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

Holck, Jesper; Fredslund, Folmer; Møller, Marie Sofie; Brask, Jesper; Krogh, Kristian B. R. M.; Lange, Lene; Weiner, Ditte Hededam; Svensson, Birte; Meyer, Anne S.; Wilkens, Casper

Published in:
Journal of Biological Chemistry

Link to article, DOI:
10.1074/jbc.RA119.009523

Publication date:
2019

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):
A carbohydrate-binding family 48 module enables feruloyl esterase action on polymeric arabinoxylan

Jesper Holck¹, Folmer Fredslund², Marie S. Møller¹, Jesper Brask³, Kristian B. R. M. Krogh³, Lene Lange⁴, Ditte H. Welner², Birte Svensson¹, Anne S. Meyer¹ and Casper Wilkens¹,*

From the ¹Section for Protein Chemistry and Enzyme Technology, Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads, Building 224, DK-2800 Kgs. Lyngby, Denmark; ²Enzyme Engineering and Structural Biology, Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, Building 220, DK-2800 Kgs. Lyngby, Denmark; ³Novozymes A/S, Biologiens Vej 2, DK-2800 Kgs. Lyngby, Denmark; ⁴LLa-Bioeconomy, Research & Advisory, Karensgade 5, DK-2500 Valby, Denmark

Running title: Arabinoxylan-binding CBM48

*To whom correspondence should be addressed: Casper Wilkens, Section for Protein Chemistry and Enzyme Technology, Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads, Building 224, DK-2800 Kgs. Lyngby, Denmark; E-mail: cwil@dtu.dk; Tel. +4545252946

Keywords: arabinoxylan; carbohydrate binding module; carbohydrate esterase family 1; crystal structure; enzyme catalysis; enzyme mechanism; ferulic acid esterase; molecular docking; molecular dynamics; structure-function

ABSTRACT

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, wtsFae1A 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-D-1,4-xylanase did enable CE1 domain–mediated FA release from arabinoxylan. 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 which 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, while double substitutions are also present in the endosperm. In contrast, dicots predominantly have single α-L-1,2-Araf substitutions (4), though 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 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) (www.cazy.org) (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 non-catalytic, 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 CBM48 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 the present 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 thus an abbreviation of 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 peptides residues, respectively. The molecular weights 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-PAGE gel corresponded to these molar masses (Fig. S1). Previous attempts to purify recombinant truncated wtsFae1A that lacks CBM48 (wtsFae1AΔCBM48) failed due to precipitation (22). This precipitation was avoided, however, by maintaining wtsFae1AΔCBM48 in the elution buffer from the HisTag 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 thermodenitrificans. When compared to 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 $0.204 \pm 0.017$ mg ml$^{-1}$ (22), while wtsFae1B displayed a $K_{d,app}$ of $1.3 \pm 0.16$ 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, xylohexaose, \([\alpha-L-\text{Araf}(1\rightarrow3)-\beta-D-\text{Xylp}(1\rightarrow4)-D-\text{Xylp}]\) (A\(^3\)X), \([\alpha-L-\text{Araf}(1\rightarrow2)-\beta-D-\text{Xylp}(1\rightarrow4)-\beta-D-\text{Xylp}(1\rightarrow4)-D-\beta-\text{Xylp}]\) (A\(^2\)XX), maltotetraose, maltohexaose and \(\beta\)-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 analogs \(\beta\)-cyclodextrin and maltotetraose and maltohexaose. However, a meaningful \(K_d\) value was obtained for xylohexaose (Fig. 1B) and found to be 14.8 \(\mu\)M.

Previously, wtsFae1A was shown to displayed minimal activity against WAX-I (22). But by using higher enzyme concentration, 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 wtsFae1A\(\Delta\)CBM48 and wtsFae1B\(\Delta\)CBM48 (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, wtsFae1A\(\Delta\)CBM48 and wtsFae1B\(\Delta\)CBM48 displayed about six-fold and four-fold higher specific activity, respectively, than the corresponding full-length enzymes towards soluble ferulated AXOS obtained by pretreating WAX-I with a glycoside hydrolase family 10 (GH10) \(\beta\)-d-1,4 xylanase (EC 3.2.1.8) (Table 1). The same trend was observed for the enzymes and truncated forms assayed on 5-O-trans-feruloyl-\(\alpha\)-L-Araf (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 \(P2_12_12_1\) with two molecules in the asymmetric units forming non-covalent homodimers (Fig. 2A–B). The interfacial surface area of wtsFae1A and wtsFae1B was 1675 and 1667 \(\AA^2\), respectively (Fig. 2A–B). According to PISA analyses (24) these observed interactions are strong enough to be of biological relevance; \(\Delta G\) 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 (rmsd) for the \(C_\alpha\) 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. 2C–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. S4A–F), which suggests these residues constitute a flexible loop. Further, electron density for the first 22 residues of wtsFae1B is missing.

