylan is likely to be an important feature of *CuGE*. Its location on the substrate may be an important factor for the main rate limiting step in natural, enzymatic conversion of LCCs.

**Conservation of carbohydrate recognition among fungal CE15s.** The molecular details of XU<sup>m4</sup>X recognition and overall substrate-binding properties are likely to be generic for a large group of highly conserved fungal enzymes (Supplementary Fig. 5) where the sequence identity between CE15-B members are in the range of 41–77%. The fungal members of CE15-A display low overall sequence identity compared to the members of CE15-B, furthermore they differ in the configuration of the catalytic triad

(Supplementary Figs. 5 and 6). The overall sequence identity within the fungal CE15-A group is 40–75%, whereas the sequence identity between the two groups is merely 25–36%.

The variations between the substrate-binding sites in the two subgroups could reflect functional differences. To investigate this, a subset of residues that defines the XU<sup>m4</sup>X-binding site was selected and compared to the corresponding sequence logos produced from a large dataset of fungal CE15 sequences (Fig. 6a). Indeed, residues involved in carbohydrate recognition are almost invariant among fungal CE15-B enzymes and most of these residues appear to be conserved in fungal CE15-A members. Currently, there are no known structures of fungal CE15-A members. However, one notable difference (on sequence level) between CE15-A and CE15-B in their residues involved in substrate recognition is the lack of E324 in fungal CE15-A members (Supplementary Fig. 6). E324 interacts with O2, and could play a role in the specificity for the  $\alpha$ -1,2-linked 4-*O*-methyl-glucuronoyl moiety. Noteworthy differences are also found at the position corresponding to Y366. In CE15-B, the latter is always a hydrophobic residue whereas CE15-A enzymes have their catalytic acid at this position. This creates distinct chemical differences between the two groups with respect to the properties of the pocket harboring the 4-*O*-methyl group (Fig. 6b) and could explain why CE15-A enzymes<sup>14</sup>, in contrast to CE15-B



**Fig. 6 Conservation of structural features implicated in recognition of natural substrates by fungal CE15 enzymes.** **a** Conservation of the carbohydrate-binding site. Residues within 5 Å from the XU<sup>m4</sup>X fragment are mapped onto the surface of CuGE. Corresponding sequence logos<sup>50</sup> are shown for the fungal subgroups CE15-A and CE15-B, respectively. The role(s) played by the individual residues in the CuGE complexes with aldouronic acids are summarized. **b** Distinct environments around the 4-O-Methyl group caused by the different configurations of the catalytic triads in CE15-A and CE15-B. The ligand from the dΔS270A:XU<sup>m4</sup>XX<sup>-OH</sup> structure is superimposed on the MZ0003 structure (6EHN)<sup>13</sup>. For clarity, only the 4-O-methyl-glucuronoyl-part of the ligand is shown. **c** Surface representations of CuGE (CE15-B) and MZ0003 (CE15-A), illustrating differences in surface topology. Notably, the αA-helix and “insert 2” are only present in the bacterial members of CE15-A. In all fungal enzymes (both CE15-A and CE15-B), a deletion in this region results in a very open, flat, and accessible-binding site. **d** Putative monolignol binding site. **e** CuGE-CBM1 homology model showing the conserved residues implicated in the adsorption onto insoluble substrates.



enzymes<sup>16,17</sup>, do not show a preference for 4-*O*-methylated substrates. A fungal CE15-A enzyme (*TtGE* from *Thielavia terrestris*) has been shown to exhibit lower activity on LCC-containing substrate compared CE15-B enzymes, but prolonged incubation results in the same product profile as CE15-B enzymes<sup>18</sup>.

### Surface properties and implications for lignin interactions.

Given the variable structure of lignin, a specific site for recognition of the alcohol moiety of the LCC substrate may not be present. Probably, the surfaces of GEs have evolved to have properties favorable for lignin interactions, rather than having conserved residues at specific positions. From a topological perspective we note that the *CuGE* fold lacks the  $\alpha/\beta$  hydrolase  $\alpha$ -helix (Fig. 3d–f), which leaves the fungal CE15 enzymes (both CE15-A and CE15-B) with a very open flat and accessible-binding site (Fig. 6c) that could allow interactions with ester substrates that have large and bulky alcohol fragments such as lignin. In *CuGE*, this putative lignin binding surface is particularly rich in serine and threonine residues. In fungal CE15-B structures, a small depletion exists at the lignin-side of the ester cleavage site, which is conserved at the structural level (Fig. 6d). Conserved residues that might interact with a monolignol moiety are S193, C269, H404, and C405.

In the CBM1 domain, a number of highly conserved residues form a flat, aromatic binding site (Fig. 6e and Supplementary Fig. 2b), which is assumed to be implicated in the absorption onto insoluble substrates, including LRP. While the SAXS study of *CuGE* shows an extended and relatively rigid linker region, the length of the linker varies significantly between different CE15 enzymes (Supplementary Fig. 7), indicating that a specific distance between the CBM1 and the catalytic domain is not crucial for the enzymatic function.

