A carbohydrate-binding family 48 module enables feruloyl esterase action on polymeric arabinoxylan

Received for publication, June 20, 2019, and in revised form, September 17, 2019 Published, Papers in Press, September 26, 2019, DOI 10.1074/jbc.RA119.009523

Jesper Holck‡, Folmer Fredslund§, Marie S. Møller*‡, Jesper Brask§, Kristian B. R. M. Krogh*, Lene Lange†, Ditte H. Welner‡, Birte Svensson†, Anne S. Meyer‡, and Casper Wilkens‡§

From the ‡Section for Protein Chemistry and Enzyme Technology, Department of Biotechnology and Biomedicine, Technical University of Denmark, Saltofts Plads, Building 224, DK-2800 Kongens Lyngby, the §Enzyme Engineering and Structural Biology, Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, Building 220, DK-2800 Kongens Lyngby, the †Novozymes A/S, BIOLOGIENS VEJ 2, DK-2800 Kongens Lyngby, Denmark, and the ‡La-Bioeconomy, Research and Advisory, KARENSGADE 5, DK-2500 Valby, Denmark

Edited by Gerald W. Hart

Feruloyl esterases (EC 3.1.1.73), belonging to carbohydrate esterase family 1 (CE1), hydrolyze ester bonds between ferulic acid (FA) and arabinose moieties in arabinoxylans. Recently, some CE1 enzymes identified in metagenomics studies have been predicted to contain a family 48 carbohydrate-binding module (CBM48), a CBM family associated with starch binding. Two of these CE1s, wastewater treatment sludge (wts) Fae1A and wtsFae1B isolated from wastewater treatment surplus sludge, have a cognate CBM48 domain and are feruloyl esterases, and wtsFae1A binds arabinoxylan. Here, we show that wtsFae1B also binds to arabinoxylan and that neither binds starch. Surface plasmon resonance analysis revealed that wtsFae1B's $K_d$ for xylohexaose is 14.8 μM and that it does not bind to starch mimics, β-cyclodextrin, or maltohexaose. Interestingly, in the absence of CBM48 domains, the CE1 regions from wtsFae1A and wtsFae1B did not bind arabinoxylan and were also unable to catalyze FA release from arabinoxylan. Pretreatment with a β-o-1,4-xylanase did enable CE1 domain-mediated FA release from arabinoxylan in the absence of CBM48, indicating that CBM48 is essential for the CE1 activity on the polysaccharide. Crystal structures of wtsFae1A (at 1.63 Å resolution) and wtsFae1B (1.98 Å) revealed that both are folded proteins comprising structurally-conserved hydrogen bonds that lock the CBM48 position relative to that of the CE1 domain. wtsFae1A docking indicated that both enzymes accommodate the arabinoxylan backbone in a cleft at the CE1–CBM48 domain interface. Binding at this cleft appears to enable CE1 activities on polymeric arabinoxylan, illustrating an unexpected and crucial role of CBM48 domains for accommodating arabinoxylan.

Arabinoxylans (AXs)§ are hemicellulose components of vascular plant cell walls. AXs are abundantly present in many key lignocellulosic biomass feedstocks and cereal-processing residues, such as wheat bran, distiller’s dried grains, and brewer’s spent grain. Enzymatic modification and degradation of AXs are important in several cereal processes, and efficient bioconversion of AXs is also imperative in the development of new sustainable biomass and biorefinery processes ranging from production of biofuels and biobased materials to xylan-based prebiotic food and feed ingredients (1–3). The AX backbone is composed of β-1,4-linked xylopyranose (Xylp) residues that may be acetylated, substituted with glucuronic acid, and/or single-substituted with either α-l-1,2- or α-l-1,3-linked arabinofuranosyl (Araf) moieties or double-substituted with both α-l-1,2- and α-l-1,3-Araf. The amount and pattern of the Araf substitutions vary among species and also their tissue, e.g. single α-l-1,3-Araf substitutions are predominant in grass (monocot) cell walls, whereas double substitutions are also present in the endosperm. In contrast, dicots predominantly have single α-l-1,2-Araf substitutions (4), although double substitutions have been observed in flax mucilage (5), and single α-l-1,3-Araf substitutions have been found in psyllium seeds (6). The Araf moieties can be further substituted in grasses and also in other plants with e.g. 5-O-ferulic acid (FA) and other hydroxycinnamic acids (4).

Complete degradation of AX requires a battery of enzymes, including feruloyl esterases (7). Feruloyl esterases catalyze the
Arabinoxylan-binding CBM48

hydrolysis of ester bonds between FA and arabinose (8). They are found in both bacteria and fungi and are grouped in carbohydrate esterase family 1 (CE1) in the Carbohydrate-Active Enzyme database (CAZy) (9). All CE1 feruloyl esterases with known structure share the α/β-hydrolase fold of a central β-sheet flanked by α-helices (8). The catalytic triad is located in a hydrophobic binding pocket that is often capped by a flexible lid (10–13).

Carbohydrate-binding modules (CBMs) are noncatalytic, individually-folded domains that are appended to the catalytic enzyme module via a linker. Like the carbohydrate-active enzymes, the CBMs are categorized into families in the CAZy database (9). CBMs have been shown to be important for the functionality of some catalytic domains by bringing the enzymes into contact with their polymeric substrate; this contact can increase both the hydrolytic rate of the enzyme and the overall conversion rate of substrate because the effective concentration of the enzyme–substrate complex is increased (14–17). Generally, CBMs are small domains consisting of ~100–150 residues and are often composed of a β-sandwich fold (18). Recently, CE1 feruloyl esterases with appended CBMs annotated as members of carbohydrate-binding module family 48 (CBM48) were identified in two metagenomics studies of beaver and moose droppings and wastewater treatment surplus sludge, which are not considered particularly rich in starch (19, 20). CBM48, however, is usually associated with starch binding (21), but starch polysaccharides are not known to contain hydroxycinnamic acids. We therefore hypothesized that these newly discovered CBM48s represent a novel type of CBM that might enable the feruloyl esterase to act on polymeric non-starch substrates and thus positively affect the function and kinetics of the feruloyl esterase. Another albeit less likely idea is that the CE1 feruloyl esterases identified would have other hitherto unknown non-FA–ester linkage targets. We selected feruloylated AX as a primary substrate type to investigate the hypothesis that the two newly-discovered CE1 feruloyl esterases and their CBM48 appendages would in fact act on AXs. We also determined the crystal structures of these two CE1 feruloyl esterases and their appended CBM48 domains in order to examine the structure–function relations of such action. The objective of this study was therefore to resolve the structure–function relations of two CE1s and their cognate CBM48s identified in a metagenomic study of anaerobic digesters fed with surplus sludge from wastewater treatment (19).

Hence, in this study, two CE1 feruloyl esterases with CBMs classified as belonging to CBM48, and with truncations lacking the CBM48 domains, were characterized to demonstrate their specificity for AX and that the CBM48 domain is essential for activity on polymeric AX. The crystal structures of the two CE1 feruloyl esterases, wtsFae1A and wtsFae1B, were solved and subjected to docking with a feruloylated arabinoxylooligosaccharide (AXOS) to identify the putative binding site particularly for the AX backbone. Additionally, molecular dynamics analysis was performed to obtain an understanding of the rigidity of the CE1 and CBM48 domains. Finally, the relationship of these CBMs to known starch-binding CBMs was investigated to reveal whether these CBM48s could belong to a new, non-starch-binding CBM48 CAZy subfamily.

Results

Sequence analysis and expression

Both wtsFae1A and wtsFae1B were identified in a metagenomic study of an anaerobic digester fed with surplus wastewater treatment sludge (wts is an abbreviation for wastewater treatment sludge) (19). wtsFae1A is the N-terminal part of a chimeric enzyme we previously studied; the larger chimeric, or triple, enzyme, in addition to the terminal wtsFae1A, also has a glycoside hydrolase family 62 α-L-arabinofuranosidase (EC 3.2.1.55) and a glycoside hydrolase family 10 β-1,4-d-xylanase (EC 3.2.1.8) (22). wtsFae1A and wtsFae1B are 343 and 386 amino acids in length, including 28 and 24 predicted signal peptide residues, respectively. The molecular masses of mature wtsFae1A and wtsFae1B, including their cognate CBMs, were calculated to be 39 and 43 kDa, respectively, and the migration of the bands in the SDS-polyacrylamide gel corresponded to these molar masses (Fig. S1). Previous attempts to purify recombinant truncated wtsFae1A that lacks CBM48 (wtsFae1ACBM48) failed due to precipitation (22). This precipitation was avoided, however, by maintaining wtsFae1ACBM48 in the elution buffer from the His-tag purification step.