The structures have a similar tertiary architecture comprised of two domains. In both enzymes the N-terminal CBM48 (residues: wtsFae1A 1–97; wtsFae1B 1–113) and the C-terminal CE1 domain (residues: wtsFae1A 115–343; wtsFae1B 131–386) are connected by a 17 residue linker. The two CBMs display the typical β-sandwich fold of CBMs (16) consisting of ten β-strands (Fig. 2A–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. 2A–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, while 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. 3A–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. 3A–B; Table 3). Moreover, potential hydrogen bonds, with one in wtsFae1A (between Asn90 and Gly158) and two in wtsFae1B (between Arg102 and Asp175; and 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. 3A–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 \( T_m \) at ~61°C (Fig. S5), 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 *A. mucronatus* (PDB ID 5CXX) (10). The Cα rmsd between AmFae1A and wtsFae1A (see Fig. 4A for structural alignment) (166 atom pairs) and wtsFae1B (see Fig. S6 structural alignment) (152 atom pairs) were 1.3 and 1.0 Å, respectively.

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. 2A–B). However, while the serine (catalytic general base) (Ser242 for wtsFae1A and Ser272 for wtsFae1B) and histidine (general acid-base catalyst) (His325 for wtsFae1A and His368 for wtsFae1B) are conserved, the general acid is Glu296 for wtsFae1A and Asp339 for wtsFae1B (Fig. 2A–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. 4A–D).

**The 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 (rmsf) as a function of the Cα shows that one region (residue 198–207) in particular 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 XynZ and wtsFae1A are not conserved at the sequence level (Fig. 6). Further, this analysis showed that presumably 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 *Bacteriodetes 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 β-clamp closes the FA binding pocket and α-helix 2 is slightly extended compared to wtsFae1A (Fig. 4A); in BiFae1A, an 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 are taken into account, the FA binding pockets of these two enzymes are practically identical (Fig. 4D). XynZ and wtsFae1A are structurally very similar (rmsd 0.7 Å for 153 Cα atom pairs).

The role of CBM48 in relation to substrate distortion to assist catalysis

To determine the role of the CBM48 we docked \([\beta-D-Xylp-(1\rightarrow4)-[5-O-feruloyl-\alpha-L-Araf-(1\rightarrow2)]-\beta-D-Xylp-(1\rightarrow4)-\beta-D-Xylp\] (XA52X) (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 to XynZ in complex with FA (PDB ID 1JT2) a slight shift in the position of the FA on XA52X could be observed (Fig. 7B). However, the displacement of the flexible loop and α-helix 6 (Fig. 4D) and the fact that FA on XA52X is constrained by its attachment to the AXOS may cause the observed differences. Further, the stacking interaction with Trp157 (Fig. 7B) and Ile90 in XynZ may also contribute to the altered orientation of FA.

The docking of XA52X 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 Xylp moieties and residues on the CBM48 (Fig. 7B), which are not conserved among the CE1-CBM48 homologs (Fig. S7). Further, no stacking interactions with Xylp moieties were observed, but Araf forms a hydrogen bond with the conserved general acid-base catalyst His325 (Fig. 7B).

Surprisingly, the active sites and binding clefts of wtsFae1A and wtsFae1B display very distinct topology and properties (Fig. 8A–E). The binding cleft and active site pocket of
wtsFae1A is negatively charged, whereas that of wtsFae1B is neutral (Fig. 8A–B). Interestingly, in wtsFae1B no aromatic residues are present, which is similar to Trp\textsuperscript{157} in wtsFae1A which can form a stacking interaction, but Phe\textsuperscript{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 towards them and thus destabilize the 5-\textit{O} linkage to Ara\textsubscript{f}.