### Discussion

GEs have received much attention during the past 10 years, primarily due to their potential use in lignocellulose bioprocessing. The basis for understanding the action of GEs in heterogeneous biocatalysis relies on structural and functional knowledge of the enzymatic function. The fact that *CuGE* has high affinity for the insoluble natural LCC-containing substrate LRP, compared to soluble ligands, implies that large lignin-rich substrates are the natural substrate structures for *CuGE*. Our structural and functional studies of *CuGE* have provided the structural basis for understanding the enzyme-substrate interactions and the well-defined complexes reveal how *CuGE* recognizes and binds the carbohydrate moiety of the LCC. An interesting finding is that the 4-*O*-methylation occurring in natural glucuronoxylan is important for fitting the substrate into the active site of the enzyme. *CuGE* possesses a flat open binding surface, where it is possible to accommodate even highly decorated xylans. In light of this binding mode, where the enzyme recognizes and binds the carbohydrate moiety of the LCC, it seems justified to conclude that the glucuronoyl esterases rightfully belong among the carbohydrate active enzymes in the CAZy database, acting on substrate on the interface between plant cell wall carbohydrates and the heterogeneous polyphenolic structures of lignin. The overall division of CE15s into two structurally different subgroups of  $\alpha/\beta$ -hydrolases provides a clear and previously unnoted distinction between the CE15 enzymes and could indicate that the two subgroups evolved for two different enzymatic roles. While the fungal enzymes belonging to CE15-B appear highly conserved and specialized in recognizing and binding ester-linked LCCs from glucuronoxylan, the CE15-A variants seem to represent a subgroup of enzymes with broader substrate specificity and

possibly even other types of activities, indicating that CE15-A enzymes have evolved to more generic types of esterases.

### Methods

**Preparation of lignin-rich precipitate from birchwood.** LCC-containing substrate, originating from milled birchwood, was prepared by a mild organosolv type of extraction with 50 vol-% ethanol and 10% dry matter at 180 °C for 1 h. The liquid phase after organosolv extraction was precipitated with water to obtain a lignin-rich precipitate (LRP). A detailed method description is found in Mosbeck et al.<sup>5</sup>. The LRP contains mainly lignin and minor amounts of covalently linked glucuronoxylan as seen from the compositional analysis of LRP, which showed that the precipitate contained ~90% Klason lignin and ~1.5% glucuronoxylan<sup>5</sup>. The concentration of 4-*O*-methyl-glucuronoyl residues was determined to be 2.39 mg g<sup>-1</sup> dry matter, which corresponds to ~1.10<sup>-2</sup>  $\mu$ mol 4-*O*-methyl-glucuronoyl mg LRP<sup>-1</sup>.

**Protein expression and purification.** Three different variants of the CE15 glucuronoyl esterase *CuGE* from *Cerrera unicolor* (Genbank accession number: MK422512) were recombinantly expressed in *Pichia pastoris* X-33; all with a C-terminal hexa histidine (His<sub>6</sub>) tag for affinity purification. (i) WT: full-length wild type, (ii)  $\Delta$ WT: truncated *CuGE* (residues G79-T458) corresponding to the catalytic domain only, and (iii)  $\Delta$ S270A: inactive truncated variant where the catalytic nucleophile S270 was replaced by alanine. The  $\Delta$ WT and  $\Delta$ S270A constructs were designed with a TEV protease cleavage site upstream of the His<sub>6</sub> tag to enable removal of the affinity tag from the purified protein. The genes encoding the *CuGE* variants were codon-optimized for *P. pastoris*, prepared by GenScript and delivered in pPICZaA vectors. Different restriction sites were used for insertion of the constructs into the vector (*EcoRI* and *XbaI* for WT vs. *EcoRI* and *SalI* for  $\Delta$ WT and  $\Delta$ S270A), resulting in a C-terminal extension with a *c-myc* epitope upstream of the His<sub>6</sub> in the WT variant only. The three *CuGE* variants were produced in *P. pastoris* and purified following the procedure described earlier<sup>5</sup>. The proteins were purified by affinity chromatography on an IMAC-column (HisTrap HP 5 mL column, GE Healthcare) using an Äkta Purifier 100 system (GE Healthcare, Uppsala Sweden).