Domain analysis using dbCAN (23) suggested that wtsFae1B contains an N-terminal CBM48 and a C-terminal CE1 domain; previously, this domain organization has been suggested for wtsFae1A, but the family of the CBM was unknown (22). wtsFae1A and wtsFae1B display 43% pairwise sequence identity with each other (see Fig. S2 for multiple alignment), and both share the highest sequence identity with XynZ from Hungateiclostridium thermocellum. When compared with other structure-determined CE1 feruloyl esterases (see Fig. S2 for multiple alignment), their identity was only 39 and 38%, respectively.

Substrate specificity and kinetics

The ability of wtsFae1A and wtsFae1B to bind to wheat starch and insoluble wheat arabinoxylan (WAX-I), respectively, was investigated by adsorption assays. The data demonstrated that none of the enzymes had affinity for wheat starch because no binding could be observed. The apparent binding affinity ($K_d,_{app}$) of wtsFae1A for WAX-I was previously determined to be 204 ± 0.017 mg ml$^{-1}$ (22), whereas wtsFae1B displayed a $K_d,_{app}$ of 1.3 ± 0.16 mg ml$^{-1}$ for WAX-I (Fig. 1A).

To further investigate the substrate-binding properties of wtsFae1A and wtsFae1B, they were subjected to surface plasmon resonance (SPR) analysis using xylobiose, xylotriose, xylotetraose, xyloligosaccharides except for the starch analogs β-cyclodextrin and maltotetraose, and maltotriaose, maltotetraose, and β-cyclodextrin as binding targets. Unfortunately, SPR analysis for wtsFae1A did not yield meaningful data, which most likely reflected the enzyme’s poor stability when immobilized on the SPR-analysis chip. However, wtsFae1B displayed weak affinity for the above-mentioned oligosaccharides except for the starch. The $K_d,_{app}$ value was obtained for xylotetraose (Fig. 1B) and found to be 14.8 μM.

5 Please note that the JBC is not responsible for the long-term archiving and maintenance of this site or any other third party hosted site.
Previously, wtsFae1A was shown to have displayed minimal activity against WAX-I (22). But by using higher enzyme concentrations, which was achieved by redissolution of crystals formed during storage at 4 °C, wtsFae1A was shown to catalyze release of FA from WAX-I (Table 1). wtsFae1B also catalyzed release of FA from WAX-I (Table 1), but no such activity was detected for wtsFae1AΔCBM48 and wtsFae1BΔCBM48 (i.e. the enzymes devoid of the CBM) (Table 1). WAX-I may contain ferulate dehydrodimers (di-FAs) (4), but none of the enzymes released any di-FA from WAX-I. Interestingly, wtsFae1AΔCBM48 and wtsFae1BΔCBM48 displayed about 6- and 4-fold higher specific activity, respectively, than the corresponding full-length enzymes toward soluble ferulated AXOS obtained by pretreating WAX-I (Table 1). The space mined to 1.63 and 1.91 Å resolution, respectively, and the data-group for both crystals a dimer in solution (Fig. S3), which suggests that the enzymes eluted from agar-filtration column at a volume consistent with 19.9 and 22.1 kcal mol⁻¹ (Table 1). Overall structures of wtsFae1A and wtsFae1B

The crystal structures of wtsFae1A and wtsFae1B were determined to 1.63 and 1.91 Å resolution, respectively, and the data-processing statistics are summarized in Table 2. The space group for both crystals was P2₁2₁2₁ with two molecules in the asymmetric units forming noncovalent homodimers (Fig. 2, A and B). The interface surface area of wtsFae1A and wtsFae1B was 1675 and 1667 Å², respectively (Fig. 2, A and B). According to PISA analyses (24), these observed interactions are strong enough to be of biological relevance; ΔG was estimated to be −19.9 and −22.1 kcal mol⁻¹. Furthermore, both enzymes eluted from a gel-filtration column at a volume consistent with a dimer in solution (Fig. S3), which suggests that the enzymes are dimers in their active form.

The root-mean-square deviation (r.m.s.d.) for the Ca atomic coordinates of wtsFae1A and wtsFae1B, chains A and B, were 0.2 (290 atom pairs) and 0.1 Å (276 atom pairs), respectively. Overall, the electron density is well-defined in both enzymes (Fig. 2, C and D), with the exception of residues 198–203 in wtsFae1A chain B and 223–237 and 300–310 in wtsFae1B chain A and B, where such density is missing (Fig. S4, A–F), which suggests these residues constitute a flexible loop. Furthermore, electron density for the first 22 residues of wtsFae1B is missing.

The structures have a similar tertiary architecture composed of two domains. In both enzymes, the N-terminal CBM48 (residues: wtsFae1A(1–97) and wtsFae1B(1–113)) and the C-terminal CE1 domain (residues: wtsFae1A(115–343) and wtsFae1B(131–386)) are connected by a 17-residue linker. The two CBMs display the typical β-sandwich fold of CBMs (16) consisting of 10 β-strands (Fig. 2, A and B). The two CE1 catalytic domains display the α/β hydrolase fold typical of CE1 domains (25) consisting of eight β-strands forming a central β-sheet flanked by six α-helices (Fig. 2, A and B).

The CE1 and CBM48 domains appear to interact extensively with one another, with an estimated interfacial surface area of 336.4 Å² for wtsFae1A, and for wtsFae1B, it is significantly lower (257.5 Å²). Nevertheless, for both wtsFae1A and wtsFae1B, the data indicate that multiple hydrogen bonds contribute to fixing the position of CBM48 relative to the CE1 domain (Fig. 3, A and B; Table 3). Interestingly, three residues (wtsFae1A: Asp¹⁰⁶, Lys¹⁰⁹, and Val¹¹¹; wtsFae1B: Asp¹¹⁹, Lys¹²⁴, and Val¹²⁷) that participate in hydrogen bonds in the linker connecting the CE1 and CBM48 domains are structurally conserved (Fig. 3A). However, these residues are not conserved among wtsFae1A and wtsFae1B homologs (Fig. 3B). Additionally, several hydrogen bonds formed by peptide backbone atoms are present in the structurally-conserved helical area of the loop connecting the CE1 and CBM48 domains (Fig. 3, A and B; Table 3). Moreover, potential hydrogen bonds, with one in wtsFae1A (between Asn⁹⁰ and Gly¹³⁸) and two in wtsFae1B (between Arg¹⁰² and Asp¹⁷⁵ and between Arg¹⁰² and Asp¹⁷⁷), diametrically opposed the linker connecting the two domains may also be involved in forming the rigid CE1 and CBM48 integral units (Fig. 3, A and B; Table 3). None of these residues are conserved among the CE1–CBM48 homologs (Fig. 3B). Differential scanning calorimetry (DSC) of wtsFae1B showed a single unfolding event with Tₘ at ~61 °C (Fig. S4), which was also found previously for wtsFae1A (22), which corroborates the view that the CE1 and CBM48 domains form an integrated unit.

Relation to other structure-determined CE1 feruloyl esterases

Structural analyses of the Protein Data Bank (26) using the DALI server (27) revealed that the closest structural homolog for the CE1 domains is AmFae1A from the rumen fungus Ade lophyrene mucronatus (PDB code 5CXX) (10). The Ca r.m.s.d. between AmFae1A and wtsFae1A (see Fig. 4A for structural alignment) (166 atom pairs) and wtsFae1B (see Fig. 5A structural alignment) (152 atom pairs) were 1.3 and 1.0 Å, respectively.