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

\textit{The 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 starch phosphatase Starch Excess4 from \textit{Arabidopsis thaliana} (AtSEX4) (PDB ID 4PYH) (28). The \textit{C}_\text{a} rmsd between AtSEX4 CBM48 and chain A of wtsFae1A and wtsFae1B CBM48s was 3.2 (43 atom pairs) and 6.2 \text{ Å} (48 atom pairs), respectively, (see Fig. 9A–C for structural alignment). The \textit{C}_\text{a} rmsd between chain A of wtsFae1A and wtsFae1B CBM48s was 1.3 \text{ Å} (52 atom pairs) (see Fig. 9A–C for structural alignment).

A superimposition of AtSEX4 in complex with maltohexaose and of wtsFae1A with XA\textsuperscript{52}X 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. 9B–D). The superimposition also revealed that AtSEX4 is unable to accommodate the XA\textsuperscript{52}X (Fig. 9C), while wtsFae1A can accommodate the maltohexaose (Fig. 9D). Maltohexaose interacts with W278, K307 W314, H330 and N332 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. S8).

It is particularly interesting that W278 and W314, which form a conserved aromatic platform at the binding site on CBM48s appended to starch acting enzymes (21), is missing in wtsFae1A and wtsFae1B (Fig. S9 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. S9 for multiple alignment)
was used to construct a maximum likelihood phylogenetic tree which 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 aromatic platform at the binding site (Fig. 10; Fig. S9). Hence it is questionable if these so-called starch binding CBM48 domains actually bind onto starch. Based on this analysis, we propose that the CBM48 modules identified in the present work 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 to 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 which 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 Ara<sub>f</sub> 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 (Fig. 4A–D; Fig. 6). These differences have been suggested to impact on the enzymes ability to accommodate mono-<sub>-</sub>, di-<sub>-</sub>, tri-<sub>-</sub> 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.

He catalytic domains of wtsFae1A and wtsFae1B and CBM48 form a rigid integral unit, and main-chain of the substrates binds at a cleft formed at the interface of the two domains (Fig. 2A–B; Fig. 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. 3A–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 Ara\textsubscript{f}-FA in the catalytic site for hydrolysis (Fig. 7B). This is supported by the complete loss of activity for wtsFae1A\textDelta CBM48 and wtsFae1B\textDelta CBM48 towards WAX-I (Table 1). This result was a bit surprising because, two CE1 feruloyl esterases from \textit{Bacteroides intestinalis} and which lack a CBM have been reported to release FA from WAX-I (12). However, the lack of interactions between the CE1 domain and the XA\textsubscript{52}X main chain in the docking experiment suggests that the CE1 domain recognizes only the FA and potentially also the Ara\textsubscript{f} 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 to xylohexaose implies that a minimum of six Xylp moieties are required for productive binding. Surprisingly, an increase in activity towards soluble ferulated AXOS and 5-\textit{O}-\textit{trans}-feruloyl-\textalpha-L-Araf was observed for wtsFae1A\textDelta CBM48 and wtsFae1B\textDelta CBM48 compared to the full-length enzymes. The reason for this remains elusive, however, maybe 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. 2A–B), but the properties and the topology of the active site pocket and the binding cleft are very different (Fig. 8A–E). This may suggest that they act on different substrates in Nature. The negatively charged binding cleft of wtsFae1A (Fig. 8A–B) could be a hint that negatively charged substrates like the glucuronic acid stretches of glucuronoarabinoxylan cannot be accommodated, whereas this is not the case for wtsFae1B. Further, 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 Trp\textsuperscript{157}; wtsFae1B Phe\textsuperscript{340}) (Fig. 8C) also suggest differences in specificity. BiFae1A from \textit{B. intestinalis} (12) and EstE from \textit{B. proteoclasticus} (13) both have an aromatic residue, namely Trp\textsuperscript{67} and Phe\textsuperscript{33}, respectively, that structurally resembles that of wtsFae1A. As already mentioned, XynZ from \textit{H. thermocellum} (11) and also AmFae1A from \textit{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 since this would require that the helical xylan structure 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 xylan-binding ones. 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 \textit{AtSEX4} (Fig. 9A–D).

The present 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 towards WAX-I, which was lost when CBM48 was not present. Altogether, the combined results of this study presents feruloyl esterases as a potential means to advance our exploitation of plant biomass.