**Deglycosylation.** The affinity-purified samples displayed very smeared bands on sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) indicating heterogeneous N- and/or O-glycosylation<sup>17</sup>. Therefore, enzyme preparations to be used for structural studies were enzymatically deglycosylated and further purified by size-exclusion chromatography (SEC). In this procedure, the elution fractions from the IMAC step were immediately buffer exchanged into 20 mM sodium acetate pH 5 using PD-10 desalting columns (GE Healthcare) followed by concentration on Vivaspinn 20 filter units with a 10 kDa MWCO (Sartorius). To trim the N-glycosylation at the single consensus site N104<sup>17</sup>, 42 mg purified  $\Delta$ S270A were treated with 1U Endoglycosidase H (Endo H, Roche) at room temperature for 17 h in a total reaction volume of 8.5 mL. Similarly, 35 mg purified full-length WT enzyme was treated with 1U Endo H at room temperature in a total reaction volume of 9.8 mL. After 17 h of Endo H reaction, 260U  $\alpha$ -1,2,3,6 mannosidase (P0768, New England Biolabs Inc.) was added in order to trim the O-glycosylation in the linker region<sup>17</sup> and the deglycosylation reaction was allowed to proceed for a total of 52 h. The final purification of the deglycosylated samples,  $\Delta$ S270A and dWT, was performed on a HiLoad 26/600 Superdex 75 column (GE Healthcare) equilibrated with 20 mM HEPES pH 7. Selected SEC fractions were pooled, concentrated to ~10–15 mg mL<sup>-1</sup> and stored at 4 °C until use. Protein concentrations were determined based on A<sub>280</sub> measured with a Nanodrop 2000 (ThermoFisher Scientific) instrument and the calculated molar extinction coefficients ( $\Delta$ S270A  $\epsilon_{280} = 82765 \text{ M}^{-1} \text{ cm}^{-1}$ ; dWT  $\epsilon_{280} = 101005 \text{ M}^{-1} \text{ cm}^{-1}$ ).

**Stability measurements.** In order to evaluate the stability of the different *CuGE* variants apparent unfolding temperatures (*T*<sub>u</sub>) were monitored based on changes in intrinsic fluorescence using a Tycho NT.6 instrument (NanoTemper Technologies). The measurements were made in 20 mM sodium acetate buffer pH 5, protein concentrations in the range 0.07–0.3 mg mL<sup>-1</sup>.

**Crystallization, data collection, and structure determination.** The deglycosylated *CuGE* variants dWT and  $\Delta$ S270A were both screened for crystallization. However, all crystals obtained contain the catalytic domain only, as crystals formed in experiments with the full-length enzyme (dWT) turned out to be a proteolytic fragment lacking the CBM1 and linker region ( $\Delta^*$ dWT). Ligand complexes were prepared by soaking aldouronic acid ligands into crystals of either the inactive variant  $\Delta$ S270A or the active variant  $\Delta^*$ dWT. Crystallization and ligand soak conditions are summarized in Supplementary Table 1. The vast majority of crystals formed at 4 °C and with high molecular weight polymers (SOKALAN CP7, PEG 10,000 or PEG 8000) as precipitants. Crystals were cryoprotected by brief soaks in either reservoir or ligand stock solutions supplemented with 26% glycerol and flash-frozen in liquid nitrogen. X-ray diffraction data were collected at cryogenic temperature (100 K) using the following beamlines: MASSIF-3 (ID30A-3) at ESRF, Grenoble, France (apo- $\Delta$ S270A,  $\lambda = 0.9677 \text{ \AA}$ ); BioMAX at MAX IV, Lund, Sweden (apo- $\Delta^*$ dWT,  $\lambda = 0.91837 \text{ \AA}$ ); P13 at EMBL, DESY, Hamburg, Germany ( $\Delta$ S270A:U<sup>m</sup>4X,  $\lambda = 0.97625 \text{ \AA}$ ), ID29 at ESRF, Grenoble, France ( $\Delta$ S270A:

$U^{m4}XX-OH$  and  $d\Delta S207A:XU^{m4}XX-OH$ ,  $\lambda = 0.99987 \text{ \AA}$ ) and ID30B at ESRF, Grenoble, France ( $\Delta^*dWT:U^{m4}XX-OH$ ,  $\lambda = 0.97625 \text{ \AA}$ ). Data collection and refinement statistics are given in Supplementary Table 2. The data sets were processed using XDS<sup>29</sup>. All crystals belong to the tetragonal space group  $P4_12_12$  with two molecules in the asymmetric unit. The structure of apo- $d\Delta S207A$  was solved by molecular replacement with PHASER<sup>30</sup> using a trimmed version of the structure of Cip2 from *H. jecorina* (3PIC<sup>9</sup>, 54% seq. id.) as search model. The initial model was rebuilt and extended using the AutoBuild<sup>31</sup> procedure in PHENIX<sup>32</sup>. The model was further rebuilt, extended and refined in iterative cycles with Coot<sup>33</sup> and PHENIX<sup>32,34</sup>, respectively. A refinement protocol using the maximum-likelihood target, real-space refinement, refinement of coordinates, individual isotropic temperature factors and translation-liberation-screw (TLS) refinement using TLS groups determined by the PHENIX software was applied. All other *CuGE* structures were solved by molecular replacement with PHASER using apo- $d\Delta S207A$  as search model and refined in a similar manner as the apo- $d\Delta S207A$  structure. All six *CuGE* structures show the complete catalytic domain (C81-T458), whereas the terminal extensions are generally not ordered. However, the two structures obtained from the SOKALAN CP7 crystallization condition (apo- $d\Delta S207A$  and  $d\Delta S207A:U^{m4}X$ ; Supplementary Table 1) are particularly well-defined and an N-terminal Glu-rich extension "EAEAEF" could be deduced from the electron density map. This sequence is derived from the pPICZaA expression vector and shows that apparently only *kex2* cleavage, but not STE13 cleavage, has occurred when recombinant *CuGE* was produced in the *P. pastoris* expression system. All structures contain one NAG residue at N104, while the two structures originating from crystallization experiments with full-length *CuGE* (apo- $\Delta^*dWT$  and  $\Delta^*dWT:U^{m4}XX-OH$ ) additionally contain a MAN residue at both S86 and T87, indicating that these sites only get O-glycosylated when the full-length enzyme is expressed. Soaks with aldouronic acids resulted in very well-defined complexes with the ligand-bound in the more accessible molecule B only. In the  $d\Delta S207A:U^{m4}X$  structure, which was obtained by soaking with a mixture of aldouronic acids (O-AMX, Megazymes), positive difference density suggested the presence of a partially occupied xylose moiety at the reducing end of the  $U^{m4}X$  ligand. This residue was, however, not included in the final model as the occupancy was too low to achieve a robust refinement. The  $\Delta^*dWT:U^{m4}XX-OH$  structure displayed continuous electron density between the catalytic nucleophile S270 and C6 of the aldouronic acid ligand and was therefore modeled as a covalent reaction intermediate; the S270(OG)-C6 distance refined to 1.62 Å. An ethylene glycol (EDO) molecule was observed in the substrate-binding pocket of all unliganded structures (molecule A and apo-structures). At the final stages of refinement, positive difference density suggested the presence of a secondary surface binding site in the structures of  $d\Delta S207A:U^{m4}XX-OH$  and  $\Delta^*dWT:U^{m4}XX-OH$ . This was modeled as a partially occupied xylobiose molecule (BXP). The data to high resolution and the complex composition of the solvent region comprising water, glycerol, ethylene glycol molecules, and the protein side chains in weakly populated alternative conformations made it non-trivial to model of the electron density in the solvent region. Only fully occupied solvent molecules were included in the final models, however, a number of partial occupied solvent molecules are visible in the electron density maps.

The Ramachandran statistics calculated by MolProbity<sup>35</sup> for the final refined models are as follows: 96.3% in the favored region, 3.7% allowed, 0% outliers for apo- $d\Delta S207A$ ; 96.0% in the favored region, 4.0% allowed, 0% outliers for apo- $\Delta^*dWT$ ; 96.2% in the favored region, 3.8% allowed, 0% outliers for  $d\Delta S207A:U^{m4}X$ ; 96.1% in the favored region, 3.9% allowed, 0% outliers for  $d\Delta S207A:U^{m4}XX-OH$ ; 96.3% in the favored region, 3.7% allowed, 0% outliers for  $d\Delta S207A:XU^{m4}XX-OH$  and 96.3% in the favored region, 3.7% allowed, 0% outliers for  $\Delta^*dWT:U^{m4}XX-OH$ , respectively. Illustrations were prepared with PyMOL (<http://www.pymol.org>).

**Small-angle X-ray scattering.** Prior to the analysis by small-angle X-ray scattering (SAXS), the deglycosylated variants of full-length (dWT) and truncated, inactive *CuGE* ( $d\Delta S207A$ ) were dialyzed into 20 mM sodium acetate pH 5 and centrifuged at  $14,000 \times g$  for 10 min at 4 °C. The permeates were used as reference solutions. For each protein, three different dilutions were prepared in 20 mM sodium acetate buffer pH 5 (dWT 3.7, 2.0, and 1.0 mg mL<sup>-1</sup> and  $d\Delta S207A$  3.6, 2.0 and 1.0 mg mL<sup>-1</sup>). SAXS measurements were performed on a Xenocs BioXolver L equipped with a GeniX3D microfocuss X-ray source ( $\lambda = 1.54 \text{ \AA}$ ) and a BioCUBE sample environment. A sample-to-detector distance of 571 mm was used, corresponding to values of the scattering vector  $q$  between 0.019 and  $0.48 \text{ \AA}^{-1}$ . Samples were kept and measured at room temperature, automatic loading was done using the sample handling robot from a 96-well tray. Scattering data was collected as 10–20 frames (depending on sample concentration) of 60 s exposure pr. frame for protein samples, as well as corresponding buffers. Two-dimensional (2D) images were radially averaged, and overlap of individual frames ensured before merging. Buffer scattering was subtracted from the total scattering. In addition, a bovine serum albumin (BSA) solution was measured for reference and molecular weight estimation. All initial data processing was done using the software RAW<sup>36,37</sup>. Subsequent primary data analysis, Guinier fit, indirect Fourier transformation and molecular weight estimation, was done using Primus<sup>38</sup>. For  $d\Delta S207A$  the solution scattering data was compared to the crystal structure of apo- $d\Delta S207A$  (molecule A) using CRY SOL<sup>39</sup>. Ab-initio shape reconstruction based on the data from dWT was performed using both DAMMIF<sup>40</sup> and GASBOR<sup>41</sup>, in both cases using the  $q$  range

from 0.02 to  $0.25 \text{ \AA}^{-1}$ . DAMMIF was run in default slow mode generating ten models. After alignment (mean NSD = 0.612), the models were averaged and refined in DAMMIN<sup>42</sup>. GASBOR was run with default settings fitting intensity in reciprocal space. Illustrations of the structural models were prepared with PyMOL (<http://www.pymol.org>).