Figure 1. wtsFae1B substrate interactions. A, adsorption analysis of insoluble wheat arabinoxylan binding to wtsFae1B. B, surface plasmon resonance analysis of xylohexaose binding to wtsFae1B.
**Arabinoxylan-binding CBM48**

### Table 1
Specific activities (milliunits mg\(^{-1}\)) of CE1 feruloyl esterases wtsFae1A and wtsFae1B on different types of substrates

| Enzyme            | 5-O-trans-Feruloyl-α-L-Araf | WAX-I     | GH10-pretreated WAX-I |
|-------------------|-----------------------------|-----------|----------------------|
| wtsFae1A          | 0.9 ± 0.1                   | 0.011 ± 0.001 | 0.028 ± 0.003 |
| wtsFae1A ∆CBM48   | 3.6 ± 0.1                   | ND\(^a\) | 0.181 ± 0.012 |
| wtsFae1B          | 0.5 ± 0.1                   | 0.011 ± 0.001 | 0.011 ± 0.001 |
| wtsFae1B ∆CBM48   | 1.7 ± 0.1                   | ND        | 0.042 ± 0.003 |

\(^a\) ND is not detected.

### Table 2
Data collection and refinement statistics

| Enzyme | wtsFae1A | wtsFae1B |
|--------|----------|----------|
| PDB code | 6RZO | 6RZN |

| Data collection | wtsFae1A | wtsFae1B |
|-----------------|----------|----------|
| Wavelength (Å)  | 0.87 Å   | 0.97 Å   |
| Resolution (Å)  | 49.6–1.63 (1.69–1.63) | 54.8–1.91 (1.98–1.91) |
| Space group     | P2\(_1\)P2\(_1\)P2\(_1\) | P2\(_1\)P2\(_1\)P2\(_1\) |
| Unit cell a, b, c (Å) | 66.6, 99.8, 114.4 | 65.7, 99.9, 133.3 |
| Total no. of reflections | 607,653 (33,386) | 324,451 (33,441) |
| No. of unique reflections | 90,065 (6422) | 68,057 (6769) |
| R\(_{merge}\)    | 0.16 (0.96) | 0.18 (0.78) |
| R\(_{sym}\)      | 0.17 (1.06) | 0.20 (0.88) |
| CC\(_{1/2}\)     | 0.064 (0.44) | 0.088 (0.38) |
| I/σ(I)          | 7.09 (1.20) | 6.73 (2.48) |
| Completeness (%) | 94.38 (68.18) | 98.49 (99.62) |
| Redundancy      | 6.7 (5.2) | 4.8 (4.9) |

| Refinement | wtsFae1A | wtsFae1B |
|------------|----------|----------|
| Reflections used in refinement | 89,974 (6422) | 68,007 (6769) |
| Reflections used for R\(_{free}\) | 4438 (305) | 3487 (351) |
| R\(_{work}\) | 0.17 (0.29) | 0.18 (0.23) |
| R\(_{free}\) | 0.19 (0.32) | 0.20 (0.25) |
| CC (work) | 0.96 (0.67) | 0.95 (0.84) |
| CC (free) | 0.96 (0.68) | 0.93 (0.80) |
| No. of refined non-hydrogen atoms | 6359 | 6308 |
| Protein | 5483 | 5366 |
| Solvent | 876 | 942 |
| B-Factors | | |
| Protein | 23.00 | 21.69 |
| Solvent | 34.80 | 31.13 |
| Wilson B-factor | 18.30 | 19.82 |
| r.m.s.d. | | |
| Bond lengths (Å) | 0.010 | 0.004 |
| Bond lengths (°) | 1.22 | 1.02 |
| Ramachandran plot | | |
| Favored | 97.78 | 98.35 |
| Allowed | 2.22 | 1.65 |
| Outliers | 0.00 | 0.00 |
| Rotamer outliers | 0.69 | 1.41 |
| No. of TLS groups | 1 | 0 |

Inspection of the structures shows that the core α/β-hydrolase fold and position of the catalytic triad is structurally-conserved in wtsFae1A and wtsFae1B (Fig. 2, A and B). However, whereas the serine (catalytic general base) (Ser\(^{242}\) for wtsFae1A and Ser\(^{277}\) for wtsFae1B) and histidine (general acid-base catalyst) (His\(^{325}\) for wtsFae1A and His\(^{368}\) for wtsFae1B) are conserved, the general acid is Glu\(^{296}\) for wtsFae1A and Asp\(^{339}\) for wtsFae1B (Fig. 2, A and B; see Fig. S2 for multiple alignment). Both Glu and Asp have been commonly observed as the key acid catalysis amino acids in feruloyl esterases (8). However, a number of loop regions potentially involved in substrate binding differ significantly among the feruloyl esterases (Fig. 4, A–D).

**Ferulic acid–binding pocket flexibility**

The electron density is missing in two loops surrounding the active sites of both enzymes, indicating that these regions are flexible. Fortunately, in one molecule (chain A of wtsFae1A), all loops were well-defined, which allowed us to perform a molecular dynamics simulation to obtain a better understanding of the dynamics of these loops. A 200-ns molecular dynamics simulation of wtsFae1A chain A showed that the core structures of both the CE1 and CBM48 domains are rigid and do not undergo noticeable movements (Fig. 5). However, the plot of the root mean square fluctuation as a function of the Cα shows that one region (residue 198–207) is apparently involved in concerted fluctuations (Fig. 5) that could result in the formation of a lid on the otherwise open active site in the apo structure (Fig. 2A). Thus, this flexible loop could act as a hinge promoting substrate binding and thus catalysis. Additionally, the Cα atoms of the residues around residue 300 close to the active site undergo fluctuations (Fig. 5), suggesting that a re-arrangement of the regions surrounding the active site could take place upon substrate binding. A multiple alignment of structure-determined CE1 feruloyl esterases showed that the residues constituting the flexible loops of

---

17342 J. Biol. Chem. (2019) 294(46) 17339–17353
XynZ and wtsFae1A are not conserved at the sequence level (Fig. 6). Furthermore, this analysis showed that presumably the flexible loop of wtsFae1B with its 36 residues differs significantly from that of the other CE1 FAEs (Fig. 6).

Concerted movements of the regions surrounding the active site have also been suggested for several other CE1 feruloyl esterases when FA is not present, including the closest structural homolog AmFae1A (10), the closest sequence homologs XynZ (11), and BiFae1A and EstE from the rumen bacteria Bacteroidetes intestinalis (12) and Butyrivibrio proteoclasticus (13), respectively. However, the topology of the FA-binding pocket lid varies significantly between the CE1 FAEs: in AmFae1A, a α-helix not present in wtsFae1A forms a lid on top of the FA-binding pocket (Fig. 4B); in Est1E, α-helices and loops not present in wtsFae1A form a flexible hinge on top of the FA-binding pocket (Fig. 4C). The FA-binding pocket of XynZ is most similar to that of wtsFae1A, and if the flexibility of the regions surrounding the FA-binding pocket is taken into account, the FA-binding pockets of these two enzymes are practically identical (Fig. 4D). XynZ and wtsFae1A are structurally very similar (r.m.s.d. 0.7 Å for 153 Ca atom pairs).

**Role of CBM48 in relation to substrate distortion to assist catalysis**

To determine the role of the CBM48, we docked [β-D-Xylp-(1→4)-[5-O-feruloyl-α-L-Araf-(1→2)]-β-D-Xylp-(1→4)-β-D-
Xylp} (XA\textsubscript{5f2X}) (Fig. 7A) to wtsFae1A chain A (Fig. 7B). As mentioned above, the electron density is unfortunately missing in certain regions near the active site of wtsFae1B chains A and B and wtsFae1A chain B, which precludes meaningful docking experiments for these. When compared with XynZ in complex with FA (PDB code 1JT2), a slight shift in the position of the FA on XA\textsubscript{5f2X} could be observed (Fig. 7B). However, the displacement of the flexible loop and \( \alpha \)-helix 6 (Fig. 4D) and the fact that FA on XA\textsubscript{5f2X} is constrained by its attachment to the AXOS may cause the observed differences. Furthermore, the stacking interaction with Trp\textsubscript{157} (Fig. 7B) and Ile\textsubscript{90} in XynZ may also contribute to the altered orientation of FA.