**Experimental procedures**
An open reading frame encoding a CE1 with a CBM48 appended (wtsFae1B) (GenBank acc. nr. BK010854.1) was identified in a metagenomic study of the anaerobic digester Randers (Whole Genome Shotgun acc. nr. 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). MW and pI were predicted by Compute pI (41) and pI was 8.82 and molecular mass was 42.7 kDa. The theoretical molar extinction coefficient, calculated using ProtParam (http://web.expasy.org/protparam), was 59820 M$^{-1}$ cm$^{-1}$. wtsFae1A is part of a chimeric enzyme (GenBank acc. nr. 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) - CBM48 from wtsFae1A (residue 1–124), wtsFae1BΔCBM48 (residue 114–386) and CBM48 from wtsFae1B (residue 1–117) - were purchased and cloned into pET-28a using the restriction sites Nco1 and Xho1 (GenScript) in frame with the HisTag. The resulting plasmids were transformed into *E. coli* strain BL21 (DE3) pLys (Novagen). Transformants were grown at 37 °C in LB medium supplemented with 50 μg ml$^{-1}$ kanamycin until cell growth reached OD$_{600}$ 0.8, 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 (2000g; 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 (20000g; 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 CV). 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 6 (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 gelfiltration step. The purity was checked on 12 % SDS-PAGE gels and the concentration of the protein samples was measured by A$_{280}$ 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 wtsFae1A (20 µM), wtsFae1B (122 µM), wtsFae1AΔCBM48 (1.67 µM) and wtsFae1BΔCBM48 (5.63 µM) with 80 µl 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 2 min preincubation at 40°C. Quantification of the released FA was determined by liquid chromatography electrospray ionization mass spectrometry (LC-ESI-MS) as described by Holck *et al.* (22). One U was defined as the amount of enzyme releasing 1 µmol min⁻¹ FA.

**Hydrolysis of wheat arabinoxylan**

Specific activity for the enzymes and truncations were determined in duplicate by mixing 50 µl wtsFae1A (14 µM), wtsFae1B (20 µM), wtsFae1AΔCBM48 (1.67 µM) and wtsFae1BΔCBM48 (5.63 µM) with 50 µl 10 mg ml⁻¹ WAX-I dissolved in 10 mM NaOAc pH 6 and incubated for 1 h at 40°C after 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 (10000g, 5 min, 4°C) after 20 h and the reactions were stopped by passing the reaction mixture over a membrane (Viaspin, 3 kDa, Sartorius, Goettingen, Germany). Subsequently, 50 µl of the resulting ferulated AXOS were mixed with 50 µl wtsFae1A (20 µM), wtsFae1B (122 µM), wtsFae1AΔCBM48 (1.67 µM) and wtsFae1BΔCBM48 (5.63 µM) and incubated for 1 h at 40°C after 2 min preincubation at 40°C. The released FA was quantified as above.

**Affinity for wheat arabinoxylan**

WAX-I (Megazyme) and wheat starch (Sigma-Aldrich) 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 enzymes (wtsFae1A 4 µM; wtsFae1B 1.35 µM) and 1 mg ml⁻¹ 95 µl 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 (20000g, 4°C, 10 min). Protein concentrations in the supernatants were determined from A₂₈₀ readings. Similarly, the \( K_{d,\text{app}} \) 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 \( K_{d,\text{app}} \) was determined by fitting the adsorption isotherm to the amount of bound enzyme, \( B = \frac{B_{\text{max}} [S]}{K_D [S]} \), where \( B \) is the amount of bound enzyme, [S] the WAX-I concentration, and \( B_{\text{max}} \) 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 RU (resonance units) onto a CM5 chip (GE Healthcare) for SPR analysis (BIACore T100, GE Healthcare). Sensorgrams (RU vs. 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 baseline-corrected by subtracting data from a parallel flow cell without enzyme. The ability of wtsFae1B to bind 1 mM β-cyclodextrin, maltotetraose, maltohexaose, xylobiose, -triose, -tetraose, -pentaose, -hexose, \( A^3X \) and \( A^2XX \) was determined. \( K_d \) was calculated (eq. 1) from RU values obtained with 4–250 µM xylohexaose by steady-state affinity fitting (BIACore T100 evaluation 2.0.3; GE Healthcare):

\[
R = \frac{R_{\text{max}} [\text{Ligand}]}{[\text{Ligand}] + K_D} \tag{1}
\]