**Homology model.** A homology model of *CuGE*-CBM1 (residues 3–38) was made with SWISS-MODEL<sup>43</sup> using the structure of the CBM1 from *Trichoderma reesei* cellobiohydrolase I (PDB code 1CBH<sup>44</sup>) as template. The sequence identity between *CuGE*-CBM1 and 1CBH is 64%.

**Microscale thermophoresis.** Binding affinities for the soluble ligands  $U^{m4}XX-OH$  (O-UXXR, MEGAZYMES), MeGlcA (4-O-methyl-D-glucuronic acid, MM46686, Carbosynth, Crompton UK), and MeMeGlcA (4-O-methyl-D-glucopyranosyluronate, purchased from Institute of Chemistry, Slovak Academy of Sciences, Bratislava, Slovakia) were measured by MST. The protein stocks,  $d\Delta S207A$  and  $\Delta WT$ , were adjusted to 200 nM with 100 mM sodium acetate buffer pH 5 supplemented with 0.1% Pluronic F-127. Ligand stocks (100 mM  $U^{m4}XX-OH$ , 43 mM MeGlcA and 250 mM MeMeGlcA) were prepared in 100 mM sodium acetate buffer pH 5, and for each ligand a series of 16 1:1 dilutions was prepared using the same buffer. For the measurement, each ligand dilution was mixed with one volume of protein, which led to a final protein concentration of 100 nM and final ligand concentrations in the nM to nM range. After 10 min incubation followed by centrifugation at  $10,000 \times g$  for 10 min, the samples were loaded into Monolith NT.LabelFree zero-background capillaries (NanoTemper Technologies). The MST experiment was performed with a Monolith NT.LabelFree instrument (NanoTemper Technologies) at 25 °C using 30% LED power and medium MST power. For MeGlcA and MeMeGlcA, the higher ligand concentrations resulted in significant variations in raw fluorescence and these data were therefore not included in the analysis. Final ligand concentration ranges used in the analysis of the MST signal were:  $U^{m4}XX-OH$  25 mM to 1.53  $\mu M$ , MeGlcA 10.8 mM to 656 nM and MeMeGlcA 15.6 mM to 3.81  $\mu M$ . For each ligand-binding curve, data from at least three independently pipetted measurements were included in the analysis with the MO. Affinity Analysis software (version 2.3, NanoTemper Technologies) using the signal from an MST-on time of 5 s.

**Isothermal titration calorimetry measurements.** Binding affinity of full-length WT *CuGE* to the insoluble lignin-rich precipitate (LRP) from birchwood was determined by ITC by measuring the heat generated upon binding of the enzyme to a ligand, in this case LRP. The measurements were performed on a NanoITC 2G instrument from TA instruments (Lindon, USA) equipped with a 950  $\mu L$  titration cell and an equal size reference cell. The reference cell was loaded with 20 mM MES buffer pH 6 throughout all titrations. Prior to ITC experiments, the WT *CuGE* preparation was buffer exchanged into 20 mM MES buffer pH 6 on a spin filter with a polyethersulfone membrane with a 10 kDa MWCO (Vivaspin 20, Sartorius). The buffer was exchanged three times prior to final round of ultrafiltration. Permeate from the last round was used as suspension buffer for LRP in the titration experiments to obtain near-identical environments in cell and syringe. Nine-hundred fifty microliters LRP-suspension ( $3.5 \text{ mg mL}^{-1}$ , corresponding to  $\sim 0.04 \text{ mM}$  4-O-methyl-glucuronoyl residues) in permeate was degassed for 5 min and loaded directly to the titration chamber. The sample syringe was loaded with 250  $\mu L$  0.2 mM WT preparation and the syringe needle was placed inside the titration cell. The needle of the sample syringe was equipped with impellers for constant rotational stirring during the titration and stirring speed was set to 300 rpm. The titration was performed at 5 °C, where hydrolysis was found to be negligible within the time frame of the assay and the enzyme preparation was added over the course of 25 injections of 10  $\mu L$  each. The time intervals between each injection was 300 s and prior to the first injection the system was allowed to equilibrate until steady baseline was achieved, however no longer than 60 min. A control experiment with permeate buffer in the sample syringe instead of enzyme preparation was performed with the same settings to assess the level of heat generation not associated with enzyme-ligand binding. The results of the enzyme-ligand measurements were corrected for background heat generation by subtracting the control measurements prior to data handling. All data handling was performed using NanoAnalyze Data Analysis v. 3.8.0 provided by TA instruments and model fitting was performed by the use of an independent data model provided by the software.