Figure 3. Hydrogen bonds keeping the CE1 and CBM48 domains in the correct orientation. A, structurally-conserved residues involved in hydrogen bonds forming the rigid wtsFae1A (cyan) and wtsFae1B (green) structures (hydrogen bonds are shown as yellow dashed lines with their length given in Å, and the residues involved are shown as sticks). B, multiple alignment of CE1–CBM48 homologs (see Fig. S6 for complete alignment). The asterisks indicate the residues involved in hydrogen bonds keeping the CE1 and CBM48 domains in the correct relative orientation (wtsFae1B above and wtsFae1A below the multiple alignment). The protein sequences are identified by their GenBank\textsuperscript{TM} accession numbers. The multiple alignment is visualized using ESPript 3.0 (57).
The docking of XA\(^{527}\) to wtsFae1A chain A further suggests that the xylan main chain is accommodated in the cleft formed at the interface between the CE1 and CBM48 domains (Fig. 7B). Interestingly, hydrogen bonds are only formed between the Xyl\(p\) moieties and residues on the CBM48 (Fig. 7B), which are not conserved among the CE1–CBM48 homologs (Fig. S6). Furthermore, no stacking interactions with Xyl\(p\) moieties were observed, but Ara\(f\) forms a hydrogen bond with the conserved general acid-base catalyst His\(^{325}\) (Fig. 7B).

### Table 3

| Donor | Acceptor | Distance (Å) |
|-------|----------|--------------|
|       |          | A  | B  |
| Asn\(^{90}\)–NH | Gly\(^{158}\)–O | 2.0 | 2.0 |
| Glu\(^{103}\)–NH | Ala\(^{100}\)–O | 2.3 | 2.4 |
| Phe\(^{102}\)–NH | Asp\(^{103}\)–O | 2.5 | 2.4 |
| Tyr\(^{106}\)–NH | Glu\(^{103}\)–O | 3.4 | 2.2 |
| Lys\(^{109}\)–H\(_2\) | Asp\(^{104}\)–O\(_2\) | 2.2 | 2.0 |
| Val\(^{111}\)–NH | Lys\(^{109}\)–O | 2.3 | 2.2 |
| Arg\(^{395}\)–H\(_2\) | Lys\(^{109}\)–O | 2.2 | 2.3 |

### Figure 4

Comparison of wtsFae1A chain A to other CE1 feruloyl esterases. Superimpositions of wtsFae1A chain A (cyan) are shown. A, AmFae1A from \(A.\) mucronatus (PDB code 5CXX) (orange); B, Bifae1A from \(B.\) intestinalis (PDB code 5VOL) (gray); C, Ets1E from \(B.\) proteoclasticus (PDB code 2WTM) (yellow); and D, XynX from \(H.\) thermocellum (PDB code 1JJF) (purple). The active site of wtsFae1A is indicated as are the flexible loops and clamps that form a lid on the active-site pocket.

### Figure 5

Flexibility of Co in wtsFae1A. Root mean square fluctuation as a function of Co in wtsFae1A chain A during a 200-ns molecular dynamics simulation is shown.

Surprisingly, the active sites and binding clefts of wtsFae1A and wtsFae1B display very distinct topology and properties (Fig. 8, A–E). The binding cleft and active-site pocket of wtsFae1A is negatively charged, whereas that of wtsFae1B is neutral (Fig. 8, A and B). Interestingly, in wtsFae1B no aromatic residues are present, which is similar to Trp\(^{157}\) in wtsFae1A, which can form a stacking interaction, but Phe\(^{340}\) in wtsFae1B may be able to form a stacking interaction with FA (Fig. 8C). We suggest that the role of both these aromatic residues may be to distort and pull the FA moiety toward them and thus destabilize the 5-O-linkage to Araf.

Similarly to what has been reported for other CE1 feruloyl esterases, the FA-binding pockets of both wtsFae1A and wtsFae1B are hydrophobic, whereas the binding clefts accommodating the xylan main chain are not (Fig. 8, D and E). This is surprising because many carbohydrates form stacking interactions with aromatic residues at binding sites in carbohydrate-active enzymes.

### Relation of feruloyl esterase CBMs to starch binding CBMs

Structural analyses of the Protein Data Bank (26) using the DALI server (27) revealed that the closest structural homolog to the wtsFae1A and wtsFae1B CBM domains is CBM48 (residues 253–338) appended to the starch phosphatase Starch Excess\(4\)
The Cα r.m.s.d. between AtSEX4 CBM48 and chain A of wtsFae1A and wtsFae1B CBM48s was 3.2 Å (43 atom pairs) and 6.2 Å (48 atom pairs), respectively (see Fig. 9, A–C, for structural alignment). The Cα r.m.s.d. between chain A wtsFae1A and wtsFae1B CBM48s was 1.3 Å (52 atom pairs) (see Fig. 9, A–C, for structural alignment).

A superimposition of AtSEX4 in complex with maltohexaose and of wtsFae1A with XA5f2X showed that the catalytic domains of the two enzymes are fixed to the CBM48 domains at
different angles; however, the substrates interact with the CBM48 domains in the same region (Fig. 9, B–D). The superimposition also revealed that AtSEX4 is unable to accommodate the XA5f2X (Fig. 9C), whereas wtsFae1A can accommodate the maltohexaose (Fig. 9D). Maltohexaose interacts with Trp²⁷⁸, Lys³⁰⁷, Trp³¹⁴, His³³⁰, and Asn³³² on the AtSEX4 CBM48 (28) that all been shown to be important for maintaining both binding and activity (28, 29). None of these residues are structurally-conserved in wtsFae1A (Fig. S7). It is particularly interesting that Trp²⁷⁸ and Trp³¹⁴, which form a conserved aromatic platform at the binding site on CBM48s appended to starch-acting enzymes (21), are missing in wtsFae1A and wtsFae1B (Fig. 8 for multiple alignment). The multiple alignment of the closely-related starch-binding CBM20, CBM48, and CBM69 (21, 30) and homologs of the CBMs of wtsFae1B and wtsFae1A (see Fig. 8 for multiple alignment) was used to construct a maximum-likelihood phylogenetic tree that showed that the CBMs appended to wtsFae1A and wtsFae1B belong to the CBM48 family (Fig. 10). Interestingly, the CBM48s appended to wtsFae1A and wtsFae1B and the homologs cluster with three CBM48s appended to starch-acting enzymes that also lack the CBM48s appended to starch-acting enzymes (Fig. 10; Fig. S8). Hence, it is questionable whether these so-called starch-binding CBM48 domains actually bind onto starch. Based on this analysis, we propose that the CBM48 modules identified in this study may constitute a separate CBM48 sub-family not belonging to the group of CBM48 starch-binding modules.

Discussion

The recent major efforts in metagenomics and genomics continue to reveal genes encoding novel carbohydrate-active enzymes (31). Enzymes that often differ from previously characterized enzymes in domain organization or differ sufficiently at the sequence level suggest that these enzymes are capable of acting on substrates not previously seen within specific enzyme families (31). Two studies have revealed CE1s with appended CBMs annotated as CBM48s (19, 20), and similar enzymes were present in GenBank™ (32); however, their CBMs, as for wtsFae1A, were not annotated as CBM48s (22). The presence of a CBM suggests that these feruloyl esterases are capable of acting on complex, polymeric substrates and potentially also on insoluble substrates, a capability that would hold vast potential in conversion of recalcitrant lignocellulosic biomass fractions (10).

Unfortunately, the commercially available polymeric feruloyl esterase substrates are limited to WAX-I in which only a small fraction of the Araf is ferulated (33). Despite this, it is clear that wtsFae1A and wtsFae1B depend on their cognate CBM48 to act on WAX-I (Table 1). Hence, both wtsFae1A and wtsFae1B could potentially act on more complex substrates and thus be tools for unlocking the potential of the recalcitrant lignocellulosic biomass fractions.

The crystal structures of wtsFae1A and wtsFae1B reveal CE1 domains that, although similar to known CE1 domains at the fold level, have a significantly different active-site topology, in particular with respect to the loops forming a lid on top of the active sites (Figs. 4, A–D, and 6). These differences have been suggested to impact the enzymes’ ability to accommodate mono-, di-, tri-, and tetra-FAs (10, 34, 35) that exist in planta (36). The FA-binding pocket of wtsFae1A resembles that of XynZ from H. thermocellum (Fig. 4D), which is exposed and thus can accommodate both mono- and di-FAs (10). Surprisingly, no di-FA was released from WAX-I by either wtsFae1A or wtsFae1B. However, wtsFae1B also has a significantly longer loop, presumably forming a lid on top of the active site (Fig. 6), which may impact the substrate specificity.