[\text{Ligand}] is the oligosaccharide concentration and \( R_{\text{max}} \) 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 protein with 2 µl 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 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 (42) (see Table 2 for statistics). Several crystals of wtsFae1A diffracted, but the best dataset was collected on a crystal from condition PACT F1 (20 w/v polyethylene glycol 3,350, 100 mM Bis-Tris 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 1.8Å 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 CC₁/₂ 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 ID 1JT2) identified through PDB-BLAST (44) with a coverage around 65% and identity of 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 *P*₂₁₂₁ 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*ₜₒₜₜ/ *R*ₜₒₜ 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, NY, USA; also used for rendering structural models) and the overall rmsd for Ca 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 (http://www.protein.osaka-u.ac.jp/rcsfp/supracryst/suzuki/jpxtal/Katsutani/en/hydrophobicity.php) in PyMol 2.2 (Schrödinger, LLC, New York, NY, USA) and colored according to the Eisenberg hydrophobicity scale (51). jsPISA v.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 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 298K for 100 ns. Data were collected every 250 ps. Snapshots were then analyzed using the built-in “md_analyzeres.mcr” macro for the rmsd of the Ca 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, XA^42X [β-D-Xylp-(1→4)-[5-O-feruloyl-α-L-Araf-(1→2)]-β-D-Xylp-(1→4)-β-D-Xylp] (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 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 to XynZ from H. thermocellum (11) in complex with FA (PDB ID 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 coworkers (21), CBM69 were from CAZy based on personal communication with Stefan Janecek and CBMs homologous to the CBM48s appended to wtsFae1A and wtsFae1B. The later 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 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 (v. 8.2.10) (55) at the CIPRES Science Gateway V 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).

**ACKNOWLEDGEMENTS**

We thank the beamline staff at ID30-A3 and ID23-2 for assistance, and ESRF for synchrotron beamtime. Štefan Janeček is gratefully acknowledged for invaluable input about carbohydrate binding modules CBM20, 48 and 69. Funding for this project was provided by the Villum Foundation (grant number VKR022796) to LL and base funding from the Technical University of Denmark. FF and DHW were supported by The Novo Nordisk Foundation through grant number NNF10CC1016517 to the Novo Nordisk Foundation Center for Biosustainability at DTU; BS and MSM were supported by The Independent Research Fund Denmark | Natural Sciences (grant number 6108-00476A) which also supported the purchase of the Biacore T100 instrument (grant number 272-06-0050). The protein sequence for the wtsFae1B gene has been deposited in the NCBI Protein Database under NCBI Accession BK010854.1. The atomic coordinates for the crystal structures of wtsFae1A and wtsFae1B are available in the Protein Data Bank (https://www.wwpdb.org/) under PDB ID 6RZO and 6RZN, respectively.
CONFLICT OF INTERESTS

The authors declare that they have no conflicts of interest with the contents of this article. JB and KBRMK are employees at Novozymes A/S, a company that produces and sells enzymes and microbes for industrial uses.

REFERENCES


49. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W.,
Towards automated crystallographic structure refinement with phenix.refine. Acta


sequences with the hydrophobic moment plot. J Mol Biol. 179, 125–142

52. Trott, 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

fast Fourier transform. Nucleic Acids Res. 30, 3059–3066

next-generation sequencing data. Bioinformatics. 28, 3150–3152

RAxML Web Servers. Syst. Biol. 57, 758–771

10.1109/GCE.2010.5676129

ENDscript server. Nucleic Acids Res. 42, 320–324

58. Letunic, I., and Bork, P. (2016) Interactive Tree Of Life (iTOL): An online tool for
phylogenetic tree display and annotation. Bioinformatics. 23, 127–128

ABBREVIATIONS AND NOMENCLATURE
Araf, arabinofuranose; AX, Arabinoxylans; AXOS, arabinxylooligosaccharide; CAZy,
Carbohydrate-Active EnZyme database; CBM, carbohydrate binding module; CBM48,
carbohydrate binding module family 48; CE1, carbohydrate esterase family 1; di-FA, ferulate
dehydrodimers; FA, ferulic acid; GH10, glycoside hydrolase family 10; rmsd, root-mean-square
deviation; SPR, surface plasmon resonance; WAX-I, insoluble wheat arabinoxylan; Xylp,
xylopyranose; XA\(^{\alpha2}\)X, β-D-Xylp-(1→4)-[5-O-feruloyl-α-L-Araf-(1→3)]-β-D-Xylp-(1→4)-β-D-
Xylp
Table 1: Specific activities (mU mg\(^{-1}\)) of CE1 feruloyl esterase wtsFae1A and wtsFae1B on different types of substrates. Data are given as averages of duplicate measurements ± standard deviations. wtsFae1AΔCBM48 and wtsFae1BΔCBM48 are abbreviations for the two CE1 feruloyl esterases devoid of their CBM.