**Binding to insoluble biopolymers, pull-down assay.** To assess the effect of the CBM1 module on the binding of *CuGE* to insoluble biopolymers a pull-down assay was performed. Suspensions of  $10 \text{ mg mL}^{-1}$  biopolymer, including Avicel® PH-101 cellulose (Sigma-Aldrich), xylan from birchwood (Sigma-Aldrich), and LRP, were prepared in 50 mM sodium acetate buffer pH 5. Full length and truncated *CuGE* variants, WT and  $\Delta WT$ , respectively, were added to a final concentration of 0.07 mg mL<sup>-1</sup> in a total reaction volume of 550  $\mu L$ . The samples were incubated for 1 h at 4 °C with rotation. After incubation the reaction tubes were centrifuged at 14,000 rpm for 10 min to separate pellet and supernatant. The pellet, containing protein bound to the insoluble biopolymer, was washed in water twice and then re-suspended in 550  $\mu L$  water. To increase the protein concentration all samples were freeze dried and re-suspended in 100  $\mu L$  prior to analysis. Samples were reduced

with DTT and analyzed by SDS-PAGE using Mini-PROTEAN® TGX™ Precast Gels (Bio-RAD).

**Topology diagrams and CE15 subgroups.** The topology diagrams were made in PowerPoint based on the description of the classical  $\alpha/\beta$ -hydrolase fold<sup>23–25</sup> combined with a manual inspection of all structurally characterized CE15 enzymes (PDB codes 6RTV, 4G4G<sup>12</sup>, 3PIC<sup>9</sup>, 6EHN<sup>13</sup>, 6GSO<sup>14</sup>, 6GRY<sup>14</sup>, and 6HSW<sup>15</sup>). The feruloyl esterase Est1 from *Butyrivibrio proteoclasticus* (PDB code 2WTM<sup>45</sup>) was included as a reference for an  $\alpha/\beta$ -hydrolase Group A fold and a canonical oxyanion hole. With reference to the  $\alpha/\beta$ -hydrolase Groups A and B<sup>25</sup>, the CE15 members were assigned to the subgroups CE15-A or CE15-B, respectively, depending on the position of the catalytic acid after either strand  $\beta 8$  (CE15-A) or  $\beta 7$  (CE15-B). The structure of the bacterial MZ0003<sup>13</sup> was used as representative for CE15-A, as no fungal CE15-A member has so far been structurally characterized, and the CuGE structure was used as representative of CE15-B. For clarity, only the major secondary structural elements were included in the topology diagrams. Structures from both CE15 subgroups contain additional small secondary structural elements, encompassing multiple helical elements inserted after the strands  $\beta 5$  and  $\beta 7$  (Supplementary Fig. 4). The bacterial CE15-A structures comprise the canonical  $\alpha/\beta$ -hydrolase helix  $\alpha A$ , as well as the three inserts I1, I2, and I3. However, a comparison with the sequence alignments in Supplementary Figs. 4–6 suggests that these structural elements are not present in the fungal CE15-A members.

**Sequence analysis.** Multiple sequence alignments were made with ProMALS3D<sup>46,47</sup>, further edited by hand and visualized using Jalview<sup>48</sup>. The sequence logos, were based on a dataset corresponding to the sequences used by Dilokpimol<sup>49</sup> complemented with the AfuGE sequence<sup>18</sup>. A subset of binding site residues was defined, within 5 Å from the XU<sup>m4</sup>X fragment in the d $\Delta$ S207A: XU<sup>m4</sup>XX-OH structure (PDB code 6RV9), and residues at the corresponding positions were extracted from the alignment comprising all fungal CE15 sequences in the dataset (Supplementary Figs. 5 and 6). Sequence logos were produced with WebLOGO<sup>50</sup> using a total of 20 sequences for the fungal subgroup CE15-A and 116 sequences for the fungal subgroup CE15-B, respectively.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The data that support this work are available from the corresponding authors upon reasonable request. The SAXS data and corresponding models presented in the manuscript has been deposited at SASBDB ([www.sasbdb.org](http://www.sasbdb.org)) with accession codes: SASDGD6 and SASDGC6. The crystal structures presented in this manuscript have been deposited in the Protein Data Bank with the PDB codes 6RTV, 6RU1, 6RU2, 6RV7, 6RV8 and 6RV9, respectively. It must be noted that the SAXS and the PDB models differ in their sequence numbering. The SAXS model use UNIPROT numbering, whereas the PDB models refer to the sequence numbering illustrated in Fig. 1. The source data underlying Fig. 1c, 5a–d, and Supplementary Fig. 1 are provided as a Source Data file.