The catalytic domains of wtsFae1A and wtsFae1B and CBM48 form a rigid integral unit, and the main chain of the
substrates binds at a cleft formed at the interface of the two domains (Figs. 2, A and B, and 7B). Our DSC, crystallographic, and molecular dynamics data suggest that the wtsFae1A and wtsFae1B domains form a rigid structure (Fig. 5), which is kept in place by conserved hydrogen bonds (Fig. 3, A–C). Our docking data suggest that the CE1 and CBM48 domains act in consort, with the CBM48 responsible for binding the Xylp moieties, whereas the CE1 integrates the AraF-FA in the catalytic site for hydrolysis (Fig. 7B). This is supported by the complete loss of activity for wtsFae1AΔCBM48 and wtsFae1BΔCBM48 toward WAX-I (Table 1). This result was a bit surprising because two CE1 feruloyl esterases from B. intestinalis that lack a CBM have been reported to release FA from WAX-I (12). However, the lack of interactions between the CE1 domain and the XA54X main chain in the docking experiment suggests that the CE1 domain recognizes only the FA and potentially also the AraF moiety. This is in line with what has been observed in other feruloyl esterase crystal structures where FA only is observed despite being linked to an AXOS (11, 37). One can speculate that the helical structure of xylan (38) somehow prevents these feruloyl esterases from accommodating the FA in the active-site pocket. The weak binding observed for the shorter xylooligosaccharides and AXOSs compared with xylohexaose implies that a minimum of six Xylp moieties are required for productive binding. Surprisingly, an increase in activity toward soluble ferulated AXOSs and 5-O-trans-feruloyl-α-L-Araf was observed for wtsFae1AΔCBM48 and wtsFae1BΔCBM48 compared with the full-length enzymes. The reason for this remains elusive; however, perhaps the removal of the CBM grants better access to the active-site pocket for small, soluble substrates.

Overall, wtsFae1A and wtsFae1B are structurally very similar (Fig. 2, A and B), but the properties and the topology of the active-site pocket and the binding cleft are very different (Fig. 8, A–E). This may suggest that they act on different substrates in Nature. The negatively charged binding cleft of wtsFae1A (Fig. 8, A and B) could be a hint that negatively charged substrates like the glucuronic acid stretches of glucuronoarabinoxylan cannot be accommodated, although this is not the case for wtsFae1B. Furthermore, both the differences in length of the flexible loops that potentially form a lid on the active-site pocket (Fig. 6) and the different positions of the aromatic residues that potentially form a stacking interaction with the FA (wtsFae1A Trp157; wtsFae1B Phe340) (Fig. 8C) also suggest differences in specificity. BiFae1A from the B. intestinalis (12) and EstE from B. proteoclasticus (13) both have an aromatic residue, namely Trp67 and Phe33, respectively, that structurally resembles that of wtsFae1A. As already mentioned, XynZ from H. thermocellum (11) and also AmFae1A from A. mucronatus (10) do not have aromatic residues that form stacking interactions with the FA. The flexibility of the active site seems to be common for feruloyl esterases (10, 12, 13). The wtsFae1A and wtsFae1B homodimers probably do not affect the activity of the enzymes because this would require that the helical xylan structure be turned 180°.

The phylogenetic tree of the three starch-binding CBM families, CBM20, CBM48 and CBM69, and the CBMs homologous to the CBMs appended to wtsFae1A and wtsFae1B unambiguously support the classification of these CBMs as CBM48s (Fig. 10). Our binding data thus suggest that family CBM48 not only contains starch binding CBMs but also
Arabinoxylan-binding CBM48

The phylogenetic tree also shows that the xylan-binding CBM48s cluster with the CBM48s appended to starch-acting enzymes that lack the starch-binding site (Fig. 10). Unfortunately, no binding data for these starch-acting enzymes have been published. However, we did demonstrate weak binding to maltotetraose for wtsFae1B. Hence, the starch-active and -related enzymes lacking an aromatic platform at the starch-binding site might bind to short maltooligosaccharides, which would not be expected for the maltooligosyltrehalose trehalohydrolase, glycogen-debranching enzyme, and branching enzyme function. The relation to CBM48 for the wtsFae1A and wtsFae1B CBMs is further strengthened by the structural similarity to the CBM48 appended to AtSEX4 (Fig. 9, A–D).

This study provides the first structural characterization of CE1 enzymes appended to a CBM and adds to our understanding of the functional role of CBMs. The CBMs were identified as belonging to CBM family 48. Yet, our findings also hint at the need for a renewed view of CBM48 functionality and classification. In addition, crystal structures enabled us to perform a molecular dynamics simulation showing that the two domains form a rigid structure and a docking study showing that the two domains act in consort. Both wtsFae1A and wtsFae1B displayed activity toward WAX-I, which was lost when CBM48 was not present. Altogether, the combined results of this study present feruloyl esterases as a potential means to advance our exploitation of plant biomass.

Experimental procedures

Genes, cloning, expression, and purification

An ORF encoding a CE1 with a CBM48 appended (wtsFae1B) (GenBankTM access no. BK010854.1) was identified in a metagenomic study of the anaerobic digester Randers (Whole Genome Shotgun accession no. MTKZ00000000) on contig 6388, bp 3157–4389 (19). wtsFae1B was analyzed for the presence of a signal peptide by SignalP 4.1 (39). Disulfide bonds were predicted by DiANNA 1.1 (40). Molecular mass and pl values were predicted by Compute pI (41), and pl was 8.82, and molecular mass was 42.7 kDa. The theoretical molar extinction coefficient, calculated using ProtParam (http://web.expasy.org/protparam)5, was 59820 M⁻¹ cm⁻¹. wtsFae1A is part of a
chimeric enzyme (GenBank™ accession no. BK010417.1) for which the above-mentioned characteristics were previously described (22). Domains were mapped with dbCan (23).

The codon-optimized mature genes (Table S1) for Escherichia coli encoding wtsFae1A and wtsFae1B truncations (wtsFae1AΔCBM48 (residue 99–343), CMB48 from wtsFae1A (residue 1–124), wtsFae1BΔCBM48 (residue 114–386), and CMB48 from wtsFae1B (residue 1–117), were purchased and cloned into pET-28a using the restriction sites NcoI and XhoI.

CBM48 from wtsFae1B (residue 1–117), were purchased and cloned into pET-28a using the restriction sites NcoI and XhoI. The resulting plasmids were transformed into E. coli strain BL21 (DE3) pLys (Novagen). Transformsants were grown at 37 °C in LB medium supplemented with 50 μg ml⁻¹ kanamycin until cell growth reached OD₆₀₀ 0.8 and cooled on ice for 30 min, and expression was induced by addition of 1 mM (final concentration) isopropyl-thio-β-D-galactoside to the LB medium. The cultures were grown at 15 °C for a further 16–18 h. The cells were then pelleted (2000 × g; 20 min; 4 °C), stored at −20 °C, thawed, and resuspended in 1:10 volume of 50 mM Tris, 500 mM NaCl, 20 mM imidazole, pH 7.5, and lysed by sonication. The lysate was centrifuged (20,000 × g; 20 min; 4 °C), and the supernatant was filtered (0.45-μm Durapore membrane filters; Millipore), applied to a 1- or 5-ml HisTrap SP HP column (GE Healthcare) equilibrated with 50 mM Tris, 500 mM NaCl, 20 mM imidazole, pH 7.5, and eluted (2 ml min⁻¹) by a linear 20–500 mM imidazole gradient (30 column volumes). Fractions containing the target enzymes were pooled, concentrated (Vivaspin, 10 kDa), and applied to a HiLoad 16/60 Superdex G75 column (GE Healthcare) equilibrated with 10 mM NaOAc, 150 mM NaCl, pH 7 (0.5 ml min⁻¹). Fractions containing pure target enzymes were concentrated (Vivaspin, 10 kDa) and stored at 4 °C. CBM48A and CBM48B were not subjected to the gel-filtration step. The purity was checked on 12% SDS-polyacrylamide gels and the concentration of the protein samples was measured by A₂₈₀ using the theoretically obtained molar extinction coefficients.

**Analytical gel filtration**

The analytical gel filtration was performed on a HiLoad 16/60 Superdex G200 column equilibrated with ferritin (440 kDa), aldose (158 kDa) conalbumin (75 kDa), and ovalbumin (44 kDa) (all GE Healthcare) dissolved in 10 mM NaOAc, 150 mM NaCl, pH 6, at a 0.5 ml/min flow with 10 mM NaOAc, 150 mM NaCl, pH 6, as running buffer.