<table>
<thead>
<tr>
<th></th>
<th>5-O-trans-feruloyl-α-L-Araf</th>
<th>WAX-I</th>
<th>GH10 pretreated WAX-I</th>
</tr>
</thead>
<tbody>
<tr>
<td>wtsFae1A</td>
<td>0.9 ± 0.1</td>
<td>0.011 ± 0.001</td>
<td>0.028 ± 0.003</td>
</tr>
<tr>
<td>wtsFae1A ΔCBM48</td>
<td>3.6 ± 0.1</td>
<td>n.d.</td>
<td>0.181 ± 0.012</td>
</tr>
<tr>
<td>wtsFae1B</td>
<td>0.5 ± 0.1</td>
<td>0.011 ± 0.001</td>
<td>0.011 ± 0.001</td>
</tr>
<tr>
<td>wtsFae1B ΔCBM48</td>
<td>1.7 ± 0.1</td>
<td>n.d.</td>
<td>0.042 ± 0.003</td>
</tr>
</tbody>
</table>

n.d. = not detected
Table 2: Data collection and refinement statistics.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>wtsFae1A</th>
<th>wtsFae1B</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB ID</td>
<td>6RZO</td>
<td>6RZN</td>
</tr>
</tbody>
</table>

**Data collection**

<table>
<thead>
<tr>
<th></th>
<th>wtsFae1A</th>
<th>wtsFae1B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength</td>
<td>0.87 Å</td>
<td>0.97 Å</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>49.64–1.63 (1.69–1.63)</td>
<td>54.86–1.91 (1.98–1.91)</td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁2₁2₁</td>
<td>P2₁2₁2₁</td>
</tr>
<tr>
<td>Unit cell</td>
<td>a, b, c (Å)</td>
<td>66.6, 99.8, 114.4</td>
</tr>
<tr>
<td>Total no. of reflections</td>
<td>607653 (33386)</td>
<td>324451 (33441)</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>90065 (6422)</td>
<td>68057 (6769)</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.16 (0.96)</td>
<td>0.18 (0.78)</td>
</tr>
<tr>
<td>Rfree</td>
<td>0.17 (1.06)</td>
<td>0.20 (0.88)</td>
</tr>
<tr>
<td>Rred</td>
<td>0.064 (0.44)</td>
<td>0.088 (0.38)</td>
</tr>
<tr>
<td>CC₁/₂</td>
<td>0.99 (0.41)</td>
<td>0.98 (0.70)</td>
</tr>
<tr>
<td>I/σ (I)</td>
<td>7.09 (1.20)</td>
<td>6.73 (2.48)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>94.38 (68.18)</td>
<td>98.49 (99.62)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>6.7 (5.2)</td>
<td>4.8 (4.9)</td>
</tr>
</tbody>
</table>

**Refinement**

<table>
<thead>
<tr>
<th></th>
<th>wtsFae1A</th>
<th>wtsFae1B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflections used in refinement</td>
<td>89974 (6422)</td>
<td>68007 (6769)</td>
</tr>
<tr>
<td>Reflections used for Rfree</td>
<td>4438 (305)</td>
<td>3487 (351)</td>
</tr>
<tr>
<td>Rwork</td>
<td>0.17 (0.29)</td>
<td>0.18 (0.23)</td>
</tr>
<tr>
<td>Rfree</td>
<td>0.19 (0.32)</td>
<td>0.20 (0.25)</td>
</tr>
<tr>
<td>CC (work)</td>
<td>0.96 (0.67)</td>
<td>0.95 (0.84)</td>
</tr>
<tr>
<td>CC (free)</td>
<td>0.96 (0.68)</td>
<td>0.93 (0.80)</td>
</tr>
<tr>
<td>No. of refined non-hydrogen atoms</td>
<td>6359</td>
<td>6308</td>
</tr>
<tr>
<td>Protein</td>
<td>5483</td>
<td>5366</td>
</tr>
<tr>
<td>Solvent</td>
<td>876</td>
<td>942</td>
</tr>
<tr>
<td>B-factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>23.00</td>
<td>21.69</td>
</tr>
<tr>
<td>Solvent</td>
<td>34.80</td>
<td>31.13</td>
</tr>
<tr>
<td>Wilson B-factor</td>
<td>18.30</td>
<td>19.82</td>
</tr>
<tr>
<td>r.m.s.d.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.010</td>
<td>0.004</td>
</tr>
<tr>
<td>Bond lengths (°)</td>
<td>1.22</td>
<td>1.02</td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored</td>
<td>97.78</td>
<td>98.35</td>
</tr>
<tr>
<td>Allowed</td>
<td>2.22</td>
<td>1.65</td>
</tr>
<tr>
<td>Outliers</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Rotamer outliers</td>
<td>0.69</td>
<td>1.41</td>
</tr>
<tr>
<td>No. of TLS groups</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 3. List of direct hydrogen bonds ensuring the conformational rigidity of CE1 and CBM48 domains.