Received: 16 September 2019; Accepted: 6 February 2020;

Published online: 24 February 2020

## References

- Terrett, O. M. & Dupree, P. Covalent interactions between lignin and hemicelluloses in plant secondary cell walls. *Curr. Opin. Biotechnol.* **56**, 97–104 (2019).
- Giummarella N., Pu, Y., Ragauskas, A. J. & Lawoko, M. A critical review on the analysis of lignin carbohydrate bonds. *Green. Chem.* **21**, 1573–1595 (2019).
- Ragauskas, A. J. et al. Lignin valorization: improving lignin processing in the biorefinery. *Science* **344**, 1246843 (2014).
- Balakshin, M., Capanema, E., Gracz, H., Chang, H. M. & Jameel, H. Quantification of lignin-carbohydrate linkages with high-resolution NMR spectroscopy. *Planta* **233**, 1097–1110 (2011).
- Mosbech, C., Holck, J., Meyer, A. S. & Agger, J. W. The natural catalytic function of CuGE glucuronoyl esterase in hydrolysis of genuine lignin-carbohydrate complexes from birch. *Biotechnol. Biofuels* **11**, 1–9 (2018).
- Agger, J. W., Busk, P. K., Pilgaard, B., Meyer, A. S. & Lange, L. A new functional classification of glucuronoyl esterases by peptide pattern recognition. *Front. Microbiol.* **8**, 309 (2017).
- Monrad, R. N., Eklof, J., Krogh, K. & Biely, P. Glucuronoyl esterases: diversity, properties and biotechnological potential. A review. *Crit. Rev. Biotechnol.* **38**, 1121–1136 (2018).
- Li, X. L., Spanikova, S., de Vries, R. P. & Biely, P. Identification of genes encoding microbial glucuronoyl esterases. *FEBS Lett.* **581**, 4029–4035 (2007).
- Pokkuluri, P. R. et al. Structure of the catalytic domain of glucuronoyl esterase Cip2 from *Hypocrea jecorina*. *Proteins* **79**, 2588–2592 (2011).
- Topakas, E., Moukoulis, M., Dimarogona, M., Vafiadi, C. & Christakopoulos, P. Functional expression of a thermophilic glucuronoyl esterase from *Sporotrichum thermophile*: identification of the nucleophilic serine. *Appl. Microbiol. Biotechnol.* **87**, 1765–1772 (2010).
- Holmquist, M. Alpha/Beta-hydrolase fold enzymes: structures, functions and mechanisms. *Curr. Protein Pept. Sci.* **1**, 209–235 (2000).
- Charavgi, M. D., Dimarogona, M., Topakas, E., Christakopoulos, P. & Chrysina, E. D. The structure of a novel glucuronoyl esterase from *Myceliophthora thermophila* gives new insights into its role as a potential biocatalyst. *Acta Crystallogr. D. Biol. Crystallogr.* **69**, 63–73 (2013).
- De Santi, C., Gani, O. A., Helland, R. & Williamson, A. Structural insight into a CE15 esterase from the marine bacterial metagenome. *Sci. Rep.* **7**, 17278 (2017).
- Arnling Baath, J. et al. Biochemical and structural features of diverse bacterial glucuronoyl esterases facilitating recalcitrant biomass conversion. *Biotechnol. Biofuels* **11**, 213 (2018).
- Baath, J. A. et al. Structure-function analyses reveal that a glucuronoyl esterase from *Teredinibacter turnerae* interacts with carbohydrates and aromatic compounds. *J. Biol. Chem.* **294**, 6635–6644 (2019).
- Spanikova, S., Polakova, M., Joniak, D., Hirsch, J. & Biely, P. Synthetic esters recognized by glucuronoyl esterase from *Schizophyllum commune*. *Arch. Microbiol.* **188**, 185–189 (2007).
- d’Errico, C. et al. Enzymatic degradation of lignin-carbohydrate complexes (LCCs): model studies using a fungal glucuronoyl esterase from *Cerrena unicolor*. *Biotechnol. Bioeng.* **112**, 914–922 (2015).
- Mosbech, C., Holck, J., Meyer, A. & Agger, J. W. Enzyme kinetics of fungal glucuronoyl esterases on natural lignin-carbohydrate complexes. *Appl. Microbiol. Biotechnol.* **103**, 4065–4075 (2019).
- Receveur, V., Czjzek, M., Schulein, M., Panine, P. & Henrissat, B. 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 (2002).
- Ruiz, D. M., Turowski, V. R. & Murakami, M. T. Effects of the linker region on the structure and function of modular GH5 cellulases. *Sci. Rep.* **6**, 28504 (2016).
- Kadowaki, M. A. S., Higasi, P., de Godoy, M. O., Prade, R. A. & Polikarpov, I. Biochemical and structural insights into a thermostable cellobiohydrolase from *Myceliophthora thermophila*. *FEBS J.* **285**, 559–579 (2018).
- de Araujo, E. A., Manzone, L. R., Piiadov, V., Kadowaki, M. A. S. & Polikarpov, I. Biochemical characterization, low-resolution SAXS structure and an enzymatic cleavage pattern of BICel48 from *Bacillus licheniformis*. *Int. J. Biol. Macromol.* **111**, 302–310 (2018).
- Ollis, D. THE alpha/beta-hydrolase fold. *Protein Eng.* **5**, 197–211 (1992).
- Carr, P. D. & Ollis, D. L. Alpha/beta hydrolase fold: an update. *Protein Pept. Lett.* **16**, 1137–1148 (2009).
- Dimitriou, P. S. et al. Alpha/beta-hydrolases: A unique structural motif coordinates catalytic acid residue in 40 protein fold families. *Proteins* **85**, 1845–1855 (2017).
- Dimitriou, P. S., Denesyuk, A. I., Nakayama, T., Johnson, M. S. & Denessiouk, K. Distinctive structural motifs co-ordinate the catalytic nucleophile and the residues of the oxyanion hole in the alpha/beta-hydrolase fold enzymes. *Protein Sci.* **28**, 344–364 (2019).
- Dilokpimol, A. et al. Fungal glucuronoyl esterases: Genome mining based enzyme discovery and biochemical characterization. *N. Biotechnol.* **40**, 282–287 (2018).
- Biely, P. Microbial glucuronoyl esterases: 10 years after discovery. *Appl. Env. Microbiol.* **82**, 7014–7018 (2016).
- Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–132 (2010).
- McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Terwilliger, T. C. et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D. Biol. Crystallogr.* **64**, 61–69 (2008).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221 (2010).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
- Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D. Biol. Crystallogr.* **68**, 352–367 (2012).
- Chen, V. B. et al. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 12–21 (2010).
- Nielsen, S. S. et al. BioXTAS RAW, a software program for high-throughput automated small-angle X-ray scattering data reduction and preliminary analysis. *J. Appl. Crystallogr.* **42**, 959–964 (2009).