**Hydrolysis of 5-O-trans-feruloyl-α-L-Araf**

5-O-trans-Feruloyl-α-L-Araf was synthesized from l-(+)-arabinose and trans-FA as described by Holck et al. (22). The specific activity for wtsFae1B was determined in duplicate by mixing 20 μl of wtsFae1A (20 μM), wtsFae1B (122 μM), wtsFae1AΔCBM48 (1.67 μM), and wtsFae1BΔCBM48 (5.63 μM) with 80 μl of 2 mM 5-O-trans-feruloyl-α-L-Araf dissolved in 10 mM NaOAc, pH 6, and incubated for 9 min at 40 °C after a 2-min preincubation at 40 °C. Quantification of the released FA was determined by LC electrospray ionization MS (LC–ESI–MS) as described by Holck et al. (22). One unit was defined as the amount of enzyme releasing 1 μmol min⁻¹ FA.

**Hydrolysis of wheat arabinoxylan**

Specific activity for the enzymes and truncations was determined in duplicate by mixing 50 μl of wtsFae1A (14 μM), wtsFae1B (20 μM), wtsFae1AΔCBM48 (1.67 μM), and wtsFae1BΔCBM48 (5.63 μM) with 50 μl of 10 mg ml⁻¹ WAX-I dissolved in 10 mM NaOAc, pH 6, and incubated for 1 h at 40 °C after a 2-min preincubation at 40 °C. The released FA was quantified as above. Additionally, 10 mg ml⁻¹ WAX-I dissolved in 10 mM NaOAc, pH 6, was incubated for 24 h at 20 °C with GH10 (50 μM), which yields soluble ferulated AXOS (22). The reaction mixture was centrifuged (10,000 × g, 5 min, 4 °C) after 20 h, and the reactions were stopped by passing the reaction mixture over a membrane (Vivaspin, 3 kDa, Sartorius, Gottingen, Germany). Subsequently, 50 μl of the resulting ferulated AXOS were mixed with 50 μl of wtsFae1A (20 μM), wtsFae1B (122 μM), wtsFae1AΔCBM48 (1.67 μM), and wtsFae1BΔCBM48 (5.63 μM) and incubated for 1 h at 40 °C after a 2-min preincubation at 40 °C. The released FA was quantified as above.

**Affinity for wheat arabinoxylan**

WAX-I (Megazyme) and wheat starch (Sigma) was washed three times in 50 mM NaOAc, pH 6. The ability of wtsFae1A and wtsFae1B to bind wheat starch was determined by mixing 5 μl of enzymes (wtsFae1A 4 μM, wtsFae1B 1.35 μM) and 1 mg ml⁻¹ 95 μl of WAX-I and wheat starch in 50 mM NaOAc, pH 6, and 0.005% (w/v) BSA, incubated in triplicate at 4 °C for 30 min, and centrifuged (20,000 × g, 4 °C, 10 min). Protein concentrations in the supernatants were determined by A₂₈₀ readings. Similarly, the Kᵯ,app was obtained for wtsFae1B and WAX-I by mixing 5 μl wtsFae1B (1.35 μM) and 95 μl WAX-I (0.078–10 mg ml⁻¹). The Kᵯ,app was determined by fitting the adsorption isotherm to the amount of bound enzyme, B = Bmax [S]/Kᵯ,app [S], where B is the amount of bound enzyme, [S] the WAX-I concentration, and Bmax the maximum adsorption capacity.

**Surface plasmon resonance**

wtsFae1B was diluted in 10 mM sodium acetate, pH 4.5, to 0.1 mg ml⁻¹, prior to amine coupling to 2872.6 resonance units (RU) onto a CMS chip (GE Healthcare) for SPR analysis (BIACore T100, GE Healthcare). Sensorgrams (RU versus time) were recorded at 25 °C in running buffer (10 mM sodium acetate, 150 mM NaCl, 0.005% Tween 20, pH 6) at a flow rate of 30 μl min⁻¹ with 120 s contact time followed by 120 s dissociation and were baseline-corrected by subtracting data from a parallel flow cell without enzyme. The ability of wtsFae1B to bind 1 mM β-cyclo-dextrin, maltotetraose, maltohexaose, xylobiose, -triose, -tetraose, -pentaose, -hexose, A’X, and A’XX was determined. The Kᵯ value was calculated (Equation 1) from RU values obtained with 4–250 μM xylohexaose by steady-state affinity fitting (BIAcore T100 evaluation 2.0.3; GE Healthcare):

$$R = \frac{R_{\text{max}} [\text{ligand}]}{[\text{ligand}]} + K_D$$

(Eq. 1)

[ligand] is the oligosaccharide concentration, and R_{max} is the maximum binding capacity in RU.
Differential scanning calorimetry

DSC was used to analyze the conformational stability of wtsFae1B using a Nano DSC calorimeter (TA Instruments). Protein samples (1 mg ml\(^{-1}\)) were dialyzed against 200 volumes of 10 mM sodium acetate, pH 6, for 24 h, degassed, and loaded into sample cells and scanned (25–90 °C, 1 °C min\(^{-1}\)) with the dialysis buffer in the reference cell. Baseline scans, collected with buffer in both reference and sample cells, were subtracted from sample scans, and the Universal Analysis software (TA Instruments) with a DSC add-on was used to model the reference cell and baseline-corrected thermograms using a two-state scaled model to determine \(T_m\) (unfolding temperature, defined as the temperature of maximum apparent heat capacity) and the calorimetric heat of unfolding \(\Delta H_{\text{cal}}\).

Crystallization, data collection, and data processing

wtsFae1A and wtsFae1B were crystallized in 48-well MRC Maxi plates (Jena Bioscience) by mixing 2 μl of protein with 2 μl of reservoir. A protein concentration of 80 mg ml\(^{-1}\) in 10 mM sodium acetate was used for wtsFae1B. wtsFae1A had spontaneously formed crystalline precipitate while stored at 4 °C, and the saturated supernatant was diluted four times in 10 mM sodium acetate, pH 6, to give a concentration of 0.8 mg ml\(^{-1}\) before crystallization trials were performed with the resulting sample. PACT (Jena Bioscience) and MCSG-1 (Anatrace) commercial screens were used, and crystals were identified in several conditions for both proteins. Crystals were cryo-cooled in liquid nitrogen and tested at the ESRF beamlines ID30-A3 and ID23-2. Data were collected for wtsFae1B on a crystal from condition F9 in the MCSG-1 screen (50 mM ammonium sulfate, 50 mM Bis-Tris/HCl, pH 6.5, 30% (v/v) pentaerythritol ethoxylate (15/4_EO/OH)) where the drop containing the crystals had first been supplemented with PEG400 for cryoprotection. This dataset was processed to 1.9 Å resolution with autoPROC (see Table 2 for statistics) (42). Several crystals of wtsFae1A diffracted, but the best dataset was collected on a crystal from condition PACT F1 (20 w/v PEG 3,350, 100 mM Bistris propane, pH 6.5, 200 mM sodium fluoride). The useful resolution range for this dataset was underestimated, and data were collected on the square detector to a resolution of 1.8 Å on the edge of the detector, but during processing with EDNAproc (43), useful data to 1.63Å could be obtained. Although the completeness in the highest-resolution shell was only 68%, the \(CC_{1/2}\) was above 0.4, which indicated a strong signal.

Structure solution, model building, and refinement

To solve the structure of wtsFae1A and wtsFae1B, a model was prepared from the closest homolog from the Protein Data Bank (www.pdb.org), which was the CE1 feruloyl esterase of the XynZ from \(H.\) thermocellum (PDB code 1JT2) identified through PD–BLAST (44) with a coverage around 65% and identity around 35% to the two targets. Sculptor (45) from the PHENIX package (46) was used to generate a search model, and PHASER (47) was used to perform molecular replacement, searching for two molecules in the asymmetric unit in both cases, and revealed the space group to be \(P2_12_12\), for wtsFae1B. Even though the search model only covered the CE1 domains, TFZ scores of 21.3 and 23.7 for wtsFae1B and wtsFae1A, respectively, were obtained. Initially, the structures were built with PHENIX.autobuild (48) and then refined with PHENIX.refine (49) and manual model rebuilding in Coot (50) to a final \(R_{work}/R_{free}\) of 0.17/0.20 and 0.18/0.21 for wtsFae1A and wtsFae1B, respectively.

Structural alignments, electrostatic plots, and hydrophobicity plots

Structural alignments were obtained using PyMOL 2.2 (Schrödinger, LLC, New York; also used for rendering structural models), and the overall r.m.s.d. for \(Co\) ranged from 0.152 to 0.188 Å. Electrostatic maps were obtained with the APBS plugin in PyMol 2.2 using default settings. Hydrophobicity plots were obtained using the color_h.py script in PyMOL 2.2 (Schrödinger, LLC, New York) and colored according to the Eisenberg hydrophobicity scale (51). IsPISA version 2.1.1 was used to calculate the interfacial surface areas and strength for the dimers and the CE1 and CBM48 domains (24).

Molecular dynamics

The molecular dynamics simulation was performed on wtsFae1A chain A in the Yasara Structure (18.4.24) with the built-in molecular dynamic simulation macro “md_run.mrc” using AMBER14 as a force field. The simulation cell was allowed to include 10 Å surrounding the protein and filled with water at a density of 0.997 g/ml. Initial energy minimization was carried out under relaxed constraints using steepest descent minimization. Simulations were performed in water at a constant pressure with temperature at 298 K. The systems were neutralized at pH 6 by counterions using 0.9% w/v NaCl. The simulation was run at constant pressure and 298 K for 100 ns. Data were collected every 250 ps. Snapshots were then analyzed using the built-in “md_analyzers.mrc” macro for the r.m.s.d. of the \(Co\) of the α-helices and β-strands and the ligand heavy atoms as distances of the catalytic residues from the scissile bond of the ligand.

Docking

wtsFae1A chain A was subjected to energy minimization in vacuo prior to docking. The ligand for the docking studies, \(\text{XA}^{312}_{\text{X}}\left[\beta\text{-D-Xylp-(1→4)}\right]-5\text{-O-feruloyl-\alpha\text{-1-Araf-(1→2)}\left[\beta\text{-D-Xylp-(1→4)}\beta\text{-D-Xylp}\right]}\) (Fig. 7A), was built using Yasara Structure (18.4.24) and subjected to steepest descent energy minimization prior to the docking studies. The docking studies were performed with wtsFae1A chain A using the Yasara Structure (18.4.24) with the built-in macro “dock_run.mrc” using the Autodock VINA (52) for 25 runs in the AMBER03 force field under default parameters. The simulation cell was set to 15 Å centered around Ser242 and then manually adjusted to include the expected binding cleft. The results were compared with XynZ from \(H.\) thermocellum (11) in complex with FA (PDB code 5JT2) to identify the most probable positioning of the ligand, which was subsequently subjected to energy minimization using Yasara Structure (18.4.24).

Multiple alignments and phylogenetic analysis

A multiple alignment was built with the CBM20 and CBM48 protein sequences previously investigated by Janeček and co-
workers (21), CBM69 was from CAZy based on personal communication with Stefan Janeček and CBMs homologous to the CBM48s appended to wtsFae1A and wtsFae1B. The latter were obtained from a BlastP against the NR database (44) using the wtsFae1A and wtsFae1B CBM48 sequences as queries. The top 100 hits from each search were pooled, and a multiple alignment was constructed using MAFFT (53), which was subsequently manually trimmed using the wtsFae1A and wtsFae1B CBM48 sequences for guidance. To reduce redundancy and the number of CBM48 sequences, the CBM48 sequences were clustered at 50% similarity by using CD-Hit (54), which resulted in an additional 15 CBM48 sequences. The multiple alignment was used for building the LG maximum likelihood phylogenetic tree using RaxML-HPC BlackBox (version 8.2.10) (55) at the CIPRES Science Gateway version 3.3 (56) with 1000 bootstrap replications.

The multiple alignments of the CE1 domain alone (manually selected from the multiple alignment of the full-length enzymes) and of the full-length CE1–CBM48 enzymes were based on the above-obtained sequences and obtained using MAFFT (53).

Author contributions—J. H., F. F., M. S. M., J. B., and C. W. investigation; J. H., F. F., B. S., A. S. M., and C. W. visualization; J. H., F. F., M. S. M., J. B., K. B. R. M. K., D. H. W., B. S., A. S. M., and C. W. methodology; J. H. and C. W. writing—original draft; F. F., M. S. M., K. B. R. M. K., L. L., D. H. W., B. S., A. S. M., and C. W. writing—review and editing; L. L., D. H. W., and B. S. funding acquisition; C. W. conceptualization; C. W. supervision; C. W. project administration.

Acknowledgments—We thank the beamline staff at ID30-A3 and ID23-2 for assistance, and the ESRF for synchrotron beamtime. Stefan Janeček is gratefully acknowledged for invaluable input about carbohydrate-binding modules CBM20, -48, and -69. The Independent Research Fund Denmark, Natural Sciences, supported the purchase of the Biacore T100 instrument through Grant 272-06-0050.

References
1. Naidu, D. S., Hlangothi, S. P., and John, M. J. (2018) Bio-based products from xylan: a review. Carbohydr. Polym. 179, 28–41 CrossRef Medline
2. Garg, S. (2016) Xylanase: applications in biofuel production. Curr. Metabolomics 4, 23–37 CrossRef
3. González-García, S., Morales, P. C., and Gullo, B. (2018) Estimating the environmental impacts of a brewery waste-based biorefinery: bio-ethanol and xylooligosaccharides joint production case study. Ind. Crops Prod. 123, 331–340 CrossRef
4. Scheller, H. V., and Ulvskov, P. (2010) Hemicelluloses. Annu. Rev. Plant Biol. 61, 263–289 CrossRef Medline
5. Naran, R., Chen, G., and Carpita, N. C. (2008) Novel rhamnogalacturonan I and arabinoxylan polysaccharides of flax seed mucilage. Plant Physiol. 148, 132–141 CrossRef Medline
6. Darvill, J. E., McNeil, M., Darvill, A. G., and Albersheim, P. (1980) Structure of plant cell walls. Plant Physiol. 66, 1135–1139 CrossRef Medline
7. Biely, P., Singh, S., and Puchart, V. (2016) Towards enzymatic breakdown of complex plant xylan structures: state of the art. Biotechnol. Adv. 34, 1260–1274 CrossRef Medline
8. Oliveira, D. M., Mota, T. R., Oliva, B., Segato, F., Marchiosi, R., Ferrarese-Filho, O., Faulds, C. B., and Dos Santos, W. D. (2019) Feruloyl esterases: biocatalysts to overcome biomass recalcitrance and for the production of bioactive compounds. Bioresour. Technol. 278, 408–423 CrossRef Medline
9. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrisat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, D490–D495 CrossRef Medline
10. Grüninger, R. J., Cote, C., McAllister, T. A., and Abbott, D. W. (2016) Contributions of a unique β-clamp to substrate recognition illuminates the molecular basis of exolysis in ferulic acid esterases. Biochem. J. 473, 839–849 CrossRef Medline
11. Schubot, F. D., Kataeva, I. A., Blum, D. L., Shah, A. K., Ljungdahl, L. G., Rose, J. P., and Wang, B. C. (2001) Structural basis for the substrate specificity of the feruloyl esterase domain of the cellulosomal xylanase Z from Clostridium thermocellum. Biochemistry 40, 12524–12532 CrossRef Medline
12. Wefers, D., Cavalcante, J. I. V., Schendel, R. R., Deveryshetty, J., Wang, K., Wawrzak, Z., Mackie, R. I., Koropatkin, N. M., and Cann, I. (2017) Biochemical and structural analyses of two cryptic esterases in Bacteroides intestinalis and their synergistic activities with cognate xylanases. J. Mol. Biol. 429, 2509–2527 CrossRef Medline
13. Goldstone, D. C., Villas-Bôas, S. G., Till, M., Kelly, W. J., Attwood, G. T., and Arcus, V. L. (2010) Structural and functional characterization of a promiscuous feruloyl esterase (Est1 E) from the rumen bacterium Butyrivibrio proteoclasticus. Proteins 78, 1457–1469 CrossRef Medline
14. Guillén, D., Sánchez, S., and Rodríguez-Sanaja, B. (2010) Carbohydrate-binding domains: multiplicity of biological roles. Appl. Microbiol. Biotechnol. 85, 1241–1249 CrossRef Medline
15. Várnai, A., Mákél, M. R., Djadji, D. T., Rahikainen, J., Hatakka, A., and Viikari, L. (2014) Carbohydrate-binding modules of fungal cellulases: occurrence in nature, function, and relevance in industrial biomass conversion. Adv. Appl. Microbiol. 86, 103–165 CrossRef Medline
16. Gilbert, H. J., Knox, J. P., and Boraston, A. B. (2013) Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. Curr. Opin. Struct. Biol. 23, 669–677 CrossRef Medline
17. Fujimoto, Z., Jackson, A., Michikawa, M., Maehara, T., Momma, M., Henrisat, B., Gilbert, H. J., and Kaneko, S. (2013) The structure of a Streptomyces avermitilis α-L-rhamnosidase reveals a novel carbohydrate-binding module CBM67 within the six-domain arrangement. J. Biol. Chem. 288, 12376–12385 CrossRef Medline
18. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. 382, 769–781 CrossRef Medline
19. Wållken, E., Busk, P. K., Pilgaard, B., Zhang, W.-J., Nielsen, K. L., Nielsen, P. H., and Lange, L. (2017) Diversity of microbial carbohydrate active enzymes in Danish anaerobic digesters fed with wastewater treatment sludge. Biotechnol. Biofuels 10, 158 CrossRef Medline
20. Wong, M. T., Wang, W., Couturier, M., Razeg, F. M., Lombard, V., Lapebie, P., Edwards, E. A., Terrapon, N., Henriisat, B., and Master, E. R. (2017) Comparative metagenomics of cellulose- and poplar hydrolysis degrade microcosms from gut microflora of the Canadian beaver (Castor canadensis) and North American moose (Alces americanus) after long-term enrichment. Front. Microbiol. 8, 2504 CrossRef Medline
21. Janeček S., Svensson, B., and MacGregor, E. A. (2011) Structural and evolutionary aspects of two families of non-catalytic domains present in starch and glycogen binding proteins from microbes, plants and animals. Enzyme Microb. Technol. 49, 429–440 CrossRef Medline
22. Holck, J., Djadji, D. T., Brask, I., Pilgaard, B., Krogh, K. B. R. M., Meyer, A. S., Lange, L., and Wållken, C. (2019) Novel xylanolytic triple domain enzyme targeted at feruloylated arabinoxylan degradation. Enzyme Microb. Technol. 129, 109353 CrossRef Medline
23. Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012) DBCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 40, W445–W451 CrossRef Medline
24. Krissinel, E. (2015) Stock-based detection of protein oligomeric states in jPSiA. Nucleic Acids Res. 43, W314–W319 CrossRef Medline
25. Nakamura, A. M., Nascimento, A. S., and Polikarpov, I. (2017) Structural diversity of carbohydrate esterases. Biotechnol. Res. Innov. 1, 35–51 CrossRef
26. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 CrossRef Medline

27. Holm, L., and Laakso, L. M. (2016) Dali server update. *Nucleic Acids Res.* **44**, W351–W355 CrossRef Medline

28. Meekins, D. A., Raththagala, M., Husodo, S., White, C. J., Guo, H.-F., Ko¨tting, O., Vander Kooi, C. W., and Gentry, M. S. (2014) Phosphoglucone-bound structure of starch phosphatase starch Excess4 reveals the mechanism for C6 specificity. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7272–7277 CrossRef Medline

29. Cockburn, D., Wilkens, C., Dilokpimol, A., Nakai, H., Lewinski, A., Abou Hackem, M., and Svensson, B. (2016) Using carbohydrate interaction assays to reveal novel binding sites in carbohydrate active enzymes. *PLoS ONE* **11**, e0160112 CrossRef Medline

30. Peng, H., Zheng, Y., Chen, M., Wang, Y., Xiao, Y., and Gao, Y. (2014) A starch-binding domain identified in α-amylase (AmyP) represents a new family of carbohydrate-binding modules that contribute to enzymatic hydrolysis of soluble starch. *FEBS Lett.* **588**, 1161–1167 CrossRef Medline

31. Helbert, W., Poulet, L., Drouillard, S., Mathieu, S., Loiodice, M., Couturier, M., Lombard, V., Terrapon, N., Turchetto, J., Vincentelli, R., and Henrisat, B. (2019) Discovery of novel carbohydrate-active enzymes through the rational exploration of the protein sequence space. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 10184–10185 CrossRef Medline

32. Clark, K., Karchs-Mizrahi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2016) GenBank. *Nucleic Acids Res.* **44**, D67–D72 CrossRef Medline

33. Sorensen, H. R., Pedersen, S., and Meyer, A. S. (2007) Characterization of solubilized arabinoxyl-o-oligosaccharides by MALDI-TOF MS analysis to unravel and direct enzyme catalyzed hydrolysis of insoluble wheat arabinoxylan. *Enzyme Microb. Technol.* **41**, 103–110 CrossRef Medline

34. Hermoso, J. A., Sanz-Aparicio, J., Molina, R., Juge, N., Gonzalez, R., and Faulds, C. B. (2004) The crystal structure of feruloyl esterase A from *Aspergillus niger* suggests evolutionary functional convergence in feruloyl esterase family. *J. Mol. Biol.* **338**, 495–506 CrossRef Medline

35. Bartolome, B., Faulds, C. B., Kropp, P. A., Waldron, K., Gilbert, H. J., Hazelwood, G., and Williamson, G. (1997) Aspergillus niger ferulic esterase (ferulic acid esterase III) and a recombinant *Pseudomonas fluorescens* subsp. *cellulosa* esterase (Xy1D) release a 5–5′-ferulic dehydrodimer. *Appl. Environ. Microbiol.* **63**, 208–212 Medline

36. Bunzel, M. (2010) Chemistry and occurrence of hydroxycinnamyl oligomers. *Phytochem. Rev.* **9**, 47–64 CrossRef Medline

37. Faulds, C. B., Molina, R., Gonzalez, R., Husband, F., Juge, N., Sanz-Aparicio, J., and Hermoso, J. A. (2005) Probing the determinants of substrate specificity of a feruloyl esterase, AnFaeA, from *Aspergillus niger* and a recombinant *Pseudomonas fluorescens* subsp. *cellulosa* esterase (Xy1D) release a 5–5′-ferulic dehydrodimer. *Appl. Environ. Microbiol.* **63**, 208–212 Medline

38. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 CrossRef Medline

39. Ferré, F., and Clote, P. (2006) DIANNA 1.1: an extension of the DIANNA web server for ternary cysteine classification. *Nucleic Acids Res.* **34**, W182–W185 CrossRef Medline

40. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, A., Wilkins, M., Appel, R. D., and Bairoch, A. (2005) in The Proteomics Protocols Handbook (Walker, J. M., ed) pp. 571–607, Humana Press Inc., Totowa, NJ

41. Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., and Brucigov, G. (2011) Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 293–302 CrossRef Medline

42. Monaco, S., Gordon, E., Bowler, M. W., Delagenière, S., Guijarro, M., Spruce, D., Svensson, O., McSweeney, S. M., McCarthy, A. A., Leonard, G., and Nano, M. H. (2013) Automatic processing of macromolecular crystallography X-ray diffraction data at the ESRF. *J. Appl. Crystallogr.* **46**, 804–810 CrossRef Medline

43. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oefner, R., Read, R. J., Richardson, D. C., et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 CrossRef Medline

44. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 CrossRef Medline

45. Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 61–69 CrossRef Medline

46. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 CrossRef Medline

47. Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* **179**, 125–142 CrossRef Medline

48. Trotta, O., and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 CrossRef Medline

49. Katoh, K., Misawa, K., Kuma, K. M., and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 CrossRef Medline

50. Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012) CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 CrossRef Medline

51. Stamatakis, A., Hoover, P., and Rougemont, J. (2008) A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* **57**, 758–771 CrossRef Medline

52. Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. in *Gateway Computing Environments Workshop*, IEEE, New Orleans, LA CrossRef Medline

53. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDScript server. *Nucleic Acids Res.* **42**, W320–W324 CrossRef Medline

54. Letunic, I., and Bork, P. (2007) Interactive tree of life (itol): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 CrossRef Medline