<table>
<thead>
<tr>
<th>wtsFae1A Donor</th>
<th>Acceptor</th>
<th>Distance (Å) A</th>
<th>Distance (Å) B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asn-90-NH</td>
<td>Gly-158-O</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Glu-103-NH</td>
<td>Ala-100-O</td>
<td>2.3</td>
<td>2.4</td>
</tr>
<tr>
<td>Phe-105-NH</td>
<td>Asp-102-O</td>
<td>2.5</td>
<td>2.4</td>
</tr>
<tr>
<td>Tyr-106-NH</td>
<td>Glu-103-O</td>
<td>2.4</td>
<td>2.2</td>
</tr>
<tr>
<td>Lys-109-H₂C₂</td>
<td>Asp-104-Oε₂</td>
<td>2.2</td>
<td>2.0</td>
</tr>
<tr>
<td>Val-111-NH</td>
<td>Lys-109-O</td>
<td>2.3</td>
<td>2.2</td>
</tr>
<tr>
<td>Arg-295-Hη₂₁</td>
<td>Asp-63-Oδ₁</td>
<td>2.2</td>
<td>2.3</td>
</tr>
<tr>
<td>Arg-295-Hε</td>
<td>Asp-63-Oδ₂</td>
<td>2.2</td>
<td>2.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>wtsFae1B Donor</th>
<th>Acceptor</th>
<th>Distance (Å) A</th>
<th>Distance (Å) B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg-102-Hη₁₂</td>
<td>Asp-175-Oδ₂</td>
<td>2.2</td>
<td>2.3</td>
</tr>
<tr>
<td>Arg-102-Hη₂₂</td>
<td>Asp-175-Oδ₂</td>
<td>2.6</td>
<td>2.6</td>
</tr>
<tr>
<td>Arg-102-Hη₂₁</td>
<td>Asp-177-Oδ₂</td>
<td>2.4</td>
<td>2.4</td>
</tr>
<tr>
<td>Ser-118-NH</td>
<td>Gly-115-O</td>
<td>2.4</td>
<td>2.4</td>
</tr>
<tr>
<td>D-119-NH</td>
<td>Gly-115-O</td>
<td>2.7</td>
<td>2.8</td>
</tr>
<tr>
<td>D-119-NH</td>
<td>Pro-116-O</td>
<td>2.7</td>
<td>2.8</td>
</tr>
<tr>
<td>Tyr-121-NH</td>
<td>Ser-118-O</td>
<td>2.1</td>
<td>2.1</td>
</tr>
<tr>
<td>Lys-124-HG₃</td>
<td>Asp-119-Oδ₁</td>
<td>2.3</td>
<td>2.3</td>
</tr>
<tr>
<td>Lys-124-Nζ₃</td>
<td>Asp-119-Oδ₂</td>
<td>3.1</td>
<td>2.7</td>
</tr>
<tr>
<td>Val-127-NH</td>
<td>Lys-124-O</td>
<td>2.9</td>
<td>-</td>
</tr>
<tr>
<td>Arg-374-Hη₁₂</td>
<td>Glu-79-Oε₁</td>
<td>2.3</td>
<td>2.4</td>
</tr>
</tbody>
</table>
**Figure 1. wtsFae1B substrate interactions.** A) adsorption analysis of insoluble wheat arabinoxylan binding to wtsFae1B, and B) surface plasmon resonance analysis of xylohexaose binding to wtsFae1B.
Figure 2. The overall structures of wtsFae1A and wtsFae1B dimers and electron density at the linker regions. A) The two domain structure of wtsFae1A chain A (cyan) and wtsFae1A chain B (green cyan), and B) the two domain structure of wtsFae1B chain A (green) and wtsFae1B chain B (lime). The catalytic triads, Ser\textsuperscript{242}-His\textsuperscript{325}-Glu\textsuperscript{396} for wtsFae1A and Ser\textsuperscript{272}-His\textsuperscript{325}-Asp\textsuperscript{339} for wtsFae1B, are highlighted in red. Composite omit map contoured to 1.0σ in blue mesh with a cut-off at 1.6 Å for C) the linker region of wtsFae1A chain A and D) the linker region of wtsFae1B chain A. The composite omit map was calculated using Phenix with the dataset for PDB IDs 6RZO and 6RZN.
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), and B) multiple alignment of CE1-CBM48 homologs (see Fig. S7 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 accession number. The multiple alignment is visualized using ESPript 3.0 (57).
Figure 4. Comparison of wtsFae1A chain A to other CE1 feruloyl esterases. Superimpositions of wtsFae1A chain A (cyan) and A) AmFae1A from *A. mucronatus* (PDB ID 5CXX) (orange), B) BiFae1A from *B. intestinalis* (PDB ID 5VOL) (grey), C) Ets1E from *B. proteoclasticus* (PDB ID 2WTM) (yellow) and D) XynX from *H. thermocellum* (PDB ID 1JJF) (purple). The active site of wtsFae1A is indicated as are the flexible loops and clamps which form a lid on the active site pocket.