37. Hopkins, J. B., Gillilan, R. E. & Skou, S. BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J. Appl. Crystallogr.* **50**, 1545–1553 (2017).
38. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS—a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Cryst.* **36**, 1277–1282 (2003).
39. Svergun, D., Barberato, C. & Koch, M. H. J. CRY SOL-A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
40. Franke, D. & Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346 (2009).
41. Svergun, D. I., Petoukhov, M. V. & Koch, M. H. J. Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80**, 2946–2953 (2001).
42. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **77**, 2896 (1999).
43. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
44. Kraulis, J. et al. Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry* **28**, 7241–7257 (1989).
45. Goldstone, D. C. et al. Structural and functional characterization of a promiscuous feruloyl esterase (Est1E) from the rumen bacterium *Butyrivibrio proteoclasticus*. *Proteins* **78**, 1457–1469 (2010).
46. Pei, J., Tang, M. & Grishin, N. V. PROMALS3D web server for accurate multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, W30–W34 (2008).
47. Pei, J., Kim, B. H. & Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, 2295–2300 (2008).
48. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
49. Dilokpimol, A. et al. Fungal feruloyl esterases: Functional validation of genome mining based enzyme discovery including uncharacterized subfamilies. *N. Biotechnol.* **41**, 9–14 (2018).
50. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

## Acknowledgements

Tobias Tandrup (remote collection of the dΔS270A dataset), Janina Sprenger (collection of the dΔS270A:U<sup>m4</sup>X dataset), Jens Christian Navarro Poulsen (collection of the dΔ<sup>1</sup>WT dataset), Christoph Mueller-Dieckmann ID30B, Fátima Herranz-Trillo (SAXS data collection), Blanca Lopez Mendez at the Novo Nordisk Foundation Center for Protein Research for the introduction and access to the MST instrument (NT.LabelFree). The work has been financed by the Novo Nordisk Foundation (grant NNF15OC0015222), the

Bio-Value Strategic Platform for Innovation and Research, co-funded by The Danish Council for Strategic Research and The Danish Council for Technology and Innovation, case no: 0603-00522B and The Lundbeck Foundation Initiative BRAINSTRUC.

## Author contributions

H.A.E. J.W.A., C.M., A.M., and S.L. developed the conceptual framework for this work. H.A.E. planned and conducted all experiments in relation to crystallography, and H.A.E. and S.L. performed interpretations on crystallography data. C.M. and J.W.A. provided all protein and appropriate ligands for the crystal complexes. J.W.A., C.M., and P.W. designed, performed, and interpreted ITC and pull-down assays, H.A.E. performed and interpreted MST assays. A.E.L., H.A.E., and S.L. designed and performed SAXS experiments and data interpretation. All authors contributed in preparation of the manuscript.

## Competing interests

The authors declare no competing interests

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-020-14833-9>.

**Correspondence** and requests for materials should be addressed to J.W.A. or S.L.

**Peer review information** *Nature Communications* thanks Igor Polikarpov, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020