Figure 5. Flexibility of $C_\alpha$ in wtsFae1A. Root mean square fluctuation as a function of $C_\alpha$ in wtsFae1A chain A during a 200 ns molecular dynamics simulation.
Figure 6. Comparison of the loops forming the lid on the ferulic acid binding pocket. Multiple alignment showing the loops forming a lid on the ferulic acid binding pocket. The lid-forming loops/β-clamp is highlighted by green boxes. Asterix indicates the catalytic residues. The residue numbering refers to wtsFae1A. The sequences were obtained from PDB ID 1JFF (H. thermocellum), PDB ID 5CXX (A. mucronatus), PDB ID 5VOL (B. intestinalis) and PDB ID 2WTM (B. proteoclasticus) (see Fig. S2 for complete alignment). The multiple alignment is visualized using ESPript 3.0 (57).
Figure 7. wtsFae1A substrate interaction. A) Schematic drawing of XA^{5O}X (xylypyranosyl moieties, black; arabinofuranosyl moiety, green; ferulic acid, purple), and B) wtsFae1A chain A (cyan) docked to XA^{5O}X (yellow) and ferulic acid from *H. thermocellum* XynZ (PDB ID 1JT2) (white) superimposed. wtsFae1A chain residues interacting directly with XA^{5O}X and Trp^{157} are labeled and hydrogen bonds are shown as dotted lines (yellow) and their length is given in Å.
Figure 8. Comparison of wtsFae1A and wtsFae1B substrate binding cleft and ferulic acid binding pocket. A) Electrostatic plot of wtsFae1A with docked XA^{5f2}X (yellow), B) Electrostatic plot of wtsFae1B, C) superimposition of wtsFae1A (cyan) and wtsFae1B (green) surface representations with docked XA^{5f2}X (yellow), D) hydrophobicity plot of wtsFae1A (increase in the red intensity equals increase in hydrophobicity) with docked XA^{5f2}X (yellow) and E) hydrophobicity plot of wtsFae1B (increase in the red intensity equals increase in hydrophobicity).
Figure 9. CBM48 structural comparison. A) Superimposition of the wtsFae1A (cyan) and wtsFae1B (green) CBM48s and the A. thaliana starch phosphatase Starch Excess4 CBM48 (pink) (PDB ID 4PYH), B) superimposition of wtsFae1A (cyan) with docked XA^{52}X (yellow) and A. thaliana starch phosphatase Starch Excess4 (pink) in complex with maltohexaose (black), C) as B with the surface of Starch Excess4 presented and D) as B with the surface of wtsFae1A presented.
Figure 10. Related starch binding carbohydrate binding domains relation to CE1 appended CBMs. Phylogenetic tree of the CBM20 (blue), CBM48 (red), CBM69 (purple), and wtsFae1A and wtsFae1B CBM48 homologs (green). wtsFae1A and wtsFae1B CBM482 are in black. Cladogram highlighting the relative position of protein sequences identified by a reference GenBank or Uniprot acc. nr and bootstraps values are shown at the nodes. The alignment used to construct the phylogenetic tree is shown in Fig. S9. The phylogenetic tree is visualized using iTOL (58).
A carbohydrate-binding family 48 module enables feruloyl esterase action on polymeric arabinoxylan


J. Biol. Chem. published online September 26, 2019

Access the most updated version of this article at doi: 10.1074/jbc.RA119.009523

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts