


RESEARCH

Open Access



# The mechanism by which a distinguishing arabinofuranosidase can cope with internal di-substitutions in arabinoxylans

Camila Ramos dos Santos<sup>1</sup>, Priscila Oliveira de Giuseppe<sup>1</sup>, Flávio Henrique Moreira de Souza<sup>1</sup>, Letícia Maria Zanphorlin<sup>1</sup>, Mariane Noronha Domingues<sup>1</sup>, Renan Augusto Siqueira Pirolla<sup>1</sup>, Rodrigo Vargas Honorato<sup>2</sup>, Celisa Caldana Costa Tonoli<sup>2</sup>, Mariana Abrahão Bueno de Moraes<sup>1</sup>, Vanesa Peixoto de Matos Martins<sup>2</sup>, Lucas Miranda Fonseca<sup>1</sup>, Fernanda Büchli<sup>1</sup>, Paulo Sergio Lopes de Oliveira<sup>2</sup>, Fábio Cesar Gozzo<sup>3</sup> and Mário Tyago Murakami<sup>1\*</sup> 

## Abstract

**Background:** Arabinoxylan is an abundant polysaccharide in industrially relevant biomasses such as sugarcane, corn stover and grasses. However, the arabinofuranosyl di-substitutions that decorate the xylan backbone are recalcitrant to most known arabinofuranosidases (Abfs).

**Results:** In this work, we identified a novel GH51 Abf (*XacAbf51*) that forms trimers in solution and can cope efficiently with both mono- and di-substitutions at terminal or internal xylopyranosyl units of arabinoxylan. Using mass spectrometry, the kinetic parameters of the hydrolysis of 3<sup>3</sup>-α-L-arabinofuranosyl-xylotetraose and 2<sup>3</sup>,3<sup>3</sup>-di-α-L-arabinofuranosyl-xylotetraose by *XacAbf51* were determined, demonstrating the capacity of this enzyme to cleave arabinofuranosyl linkages of internal mono- and di-substituted xylopyranosyl units. Complementation studies of fungal enzyme cocktails with *XacAbf51* revealed an increase of up to 20% in the release of reducing sugars from pretreated sugarcane bagasse, showing the biotechnological potential of a generalist GH51 in biomass saccharification. To elucidate the structural basis for the recognition of internal di-substitutions, the crystal structure of *XacAbf51* was determined unveiling the existence of a pocket strategically arranged near to the — 1 subsite that can accommodate a second arabinofuranosyl decoration, a feature not described for any other GH51 Abf structurally characterized so far.

**Conclusions:** In summary, this study reports the first kinetic characterization of internal di-substitution release by a GH51 Abf, provides the structural basis for this activity and reveals a promising candidate for industrial processes involving plant cell wall depolymerization.

**Keywords:** Glycoside hydrolase family 51, Arabinofuranosidase, Arabinoxylan, Molecular mechanism, Oligomerization, Crystal structure

\*Correspondence: mario.murakami@ctbe.cnpem.br

<sup>1</sup> Brazilian Bioethanol Science and Technology Laboratory (CTBE), Brazilian Center for Research in Energy and Materials (CNPem), Campinas, Sao Paulo 13083-970, Brazil

Full list of author information is available at the end of the article



## Background

Arabinoxylan is a hemicellulosic polysaccharide composed of a  $\beta$ -1,4-linked xylose backbone, which can be mono-substituted (at O-3) or di-substituted (at O-2 and O-3) with  $\alpha$ -L-arabinofuranosyl residues (Araf) and eventually with (4-O-methyl) glucuronic acid [1, 2]. Industrially relevant biomasses such as sugarcane [3], corn stover [4] and grasses are rich in arabinoxylans, which can represent up to 50% (w w<sup>-1</sup>) of their polysaccharides in the secondary wall [2]. Moreover, arabinoxylans from cereals stimulate the activity of beneficial bacteria in the colon of humans and animals, being considered a source of prebiotic oligosaccharides with promising health-promoting properties [5, 6].

To consume arabinoxylans, the microorganisms produce a set of glycoside hydrolases including  $\alpha$ -L-arabinofuranosidases (EC 3.2.1.55) to release the Araf decorations, endo- $\beta$ -1,4-xylanases (EC 3.2.1.8), which depolymerize the backbone, and  $\beta$ -xylosidases (EC 3.2.1.37) to convert xylooligosaccharides into xylose. Xylanases are mainly categorized into families 10 and 11 of glycoside hydrolases (GH) and often display low tolerance to substitutions [7]. GH10 xylanases can accommodate substitutions at +1 subsite, but not at -2, -1, and +2 subsites [8], whereas the active site of GH11 enzymes requires at least three non-substituted residues *in tandem* for catalysis [9]. Thus, GH10 and GH11 xylanases demand the prior removal of Araf decorations by  $\alpha$ -L-arabinofuranosidases to best convert xylan into xylooligosaccharides.

$\alpha$ -L-Arabinofuranosidases (Abfs) hydrolyze non-reducing Araf groups of polysaccharides such as arabinoxylans and arabinans. They are mainly found in the GH families 43, 51, 54 and 62 and proven to have a positive effect on the enzymatic hydrolysis of pretreated wheat straw [7, 10]. In general, the characterized Abfs from family 62 seem to be specialized in cleaving Araf residues from mono-substituted xylopyranosyl (Xylp) units [11], whereas, in family GH43, some enzymes are specific for mono-substitutions [12, 13] and others recognize O3-linked Araf moieties from di-substituted xylan [14–16]. More generalist Abfs that release Araf from both mono- and di-substituted Xylp residues have been found in GH51 and GH54 families [17–20]. However, the molecular adaptations that allowed some GH51 Abfs to cleave di-substitutions remain elusive. In addition, the capacity of these enzymes to cleave di-substitutions has been mainly analyzed qualitatively without a kinetic characterization using di-substituted arabinoxylooligosaccharides (AXOs) [13, 19–22].

Thus, in this work, we reveal a novel generalist GH51 enzyme that forms trimers in solution and can cope with both mono- and di-substitutions in arabinoxylans, with

biotechnological potential for biomass saccharification. For the first time, the kinetic characterization by mass spectrometry was described for a di-substituted AXO and the structural basis for di-substitution recognition in the GH51 family was elucidated.

## Results

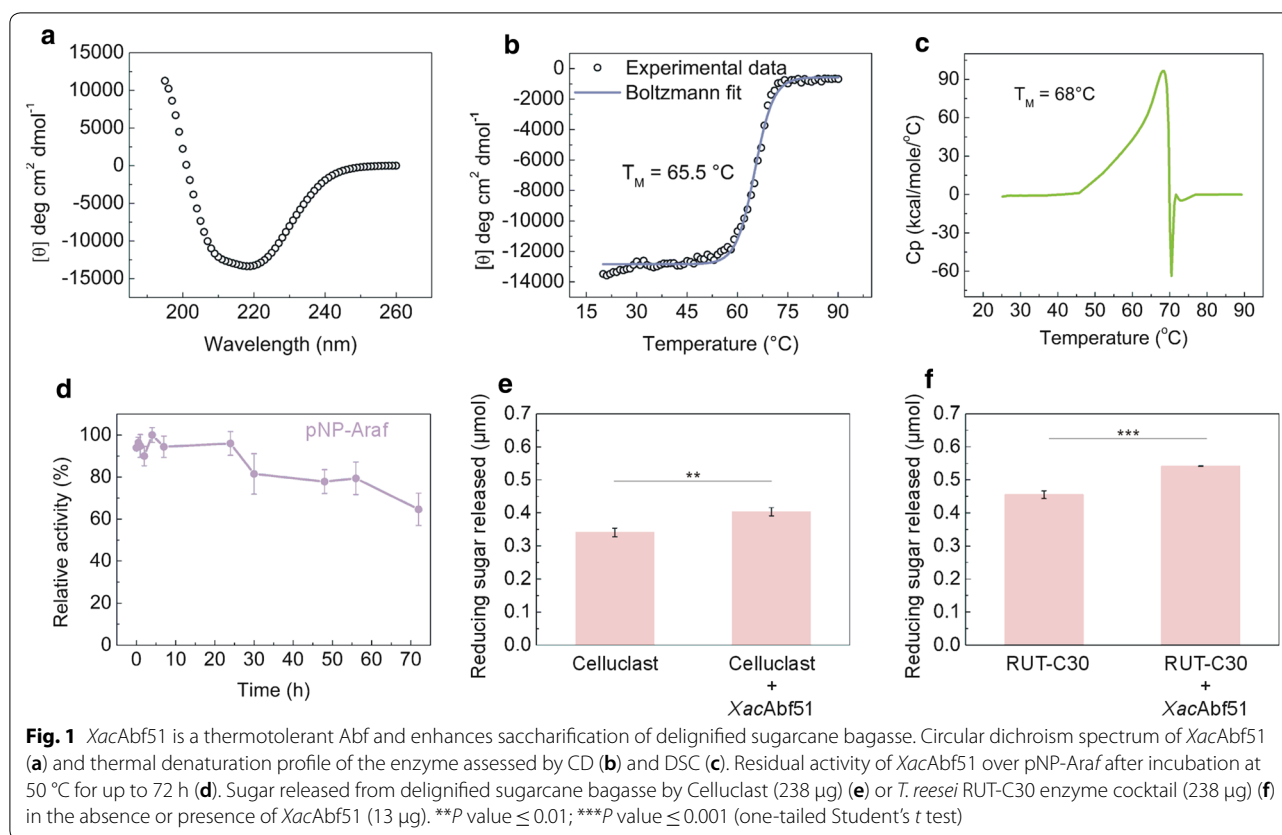
### *XacAbf51* is a thermotolerant $\alpha$ -L-arabinofuranosidase and enhances sugarcane bagasse saccharification

The enzyme *XacAbf51* fused to an N-terminal His-tag was recombinantly expressed in *Escherichia coli* cells and purified to homogeneity by metal-affinity and size-exclusion chromatography. The melting temperature ( $T_m$ ) assessed by circular dichroism spectroscopy (CD) and differential scanning calorimetry (DSC) is around 67 °C (Fig. 1a–c), indicating enhanced thermotolerance compared to other glycoside hydrolases from *X. axonopodis* pv. *citri*, which usually have a  $T_m$  between 45 and 55 °C [23]. *XacAbf51* cleaves the synthetic substrate pNP-Araf, which confirms its  $\alpha$ -L-arabinofuranosidase activity (EC 3.2.1.55). It is very stable over time, remaining active up to 45 days, when stored at 4 °C (not shown), and retaining more than 80% of its activity after 55 h incubated at 50 °C (Fig. 1d).

The prominent thermotolerance and activity of *XacAbf51* in conditions akin to those used for enzymatic hydrolysis in biorefineries led us to evaluate the biotechnological potential of this novel Abf as a complement in fungal enzyme cocktails used for sugarcane bagasse degradation, since arabinoxylan is an important component of this biomass [3]. As expected, the addition of *XacAbf51* in cellulast and RUT-C30 enzyme cocktails enhanced the hydrolysis of delignified sugarcane bagasse in near 20%, indicating that *XacAbf51* might be a useful additive in enzyme formulations for sugarcane bagasse saccharification (Fig. 1e and f).

### *XacAbf51* recognizes internal di-substituted Xylp residues

To better understand the catalytic properties of *XacAbf51*, we characterized the influence of pH and temperature on enzyme activity and investigated its substrate specificity. Maximum catalytic rates were observed at pH 5.5 (Fig. 2a) and temperature between 55 and 60 °C (Fig. 2b), which is fully compatible with the reaction conditions of commercial fungal enzyme cocktails. Besides pNP-Araf, *XacAbf51* also cleaves natural polysaccharides such as arabinoxylan and arabinan (Table 1). A comparison of the reaction with arabinan and arabinoxylan at 10 mg mL<sup>-1</sup> indicates that the enzyme cleaves arabinan better than arabinoxylan. The enzyme was not able to cleave pNP-Xylp and arabinogalactan, indicating a high specificity for Araf residues linked to xylan or arabinan backbones.



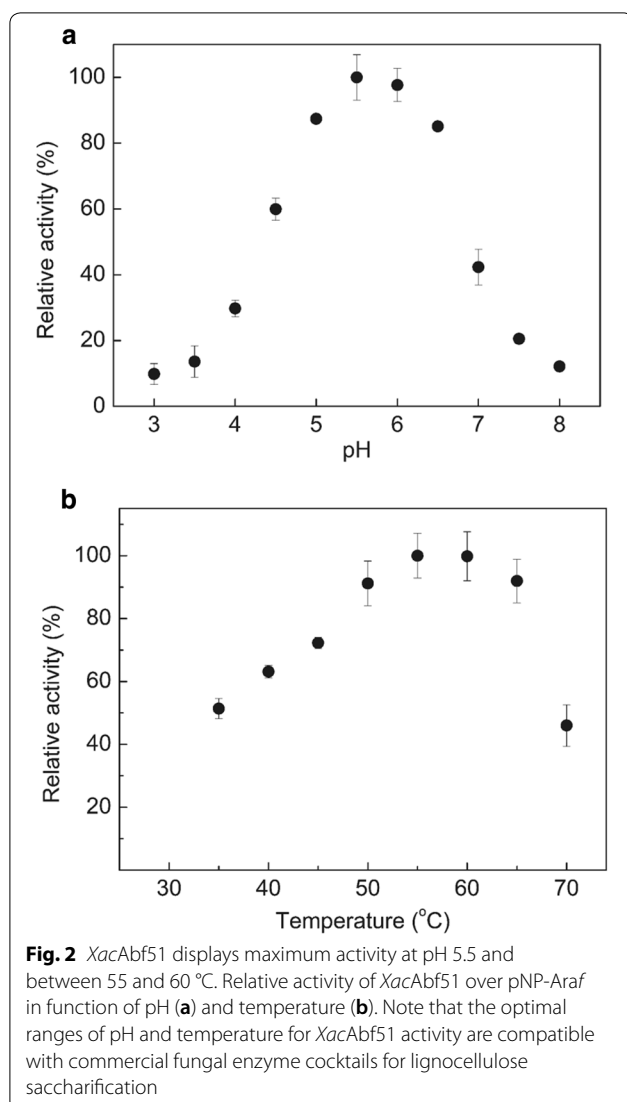
The higher activity of *XacAbf51* on arabinan as compared to arabinoxylan prompted us to investigate whether the enzyme *TxAbfD3* (EC 3.2.1.55) from *T. xylanilyticus*—a GH51 member highly active on arabinoxylan [24]—displays the same behavior. In contrast to *XacAbf51*, the enzyme *TxAbfD3* was more active on arabinoxylan than on arabinan, showing that distinct substrate preferences occur within the family GH51, despite their capacity to recognize several substrates.

According to capillary zone electrophoresis data, *XacAbf51* releases arabinose from arabinoxylan as well as from mono- and di-substituted *Xylp* residues located at the non-reducing end or within the backbone of AXOS (Fig. 3). To better characterize the action of *XacAbf51* on internal (di)-substitutions, we monitored the enzymatic hydrolysis of 2<sup>3</sup>, 3<sup>3</sup>-di- $\alpha$ -L-arabinofuranosyl-xylotetraose (XA<sup>2+3</sup>XX) and 3<sup>3</sup>- $\alpha$ -L-arabinofuranosyl-xylotetraose (XA<sup>3</sup>XX) using mass spectrometry. In the reactions containing XA<sup>2+3</sup>XX as substrate, we detected the product xylotetraose but not mono-substituted AXOS, indicating that *XacAbf51* cleaves both Araf moieties from internal di-substitutions (Fig. 4a). The enzyme showed a  $k_{\text{cat}}$  of  $9.6 \pm 0.5 \text{ s}^{-1}$  and  $K_{\text{m}} = 4.97 \pm 0.48 \text{ mM}$  against the di-substituted substrate ( $k_{\text{cat}}/K_{\text{m}} = 1.8 \times 10^3 \text{ s}^{-1} \text{ M}^{-1}$ ), although higher specificity was observed for the

O3-linked mono-substitution ( $k_{\text{cat}} = 73.3 \pm 0.9 \text{ s}^{-1}$ ,  $K_{\text{m}} = 2.82 \pm 0.08 \text{ mM}$ ,  $k_{\text{cat}}/K_{\text{m}} = 2.5 \times 10^4 \text{ s}^{-1} \text{ M}^{-1}$ ), which agrees with the lack of detection of mono-substituted intermediates in the XA<sup>2+3</sup>XX hydrolysis (Fig. 4a). The homologous enzyme *TxAbfD3* ( $15 \mu\text{g mL}^{-1}$ ), which was active against XA<sup>3</sup>XX ( $v_0/[E]_t = 88.5 \pm 2.6 \text{ s}^{-1}$  at 10 mM substrate), displayed 20-fold lower activity against the di-substituted substrate XA<sup>2+3</sup>XX ( $v_0/[E]_t = 0.40 \pm 0.03 \text{ s}^{-1}$  at 10 mM substrate) compared to *XacAbf51* ( $v_0/[E]_t = 7.96 \pm 0.85 \text{ s}^{-1}$ ) assayed in the same conditions, indicating that *XacAbf51* underwent molecular adaptations to better cleave di-substituted AXOS.

#### Structural basis for the cleavage of AX di-substitutions by *XacAbf51*

To investigate the molecular mechanisms by which *XacAbf51* cleaves AX di-substitutions, we solved and analyzed its crystal structure. As a typical GH51 enzyme, *XacAbf51* harbors the active site in a ( $\beta/\alpha$ )<sub>8</sub>-barrel that is tightly associated with a  $\beta$ -sandwich domain. The  $\beta$ -sheets of this  $\beta$ -sandwich put the N- and the C-terminal regions of the barrel together, stabilizing these two regions that otherwise would be labile (Fig. 5a). Thus, although not participating in the catalysis, the



$\beta$ -sandwich domain seems to be essential for the catalytic domain stability.

Structural comparisons revealed that *XacAbf51* displays all structural features required for the retaining

mechanism of hydrolysis conserved in GH51 enzymes [25–27]. The catalytic residues Glu182 (acid–base) and Glu304 (nucleophile) are positioned 3.7 Å apart from each other within the active site pocket (Fig. 5a). A glycerol molecule occupied the –1 subsite in a conformation that mimics part of the Araf ring (Fig. 5b). All residues from this subsite are identical or semi-conserved between *XacAbf51* and GH51 structures known so far, except for Cys80 and Cys186. These cysteine residues form a disulfide bridge in *XacAbf51* and *TxAbfD3*, which likely contributes to the high thermostability of these enzymes [24]. In other GH51 Abfs, Cys80 and Cys186 residues are replaced by asparagine and glutamine (Fig. 5c). Although Asn181 is fully conserved between the compared GH51 Abfs, it adopts a different rotamer in *XacAbf51* (Fig. 5b).

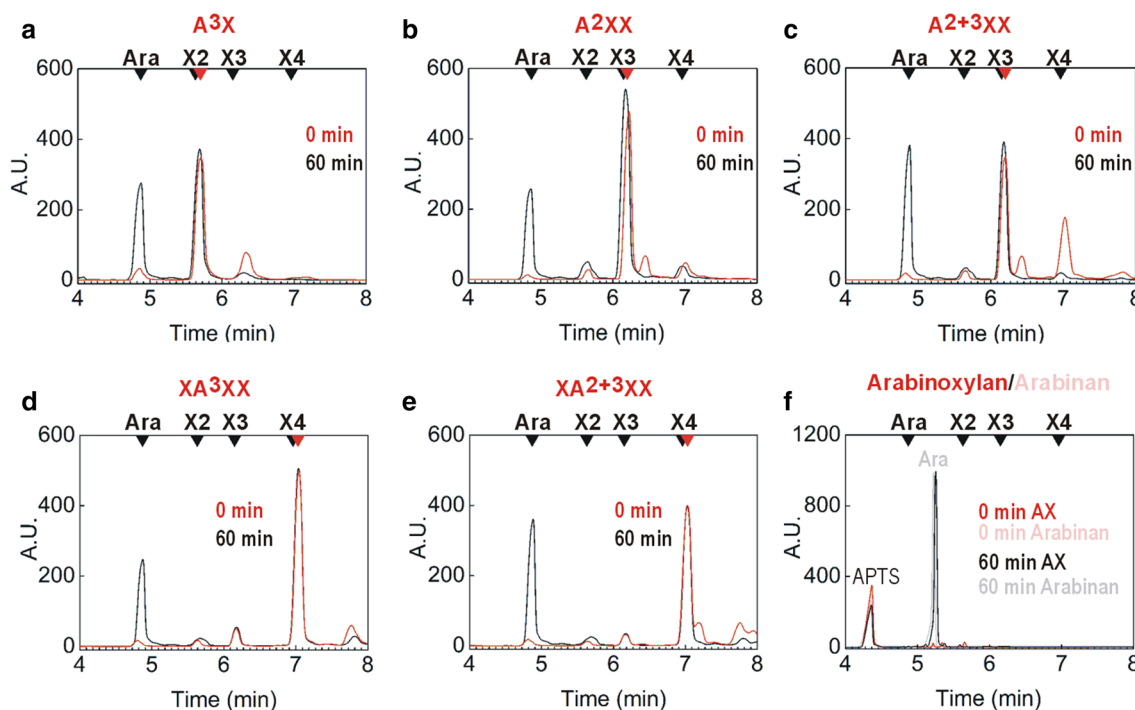
Structural superimposition of *XacAbf51* with *TxAbfD3* in complex with 3<sup>2</sup>- $\alpha$ -L-arabinofuranosyl-xylotriose (XA<sup>3</sup>X) evidenced the presence of a cavity near to the –1 subsite that could potentially accommodate the second Araf substitution of a di-substituted substrate (Fig. 6a). To gain insights into the molecular events involved in binding and hydrolysis of Araf from internal di-substituted Xylp residues, we appended an O<sub>2</sub>-linked Araf at XA<sup>3</sup>X, thus generating XA<sup>2+3</sup>X, and carried out a molecular dynamics (MD) simulation of *XacAbf51* complexed with this di-substituted substrate. According to this simulation, the side chains of Ser222 and Asp223 adopted different rotameric conformations to better accommodate the O<sub>2</sub>-linked Araf at the +2NR\* subsite (Fig. 6b). The side chain of Asn181 rotated 180° around C $\beta$  to interact with the O<sub>2</sub> atom of the arabinofuranosyl residue at the –1 subsite. Trp254 formed hydrophobic interactions with the +2R Xylp residue, but no hydrogen bonds were observed between the enzyme and the xylan backbone, which correlates with the versatility of *XacAbf51* in recognizing both arabinoxylan and arabinan. Selected inter-atomic distances between enzyme and XA<sup>2+3</sup>X remained stable over the simulation, indicating favorable interactions for substrate binding (Fig. 6c). Thus, the MD simulation data support that the pocket adjacent to –1 subsite can accommodate the O<sub>2</sub>-linked Araf from

**Table 1** Kinetic parameters of *XacAbf51* and *TxAbfD3* on pNP-Araf and arabinan and comparative activity of *XacAbf51* and *TxAbfD3* on arabinoxylan.

	pNP-Araf			Arabinan			Arabinoxylan <sup>a</sup>
	$K_m$ (mM)	$k_{cat}$ (s <sup>-1</sup> )	$k_{cat}/K_m$ (s <sup>-1</sup> M <sup>-1</sup> )	$K_m$ (mg mL <sup>-1</sup> )	$k_{cat}$ (s <sup>-1</sup> )	$k_{cat}/K_m$ (s <sup>-1</sup> mg <sup>-1</sup> mL)	$v_0/[E]_t$ (s <sup>-1</sup> )
<i>XacAbf51</i>	0.19 ± 0.01	56.4 ± 0.8	3 × 10 <sup>5</sup>	20.4 ± 7.8	56.7 ± 6.4	2.8	9.1 ± 0.55
<i>TxAbfD3</i>	0.52 ± 0.06	154.2 ± 4.2	3 × 10 <sup>5</sup>	41 ± 9.5	27.15 ± 2.2	0.7	30.0 ± 1.31

For comparative purposes, the assays were performed at 50 °C, pH 5.5, which is compatible with industrial processes of biomass saccharification. The  $K_m$  and  $k_{cat}$  could not be estimated for arabinoxylan because the maximum velocity is not reached at the highest possible concentration of the substrate

<sup>a</sup> Activity measured using substrate at 10 mg mL<sup>-1</sup>



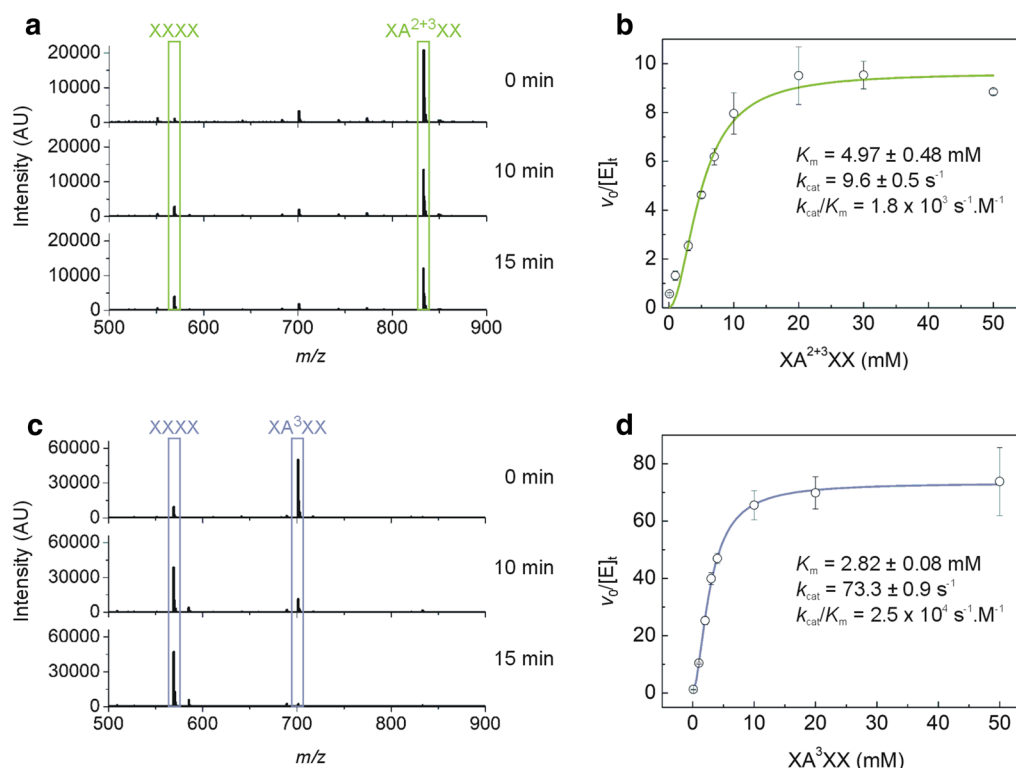
**Fig. 3** *XacAbf51* releases Ara<sub>f</sub> from mono- and di-substituted AXOS and from arabinoxylan. Capillary zone electrophoresis profiles of AXOS before (red lines) and after (black lines) incubation with *XacAbf51*. Although the peaks of decorated and undecorated oligosaccharides were indistinguishable in this assay, the increase of arabinose (Ara) peak after enzyme treatment shows the capacity of *XacAbf51* to release Ara<sub>f</sub> from several AXOS and from arabinoxylan and arabinan. **a** A<sup>3</sup>X = 3<sup>2</sup>-α-L-arabinofuranosyl-xylobiose; **b** A<sup>2</sup>XX = 2<sup>3</sup>-α-L-arabinofuranosyl-xylotriose; **c** A<sup>2+3</sup>XX = 2<sup>3</sup>, 3<sup>3</sup>-di-α-L-arabinofuranosyl-xylotriose; **d** XA<sup>3</sup>XX = 3<sup>3</sup>-α-L-arabinofuranosyl-xylotetraose; **e** XA<sup>2+3</sup>XX = 2<sup>3</sup>, 3<sup>3</sup>-di-α-L-arabinofuranosyl-xylotetraose; **f** arabinoxylan from wheat flour and arabinan from sugar beet. Black arrowheads represent the migration time of arabinose (Ara), xylobiose (X2), xylotriose (X3) and xylo-tetraose (X4) standard runs. Red arrowheads represent the substrate migration time (0 min, without enzyme). In (**f**), the Ara released from arabinan was used as a reference for the analysis of arabinoxylan cleavage, due to the anomalous migration of Ara in these conditions, compared to the standard run

internal di-substituted Xyl<sub>p</sub> residues, while the O3-linked decoration is placed into −1 subsite for catalysis. Considering the pseudosymmetry of xylan and the design of catalytic interface, the backbone might also bind to the active site in the inverted direction, placing, in this case, the internal O2-linked Ara<sub>f</sub> (from mono- or di-substitutions) into −1 subsite for cleavage.

In TxAbfD3, we observed variable regions at β6–α6 and β5–α5 loops that might explain its low activity against internal di-substitutions. The β6–α6 loop contains the tryptophan residue that interacts with the +1 Xyl<sub>p</sub> unit in TxAbfD3, but makes hydrophobic contacts with the +2 Xyl<sub>p</sub> residue in *XacAbf51* (Fig. 6d). To test the influence of β6–α6 loop in substrate preference, the sequence TIPGGWPPRASST (Thr249–Thr261) and two extra residues (Ala310–Pro311) of *XacAbf51* were replaced by the sequence TVPGPWEKKGPAT and DV of TxAbfD3, because the aspartic residue from the DV motif interacts with β6–α6 loop in TxAbfD3. CD analysis indicated a folded conformation of the mutant (data not shown); however, it was inactive against arabinan

and arabinoxylan and poorly active against pNP-Ara<sub>f</sub>. Another point of divergence between *XacAbf51* and TxAbfD3 is the sequence SDD (Ser222–Asp224, β5–α5 loop) of *XacAbf51*, which is replaced by the NTA (Asn216–Ala218) motif in TxAbfD3, attracting the +2 Xyl<sub>p</sub> unit via a hydrogen bond donated by Asn216 (Fig. 6d). This three-residue replacement caused enzyme aggregation, as assessed by Dynamic Light Scattering (DLS), and disrupted the enzyme activity against arabinoxylan and arabinan (results not shown). We also tested whether the triple replacement of β6–α6 loop, DV and SDD motifs would convert the substrate preference of *XacAbf51* to that of TxAbfD3. Although the mutant showed a folded conformation with a similar hydrodynamic radius (*R<sub>h</sub>*) to the WT enzyme, the triple modification also abolished the *XacAbf51* activity against arabinoxylan and arabinan, indicating that other structural features might affect the positioning and dynamics of β6–α6 and β5–α5 loops, impairing activity when associated with transplanted loops.





**Fig. 4** *XacAbf51* hydrolyzes internal mono- and di-substitutions of AXOS. **a** Mass spectra of  $XA^{2+3}XX$  ( $2^3, 3^3$ -di- $\alpha$ -L-arabinofuranosyl-xylotetraose) after 0, 10 and 15 min of reaction with *XacAbf51* (15  $\mu g mL^{-1}$ ) at 50 °C, pH 5.5, using 10 mM substrate. Boxed peaks correspond to the sodiated forms of the substrate  $XA^{2+3}XX$  ( $m/z = 810 + 23$  ( $Na^+$ ) = 833) and the product xylotetraose (XXXX,  $m/z = 546 + 23$  ( $Na^+$ ) = 569). **b** Kinetic parameters of  $XA^{2+3}XX$  hydrolysis by *XacAbf51* (15  $\mu g mL^{-1}$ ) at 50 °C, pH 5.5, assessed by mass spectrometry in triplicate. **c** Mass spectra of  $XA^3XX$  ( $3^3$ - $\alpha$ -L-arabinofuranosyl-xylotetraose) after 0, 10 and 15 min of reaction with *XacAbf51* in the same conditions described in (a). Boxed peaks correspond to the sodiated forms of the substrate  $XA^3XX$  ( $m/z = 678 + 23$  ( $Na^+$ ) = 701) and the product XXXX ( $m/z = 546 + 23$  ( $Na^+$ ) = 569). **d** Kinetic parameters of  $XA^3XX$  hydrolysis by *XacAbf51* (15  $\mu g mL^{-1}$ ) at 50 °C, pH 5.5, assessed by mass spectrometry in triplicate. Error bar represents standard deviations of the mean

### The biological unit of *XacAbf51* is a trimer

In the crystal structure of *XacAbf51*, six protein chains compose the asymmetric unit, but in a different spatial disposition from that observed for known GH51 hexamers such as *TxAbfD3* [28] (Fig. 7a). Analysis of the crystal interfaces using jsPISA [29] indicates that trimers, composed by ABC or DEF chains, are the most stable quaternary structure of *XacAbf51*. Moreover, the interface between the dimer of trimers that compose the *TxAbfD3* hexamer is not conserved in *XacAbf51*.

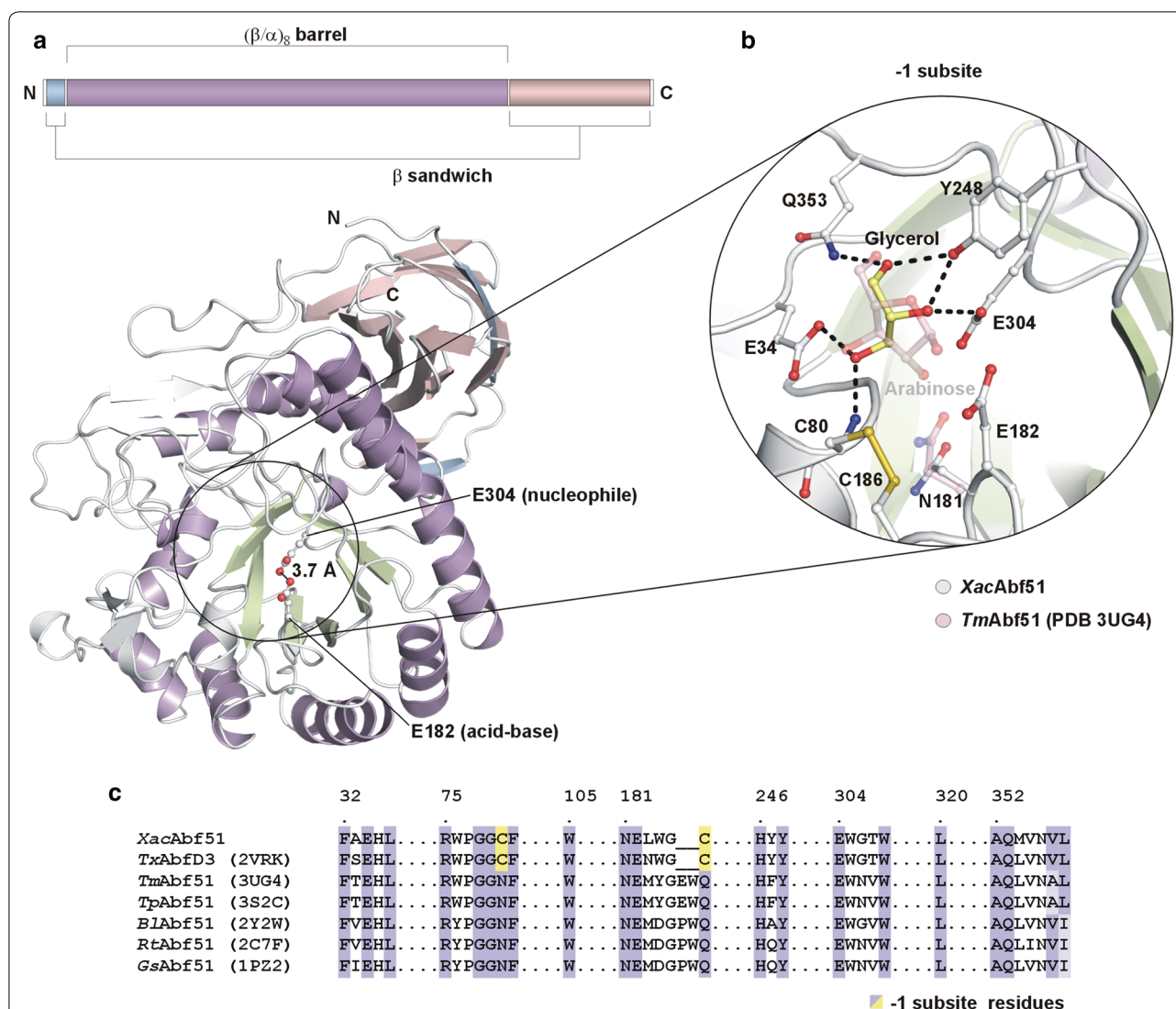
To determine the oligomeric state of *XacAbf51* in solution, several experiments were carried out with the purified protein. The small angle X-ray scattering (SAXS) curve of *XacAbf51* revealed a radius of gyration (4.5 nm) and a low-resolution molecular envelope that are consistent with the crystallographic trimer (Fig. 7b). Moreover, the sedimentation coefficient estimated from analytical ultracentrifugation (AUC) at different protein concentrations (Fig. 7c and d) corresponds to a particle of 161 kDa, which is in accordance with the theoretical mass of the

trimer (171 kDa). Estimation of  $R_h$  using DLS (Fig. 7e) further supported that the biological unit of *XacAbf51* is a trimer.

### Evolution of GH51 enzymes

To gather insight into the evolution of GH51 Abfs, a phylogenetic tree was constructed based on the catalytic domain of characterized GH51 enzymes and their respective paralogues (Fig. 8). This phylogenetic reconstruction shows two major clades (clades I and II) referent to a gene duplication that occurred early in evolution, as indicated by the presence of genes from the two clades in *Thermotoga petrophila*, a species from a deep phylogenetic branch in the tree of life [30]. Members of clade I are abundant in bacteria, whereas those of clade II are found mainly in plants and fungi.

The division in two major clades reflects two main types of modular architecture. In clade I, most enzymes display the  $(\beta/\alpha)_8$  barrel +  $\beta$ -sandwich composition, but, in clade II, the proteins have an extra N-terminal domain which

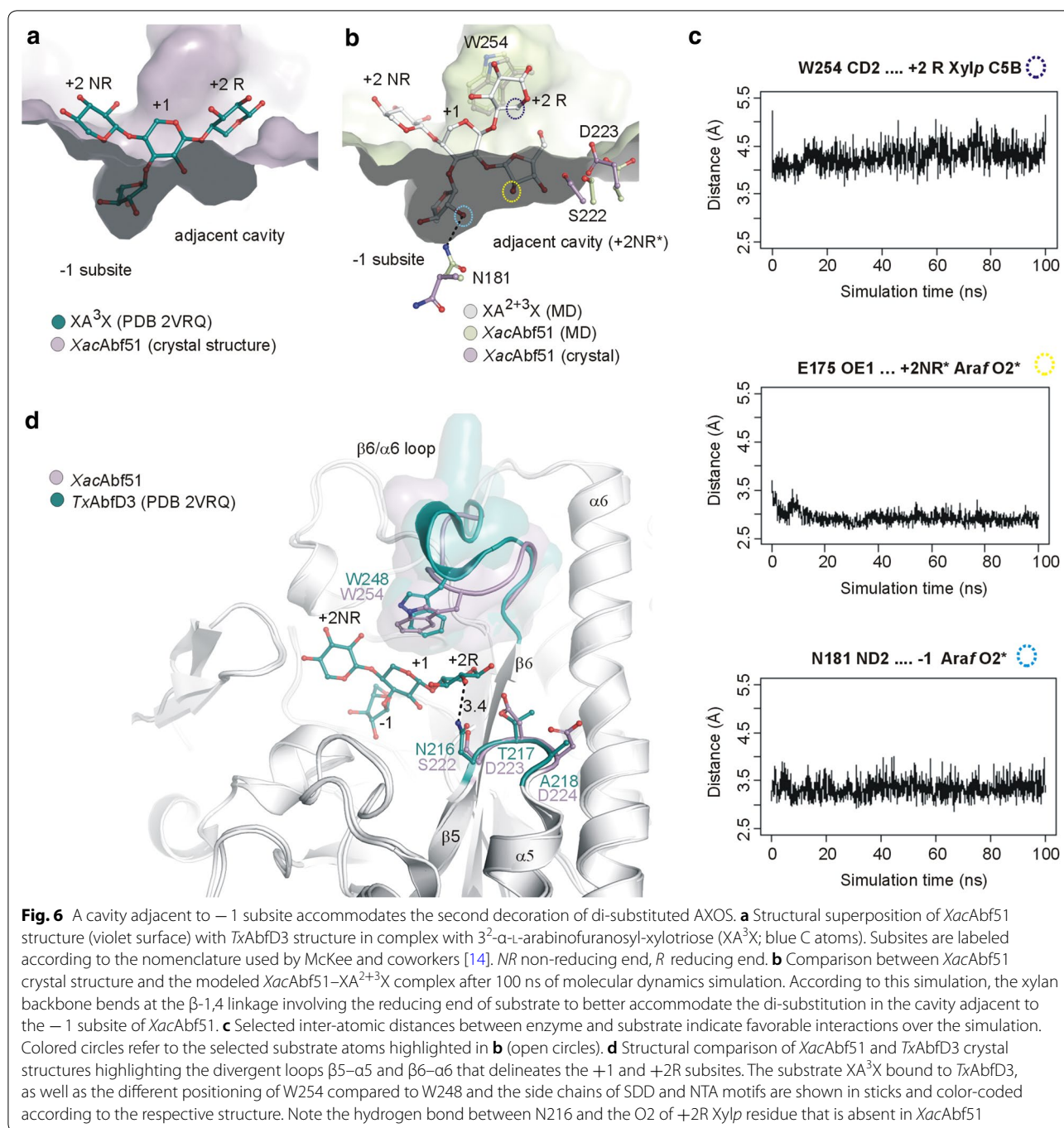


**Fig. 5** Crystallographic structure of *XacAbf51* reveals a typical fold of GH51 arabinofuranosidases and a disulfide bridge at -1 subsite conserved in *TxAbfD3*, but divergent in other structurally characterized GH51 enzymes. **a** Scheme of *XacAbf51* domain architecture (top) and cartoon representation of the 3D structure (bottom) highlighting the distance (3.7 Å) between the catalytic residues (sticks) compatible with the retaining mechanism of hydrolysis found in GH51 family. **b** Magnified view of -1 subsite (ball and sticks, light gray C atoms) in which a glycerol molecule (yellow C atoms) is bound mimicking part of the arabinose scaffold observed in the crystallographic structure of *TmAbf51*-arabinose complex (pink C atoms). **c** Structure-based sequence alignment of -1 subsite (boxed residues) from the GH51 enzymes of known structure. Dark violet represents identical residues, light violet semi-conserved and yellow highlights the cysteine residues that form a disulfide bridge only in the *XacAbf51* and *TxAbfD3* enzymes of the presented comparison. *Tm*: *Thermotoga maritima*; *Tp*: *Thermotoga petrophila*; *Bl*: *Bifidobacterium longum*; *Rt*: *Ruminiclostridium thermocellum*; *Gs*: *Geobacillus stearothermophilus*

resembles carbohydrate-binding modules (CBM) from families 4, 6 or 11 (Fig. 8). Interestingly, enzymes with β-1,4-glucanase activity, found only in specific bacteria from *Fibrobacter* and *Alicyclobacillus* genera (clade Ib), have peculiar and diverse domain arrangements, indicating they emerged from gene duplication and recombination events. In these enzymes, the (β/α)<sub>8</sub> barrel is usually fused to one or more copies of putative cellulose-binding

modules (CBM 3, 11 and 30). Moreover, unconventional domains (Gp9-like and cupredoxin-like) are detected in two endoglucanases from *Alicyclobacillus* sp.

To date, the only structures available for the GH51 family comprise Abfs from clade Ia with the (β/α)<sub>8</sub> barrel + β-sandwich composition. Except for *XacAbf51*, which is a trimer, the other structures reported so far



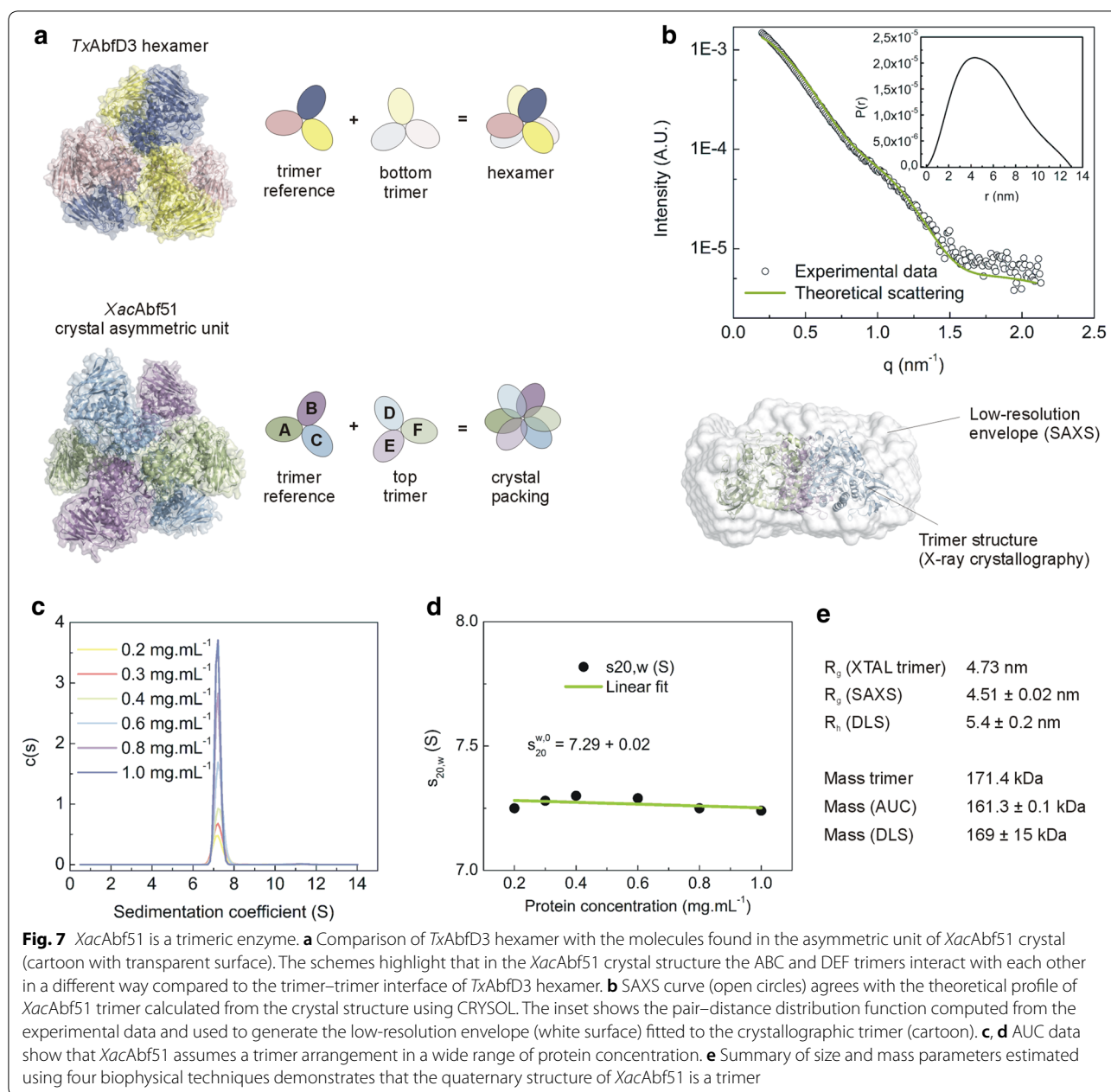
are hexamers, indicating that the molecular diversity of GH51 enzymes include changes in quaternary structure besides modular rearrangements. The capacity to cleave  $\alpha$ -1,2 and  $\alpha$ -1,3 *Araf* decorations in arabinoxylan and/or arabinan as well as  $\alpha$ -1,5 bonds in arabinan is observed in Abfs from both clades I and II, evidencing

the structural plasticity of the GH51 active site [13, 20, 31–37].

## Discussion

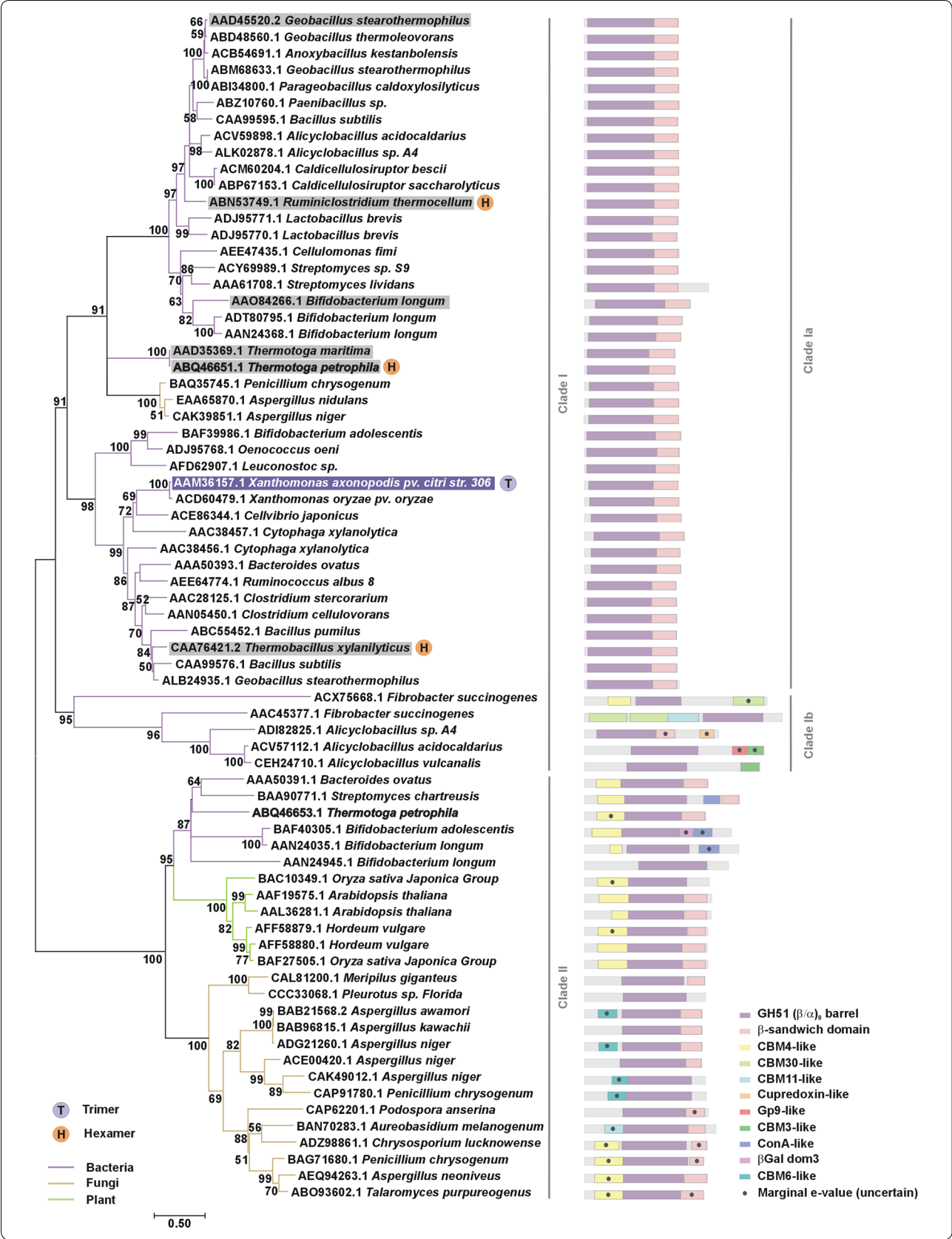
This study reports the first Michaelis–Menten kinetic parameters for the cleavage of internal *Araf* di-substitutions by a GH51 Abf and provides the structural basis





(See figure on next page.)

**Fig. 8** Molecular phylogenetic analysis of GH51 family. Phylogenetic tree (unrooted) based on a multiple sequence alignment of the  $(\beta/\alpha)_8$  barrel of characterized GH51 enzymes present in the CAZY database [7] and the respective paralogues. The evolutionary history was inferred using the maximum likelihood method implemented in the MEGA7 software [65, 67]. The tree with the highest log likelihood ( $-23,649.75$ ) is shown and the percentage of trees in which the associated taxa clustered together are shown next to the branches (except for those with values below 50%). Branch lengths represent the number of substitutions per site. The right panel shows the domain architecture predicted for each sequence using the webserver SUPERFAMILY [63]. Proteins with known 3D structure are highlighted with purple (XacAbf51) or gray boxes. Paralogous sequences from *T. petrophila* are shown in bold



for this activity. Cleavage of terminal di-substitutions in AXOS has been reported for some GH51 enzymes, but internal di-substitutions have been described as poor or non-cleavable substrates [17–20]. Our data reveal a novel GH51 enzyme that releases both Araf residues from internal di-substitutions with a catalytic constant of  $\sim 10 \text{ s}^{-1}$ . Although our data do not resolve the *XacAbf51* preference between O2 or O3 linkages, they reveal that the first cleavage of a di-substitution is the rate-limiting step of the reaction catalyzed by *XacAbf51*, leading to a tenfold lower  $k_{\text{cat}}/K_{\text{m}}$  for the di-substituted compared to the O3-mono-substituted substrate.

For almost all GH51 enzymes characterized so far, kinetic parameters have only been assessed using synthetic substrates (pNP derivatives), probably because of the high-cost and limited availability of AXOS allied to the low response stability and time-consuming characteristic of HPAEC-PAD analyses [38]. To overcome such bottlenecks, we used mass spectrometry to monitor the enzymatic hydrolysis of mono(di)-substituted arabinoxylotetraoses—a fast, direct and highly sensitive approach that requires minimum amounts of substrate (in this study, we acquired each data point in 1 min and used less than 10 mg of substrate for a complete enzyme characterization). Thus, we envisage the mass spectrometry as a useful, fast and precise alternative, not only for future studies of GH51 enzymes, but also to assess Michaelis–Menten kinetics of oligosaccharide hydrolysis by other GHs, as previously reported for xylanases [39].

The positive effect of *XacAbf51* in the saccharification of delignified sugarcane bagasse may be useful for the development of enzyme cocktails optimized for this biomass. Supplementation of fungal cellulases mixtures with hemicellulases and auxiliary enzymes, including a GH51 Abf, has already been shown to increase the conversion of AFEX pretreated corn stover into monosaccharides [40]. Here we evidence that this approach is also valuable to increase the hydrolysis yield of pretreated sugarcane bagasse. The cellulolytic fungi *T. reesei* displays three Abfs (GH43, GH62 and GH54), but is devoid of GH51 enzymes [41]. Thus, our data support that the *XacAbf51* capacity of releasing terminal and internal di-substitutions of AXOS might improve the performance of widely

used cellulolytic enzyme cocktails over arabinoxylan-rich biomasses.

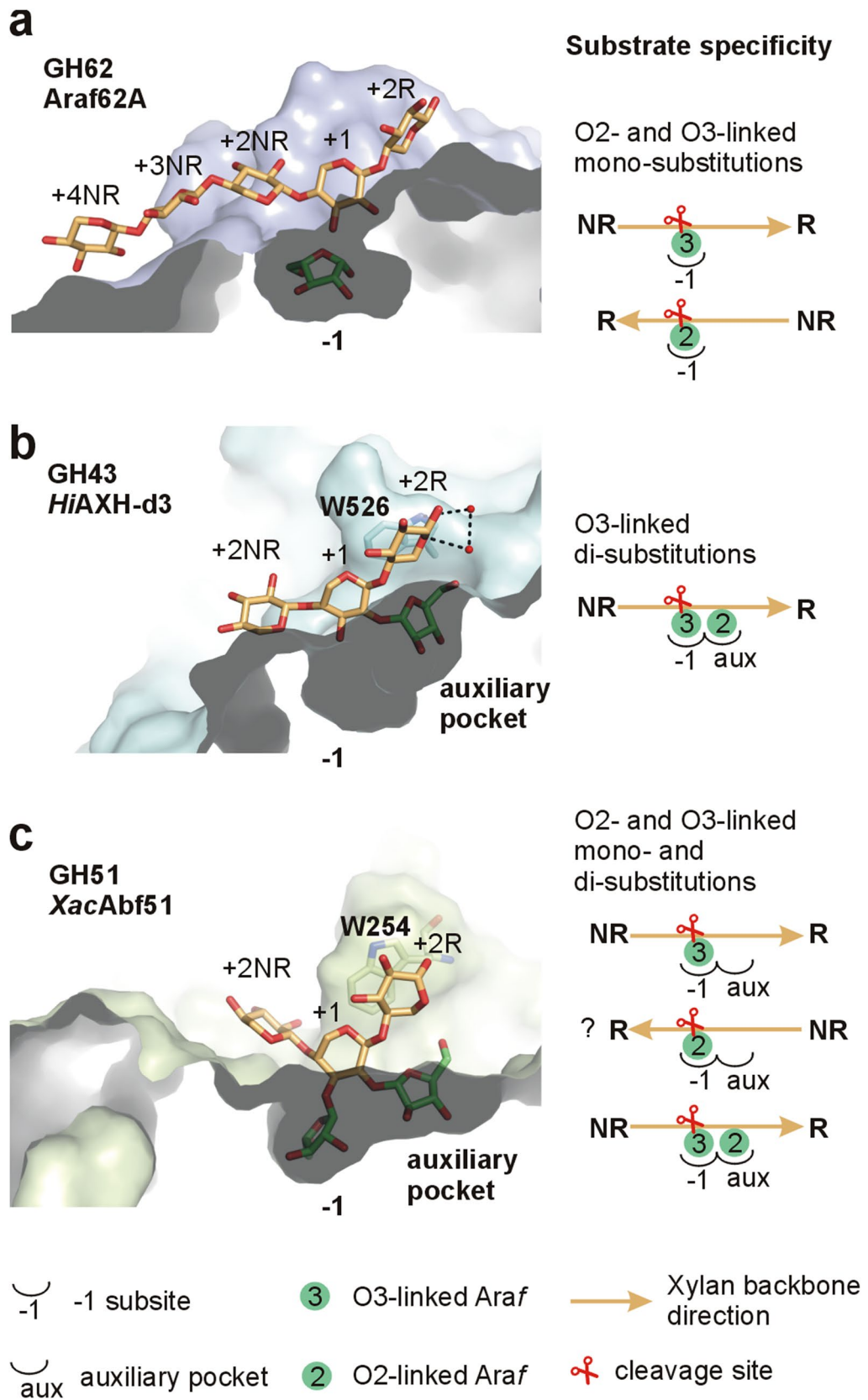
Our structural data compared to those of GH62 and GH43 Abfs (EC 3.2.1.55) contribute to a better understanding of the molecular determinants for distinct substrate specificities in Abfs. GH62 enzymes specialized in mono-substitutions display a single arabinose-binding pocket in the middle of a long cleft where the xylan backbone binds (Fig. 9a). As proposed by Maehara and coworkers, the pseudosymmetry of xylan backbone and the active site topology of Araf62A likely allows arabinoxylan to bind into the cleft in two opposite directions to, respectively, allocate the O3- and O2-linked mono-substitutions at the  $-1$  subsite [42]. Differently, in the GH43 enzyme *HiAXH-d3*, which is specific for O3-linked Araf from di-substitutions, an auxiliary pocket accommodates the second Araf decoration and solvent-mediated hydrogen bonds (involving Trp526 and the ring oxygen of  $+2\text{R Xylp}$ ) selects a single orientation of the xylan backbone, in a manner that the catalytic pocket is always occupied by the O3-Araf moiety (Fig. 9b) [14]. Similar to *HiAXH-d3*, *XacAbf51* also displays an auxiliary pocket to accommodate the second substitution of di-substituted substrates (Fig. 9c). However, the residue Trp254 (equivalent to Trp526 of *HiAXH-d3*) makes a  $\pi$ -stacking interaction with  $+2\text{R Xylp}$ , which does not depend on the endocyclic oxygen, the only asymmetric feature of xylan. Thus, according to these analyses, it is plausible to suggest that the active site of *XacAbf51* allows the bidirectional binding of arabinoxylan and AXOS to cleave O2- and O3-linked Araf from mono- or di-substitutions.

The positioning of Trp254 seems to play a role in di-substitution recognition. However, our mutational strategy to test this hypothesis ( $\beta 6$ – $\alpha 6$  and/or  $\beta 5$ – $\alpha 5$  loops transplantation from *TxAbfD3* to *XacAbf51*) inactivated *XacAbf51* instead of changing its substrate specificity, indicating an incompatibility that may require secondary mutations or the reverse transplant (from *XacAbf51* to *TxAbfD3*) to attain the expected functional changes.

All GH51 proteins whose structure is currently available are bacterial enzymes from Clade Ia (Fig. 8). The oligomeric state of only three of them has been validated *in solution* [*TrAbf51*, [43]; *TxAbfD3* (AUC data not shown)

(See figure on next page.)

**Fig. 9** Molecular diversity of arabinoxylan-degrading mechanisms by Abfs. **a** The active site of Araf62A (GH62) is composed by a cleft that accommodates the xylan backbone and a  $-1$  subsite that binds specifically to mono-substitutions of Araf (O2- or O3-linked). The arabinose (green C atoms) and protein surface are from PDB 3WN0, while the xylan backbone (orange C atoms) is from PDB 3WN2 [42]. NR non-reducing end, R reducing end. **b** *HiAXH-d3* from *Humicola insolens* (GH43) cleaves specifically the O3 linked Araf substitution from di-substituted Xylp units and displays an auxiliary pocket to accommodate the di-substitution. The residues W526 selects a single orientation for arabinoxylan binding into the active site via solvent-mediated interactions (dashed lines) with the endocyclic oxygen of  $+2\text{R Xylp}$  (PDB 3ZXK, [14]). **c** The active site of *XacAbf51* also has an auxiliary pocket to accommodate di-substitutions, but the positioning of W254 seems to accept the binding of arabinoxylan in the direct and reverse direction to allow the cleavage of O3 and O2 substitutions, respectively, making this enzyme a generalist Abf





and *XacAbf51* (Fig. 7)] and served as a guide to map how the quaternary structure of GH51 enzymes evolved. The hexameric arrangement, which can be seen as dimer of trimers, seems to have appeared early during evolution of GH51 family, being found in the *Thermotoga* genus, a deep lineage back to the early forms of bacteria [30, 43]. The hexameric arrangement remained stable in other thermophilic bacteria, such as *Ruminiclostridium thermocellum* (jsPISA prediction, [29]) and *T. xylanilyticus* [28]), but, in the mesophilic *X. axonopodis* pv. *citri*, the dimer of trimers was disrupted, giving rise to a trimeric enzyme. Based on these data, we suggest that the ancient GH51 arabinofuranosidases from clade I formed hexamers—possibly to withstand extreme conditions of high temperature—and that colder environments favored the emergence of trimeric enzymes, at least during *X. axonopodis* pv. *citri* speciation, changing the paradigm that GH51 Abfs are exclusively hexameric. According to ConSurf analyses [44], the trimer interface, which is close to the active site, harbors residues more conserved than those assembling trimers into hexamers, indicating that the trimeric arrangement may be more crucial than the hexameric configuration for enzyme function.

## Conclusions

In summary, our study expands our knowledge about the diversity of GH51 Abfs in terms of tertiary and quaternary structure and provides the structural basis for the release of internal Araf di-substitutions by a generalist Abf that copes with all types of Araf decorations in arabinoxylan and arabinan. The rare mode of action of *XacAbf51*, along with full pH and temperature compatibility with current fungal enzyme cocktails, is very attractive for industrial applications, especially in technologies for the production of fermentable sugars using arabinoxylan-rich biomasses such as sugarcane, corn stover and grasses.

## Methods

### Molecular cloning

The nucleotide sequence encoding *XacAbf51* (GenBank AAM36157.1) was amplified from the genomic DNA of *X. axonopodis* pv. *citri* str. 306 using the following oligonucleotides: 5'- CAT ATG CCG GCC AGC TTC GCT G -3' and 5'- AAG CTT TCA TTG CAG CTT GAG CAT CAC GAT CG -3'. It was cloned into pET28a after digestion with *NdeI* and *HindIII* restriction enzymes. The annotated sequence begins with GTG codon but upstream sequence analysis indicated additional 27 nucleotides (beginning at ATG) that are part of the signal peptide according to PROSECTO (<http://www.lge.ibi.unicamp.br/lnbio/prosecto.htm>). The signal peptide was removed during cloning for expression

in *E. coli*. The DNA sequences of mutants I (containing 249-TVPGPWEEKGPAT-261 and 310-DV-311 instead of 249-TIPGGWPPRASST-261 and 310-AP-311), II (containing 222-NTA-224 instead of 222-SDD-224) and III (mutations I+II) were produced in pET28a vector, between *NdeI* and *XhoI* restrictions sites, by GenScript (Piscataway, NJ). The construct *TxAbfD3* (GenBank CAA76421.2) cloned into pET21a between *NdeI* and *HindIII* restriction sites was also purchased from GenScript (Piscataway, NJ).

### Protein production and purification

*XacAbf51* and mutants were expressed in *Escherichia coli* Origami™2(DE3) cells in Terrific broth medium (1.2% (m v<sup>-1</sup>) tryptone, 2.4% (m v<sup>-1</sup>) yeast extract, 0.4% (v v<sup>-1</sup>) glycerol, 17 mM sodium phosphate monobasic monohydrate, 72 mM sodium phosphate dibasic) supplemented with 50 µg mL<sup>-1</sup> kanamycin. The culture was grown at 37 °C, 225 rpm, until the O.D.<sub>600nm</sub> has reached 1.0, transferred to 20 °C for 1 h, and then incubated with 0.25 mM isopropyl β-D-1-thiogalactopyranoside (Sigma-Aldrich, St. Louis, MO) for 18 h at 20 °C, 170 rpm. The *TxAbfD3* protein was produced in *E. coli* BL21 (DE3) cells grown in Luria–Bertani medium supplemented with 100 µg mL<sup>-1</sup> ampicillin, following the same protocol used for *XacAbf51*. The cells were collected, resuspended in lysis buffer (20 mM sodium phosphate, pH 7.5, 500 mM NaCl, 5 mM imidazole, 1 mM PMSF, 0.5 mg mL<sup>-1</sup> lysozyme), incubated on ice for 30 min and disrupted by sonication. The soluble extract was applied into a 5-mL HiTrap Chelating HP column (GE Healthcare, Little Chalfont, UK), previously charged with Ni<sup>2+</sup>, coupled to an ÄKTA purifier (GE Healthcare, Little Chalfont, UK), at a flow rate of 2 mL min<sup>-1</sup>. The target proteins were eluted using a non-linear (0–0.5 M) gradient of imidazole. The fractions containing pure proteins were pooled, concentrated and applied into a HiLoad 16/600 Superdex 200 pg column (GE Healthcare, Little Chalfont, UK), previously equilibrated with 20 mM sodium phosphate, pH 7.5, and 150 mM NaCl, coupled to an ÄKTA purifier (GE Healthcare, Little Chalfont, UK) at a flow rate of 1 mL min<sup>-1</sup>.

### Enzyme activity assays monitored by colorimetric methods

Arabinose, pNP and 4-nitrophenyl-α-L-arabinofuranoside (pNP-Araf) were purchased from Sigma-Aldrich, St. Louis, MO. Wheat flour arabinoxylan, sugar beet arabinan and AXOS were purchased from Megazyme, Co. Wicklow, IE.

For the thermotolerance assay, *XacAbf51* was incubated at 50 °C for up to 72 h and samples were collected to measure activity against pNP-Araf in McIlvaine buffer (pH 5.5) at 50 °C, for 10 min, using 0.5 µg mL<sup>-1</sup> (9 nM) enzyme and 10 mM substrate. The



generation of *p*-nitrophenolate from pNP-conjugated monosaccharides was monitored at  $A_{400\text{nm}}$  ( $\epsilon_{400\text{nm}, \text{pH } 12} = 17,500 \text{ mol}^{-1} \text{ L cm}^{-1}$ ). Activity against arabinoxylan at  $10 \text{ mg mL}^{-1}$  was measured in McIlvaine buffer (pH 5.5) at  $50^\circ\text{C}$ , for 10 min, using  $15 \text{ }\mu\text{g mL}^{-1}$  *XacAbf51* (263 nM) or  $16 \text{ }\mu\text{g mL}^{-1}$  (278 nM) *TxAbfD3* and the generation of arabinose from polysaccharides was determined by the 3,5-dinitrosalicylic acid (DNS) method [45]. To determine the kinetic properties of *XacAbf51* and *TxAbfD3*, the reactions were performed in McIlvaine buffer (pH 5.5) at  $50^\circ\text{C}$ , for 10 min, in the range from  $7 \text{ }\mu\text{M}$  to  $14 \text{ mM}$  of pNP-Araf using  $0.5 \text{ }\mu\text{g mL}^{-1}$  (9 nM) *XacAbf51* or  $0.1 \text{ }\mu\text{g mL}^{-1}$  (2 nM) *TxAbfD3* and from  $0.3$  to  $140 \text{ mg mL}^{-1}$  of arabinan using  $263 \text{ nM}$  *XacAbf51* or  $278 \text{ nM}$  *TxAbfD3*. The kinetic parameters were calculated by non-linear regression analysis of the Michaelis–Menten plot using the program OriginPro 8.1.

#### Enzyme activity assays monitored by mass spectrometry

To estimate initial rates, the reactions were made in triplicate with  $5 \text{ }\mu\text{L}$  of oligosaccharide ( $\text{XA}^3\text{XX}$  or  $\text{XA}^{2+3}\text{XX}$ ) in different concentrations ( $0.2$ – $100 \text{ mM}$ ),  $3 \text{ }\mu\text{L}$  of McIlvaine buffer at pH 5.5,  $1 \text{ }\mu\text{L}$  of water and  $1 \text{ }\mu\text{L}$  of *XacAbf51* stock at  $150 \text{ }\mu\text{g mL}^{-1}$  (final enzyme concentration =  $15 \text{ }\mu\text{g mL}^{-1}$ ). After 5 min of incubation at  $50^\circ\text{C}$ ,  $700 \text{ rpm}$ ,  $40 \text{ }\mu\text{L}$  of methanol was added to quench the reaction. Assays to compare the relative activity of *XacAbf51* and *TxAbfD3* on  $2^3,3^3$ -di- $\alpha$ -L-arabinofuranosyl-xylotetraose ( $\text{XA}^{2+3}\text{XX}$ ) or  $3^3$ - $\alpha$ -L-arabinofuranosyl-xylotetraose ( $\text{XA}^3\text{XX}$ ) were performed incubating *XacAbf51* ( $15 \text{ }\mu\text{g mL}^{-1}$ ) or *TxAbfD3* ( $15 \text{ }\mu\text{g mL}^{-1}$ ) with  $10 \text{ mM}$  substrate, at pH 5.5,  $50^\circ\text{C}$ ,  $700 \text{ rpm}$ , 5 min (reaction volume =  $10 \text{ }\mu\text{L}$ ) and quenched with  $40 \text{ }\mu\text{L}$  of methanol in triplicate.

The kinetic assays were monitored on a Waters Synapt HDMS, at *V* mode and ESI(+) with a spray voltage maintained at  $3.0 \text{ kV}$  and heated to  $130^\circ\text{C}$  in the source. A total of  $15 \text{ }\mu\text{L}$  of the quenched reactions and  $2 \text{ }\mu\text{L}$  of  $1 \text{ mM}$  xylotriose (used as the internal standard) were added to  $183 \text{ }\mu\text{L}$  of water and injected into the mass spectrometer in scan mode ( $m/z$   $300$ – $900$ ) with direct infusion at a flow rate of  $50 \text{ }\mu\text{L min}^{-1}$ . An internal standard with ionization similar to analytes (xylotriose) was used to increase the reliability of the method [46]. A calibration curve was made to determine the concentrations of the products of the enzymatic reaction. The kinetic parameters of the reactions ( $k_{\text{cat}}$ ,  $K_{\text{m}}$  and  $V_{\text{max}}$ ) were determined by non-linear regression analysis (Hill model) of the Michaelis–Menten plot using the software Origin8.1.

#### Capillary zone electrophoresis

For capillary zone electrophoresis analysis, reactions were incubated for 60 min at  $50^\circ\text{C}$ , pH 5.5, using  $100 \text{ }\mu\text{g mL}^{-1}$  *XacAbf51* and  $10 \text{ mM}$  AXOS, or  $20 \text{ mg mL}^{-1}$  arabinan, or  $20 \text{ mg mL}^{-1}$  arabinoxylan. Arabinose, xylobiose, xylotriose and xylotetraose ( $5 \text{ mM}$ ) and reactions without enzyme were used as standard references. Samples were heated at  $95^\circ\text{C}$ , for 5 min, centrifuged and  $30 \text{ }\mu\text{L}$  of the supernatant was dried using Speed-Vac (Thermo Fisher Scientific, Waltham, MA). Reaction products were derivatized with 8-aminopyrene-1,3,6-trisulfonic acid (APTS) [47] by incubation with  $20 \text{ }\mu\text{L}$   $2.5 \text{ M}$  citric acid,  $8 \text{ }\mu\text{L}$   $1 \text{ M}$  sodium cyanoborohydride (in THF) and  $1.5 \text{ }\mu\text{L}$   $100 \text{ mM}$  APTS (in  $25\%$  ( $v/v$ ) acetic acid), for 2 h at  $75^\circ\text{C}$ . Labeled reactions ( $4 \text{ }\mu\text{L}$ ) or negative controls without enzyme ( $2 \text{ }\mu\text{L}$ ) were diluted to a final volume of  $60 \text{ }\mu\text{L}$  using run buffer ( $0.04 \text{ M}$  potassium phosphate, pH 2.5) and injected into an uncoated fused-silica capillary of  $75 \text{ }\mu\text{m}$  internal diameter and  $20 \text{ cm}$  effective length (Beckman Coulter, Brea, CA), by application of  $0.5 \text{ psi}$ , for 5 s, using a P/ACE MDQ instrument configured with a laser-induced fluorescence detection system (Beckman Coulter, Brea, CA). Electrophoretic conditions were  $20 \text{ kV}/70$ – $100 \text{ mA}$  with reverse polarity at a controlled temperature of  $25^\circ\text{C}$ . Carbohydrates labeled with APTS were excited at  $488 \text{ nm}$  and emission was collected through a  $520\text{-nm}$  band pass filter.

#### Preparation of enzyme cocktail from *T. reesei* Rut-C30

The strain *T. reesei* Rut-C30 was cultivated on Petri dishes containing potato dextrose agar (Difco, MI). After 6–10 days, spores were collected in spore solution ( $20\%$  ( $v/v$ ) glycerol,  $0.8\%$  ( $m/v$ ) NaCl,  $0.025\%$  ( $v/v$ ) Tween 20), filtered through sterile cotton, quantified using a hemocytometer and frozen at  $-80^\circ\text{C}$  for long-term storage.

Fermentations were performed using the BioFlo/CelliGen 115 system (Eppendorf, Hamburg, Germany) and water-jacketed  $3.0\text{-L}$  vessels. The fermentation medium comprised of  $5\%$  ( $m/v$ ) milled soybean hulls,  $5\%$  ( $v/v$ ) milk whey,  $2\%$  ( $m/v$ )  $(\text{NH}_4)_2\text{SO}_4$  and  $1 \text{ mL L}^{-1}$  of J647 antifoam (Struktol, Hamburg, Germany) in the batch phase and milk whey with lactose concentration of  $177 \text{ g L}^{-1}$  were fed from 72 to 170 h at an average rate of  $0.5 \text{ g L}^{-1} \text{ h}^{-1}$  total sugar. Aeration was maintained at  $1.0 \text{ VVM}$  compressed air, pH between 3.8 and 4.8 using  $2 \text{ M}$  phosphoric acid and  $10\%$  ammonia, and DO above  $30\%$  with an agitation cascade ( $400$ – $950 \text{ rpm}$ ). The initial volume was  $1 \text{ L}$ , and the reactors were inoculated with  $1:10$  volume of 7-day-old shake flask preculture using the same media composition as the fermentation batch medium, spore concentration in the inoculum bottle was  $2.5 \times 10^7$  in  $100 \text{ mL}$ . Samples were withdrawn every 24 h,

centrifuged at  $21,000\times g$  for 10 min and the supernatants stored at  $-20\text{ }^{\circ}\text{C}$  for analysis. Whole broth samples were adjusted to pH 5.0, frozen at  $-20\text{ }^{\circ}\text{C}$  and used for hydrolysis assays. Fermentations were terminated after 170 h when the feeding was stopped.

For quantifying protein, the sample was first diluted to a final concentration of  $0.3\text{--}1.5\text{ g L}^{-1}$  in 50 mM Na citrate buffer, pH 5.0. A 200  $\mu\text{L}$  sample was combined with 800  $\mu\text{L}$  ice-cold acetone, mixed by inverting the tube several times and then maintained at  $-20\text{ }^{\circ}\text{C}$  for 1 h. The precipitated proteins were pelleted by centrifugation at  $14,000\times g$  and  $4\text{ }^{\circ}\text{C}$  for 5 min. The supernatant was removed and the pellet was air-dried for 5 min before resuspending in the original volume (200  $\mu\text{L}$ ) of buffer. The protein concentration was then quantified using the DC protein kit (BioRad, Hercules, CA) based on the method of Lowry [48] using bovine serum albumin as standard.

#### Complementation assays

Delignified sugarcane bagasse was prepared using an alkaline pretreatment ( $130\text{ }^{\circ}\text{C}$ , 30 min,  $1.5\%\text{ m v}^{-1}$  NaOH), yielding a material composed by 58.6% cellulose, 22.1% hemicellulose, and 8.8% lignin. Enzymatic hydrolysis reactions were performed in samples of 1 mL containing 5% of dry biomass (50 mg) and 237.5  $\mu\text{g}$  of enzyme cocktail, supplemented or not with 12.5  $\mu\text{g}$  *XacAbf51*, in buffer 50 mM sodium citrate, pH 5.5, with 0.02% sodium azide. The reactions were done in triplicate and incubated in a hybridization oven at  $50\text{ }^{\circ}\text{C}$  with agitation during 24 h. The enzyme cocktails used were Celluclast (Novozymes, Krogshoejvej, Denmark) and the whole broth from *T. reesei* RUT-C30, prepared as described above. Protein concentration was estimated by the Lowry method [48] using the DC protein kit (BioRad, Hercules, CA).

#### Circular dichroism

CD spectra were acquired on a JASCO J-815 CD spectrometer (Jasco, Tokyo, Japan) controlled by a CDF-426S/15 Peltier temperature control system using a quartz cuvette with a 1-cm path length. The enzyme was prepared in phosphate buffer (20 mM sodium phosphate, 150 mM NaCl, pH 7.5) at a final concentration of 8  $\mu\text{M}$ . All spectra were obtained at  $20\text{ }^{\circ}\text{C}$  in the range 195–260 nm with a bandwidth of 2 nm and a response time of  $4\text{ s nm}^{-1}$ . CD spectra were buffer subtracted and normalized to mean residue ellipticity. Thermal unfolding experiments were monitored at 220 nm in the temperature range  $20\text{--}90\text{ }^{\circ}\text{C}$  with a scan rate of  $1\text{ }^{\circ}\text{C min}^{-1}$ . The melting temperature was determined according to the sigmoidal-Boltzmann fitting of the CD denaturation curve.

#### Differential scanning calorimetry

Thermal stability was also analyzed by DSC using a VP-DSC device (Microcal, GE Healthcare, Northampton, MA). The enzyme was prepared in phosphate buffer (20 mM sodium phosphate, 150 mM NaCl, pH 7.5) at a final concentration of  $2\text{ mg mL}^{-1}$ . A temperature rate of  $1\text{ }^{\circ}\text{C min}^{-1}$  was used and the reversibility of protein denaturation was tested. Denaturation curves were buffer subtracted, concentration normalized and the resultant endotherms integrated following assignment of pre- and post-transition baselines.

#### Dynamic light scattering

Size distribution of the purified enzyme in solution was evaluated using DLS. Measurements were acquired at  $20\text{ }^{\circ}\text{C}$  on a Malvern Zetasizer Nano ZS 90 (Model no. ZEN3690, Malvern, Worcestershire, UK) with a 633-nm laser, in a quartz cell with a scattering angle of  $90^{\circ}$ . The protein was analyzed at a concentration of  $0.5\text{ mg mL}^{-1}$  in phosphate buffer (20 mM sodium phosphate, 150 mM NaCl, pH 7.5). An average of 20 runs was used to estimate the  $R_h$  through Stokes–Einstein equation.

#### Analytical ultracentrifugation

Sedimentation velocity experiments were performed on a Beckman Optima XL-A analytical ultracentrifuge (Beckman Coulter, Indianapolis, IN) at  $20\text{ }^{\circ}\text{C}$ . Spectra were collected at both 220 and 280 nm. The protein was prepared in different concentrations ranging from 0.2 to  $0.9\text{ mg mL}^{-1}$  in phosphate buffer (20 mM sodium phosphate, 150 mM NaCl, pH 7.5). AUC data were analyzed using the continuous sedimentation distribution method in the SEDFIT program [49]. The  $s_{020,w}$  value at infinite dilution was calculated by linear regression of  $s_{20,w}$  as a function of protein concentration.

#### Small angle X-ray scattering

Small angle X-ray scattering measurements were performed at three different concentrations (2, 4 and  $6\text{ mg mL}^{-1}$ ) in 20 mM Tris buffer, pH 7.5. Data were collected at SAXS2 beamline (LNLS, Campinas, Brazil), integrated using Fit2D [50] and analyzed using GNOM [51]. The molecular envelope was calculated from the experimental SAXS data using the program DAMMIN [52]. Ten runs of ab initio shape determination yielded highly similar models (normalized spatial discrepancy values  $<1$ ), which were then averaged using the package DAMAVER [53]. The theoretical scattering curves of crystallographic structures were calculated and compared with the experimental SAXS curves using the program CRY SOL [54]. The crystallographic structure

was fitted into the SAXS molecular envelope using the program SUPCOMB [55].

#### Protein crystallization, X-ray data collection and structure determination

*XacAbf51* (27 mg mL<sup>-1</sup>) crystallized by vapor diffusion method in sitting drops containing 17% (w v<sup>-1</sup>) polyethylene glycol 3350 and 0.2 M ammonium chloride. Crystals were cryoprotected using the reservoir solution added of 20% (v v<sup>-1</sup>) glycerol. Diffraction data were collected at the BL12-2 beamline from the Stanford Synchrotron Radiation Lightsources (Stanford, CA). Data were processed using XDS [56] and the structure was solved by molecular replacement method using the program MOLREP and the atomic coordinates of *TxAbfD3* (PDB ID: 2VRQ) as search model. Six chains were found in the asymmetric unit and the model was refined against electron density using COOT [57] and against X-ray data using phenix.refine [58] and REFMAC [59]. Final model was validated using MolProbity [60]. Data collection, processing and refinement statistics are summarized in Table 2.

#### Molecular dynamics simulation

Superimposition of *XacAbf51* crystal structure with that of *TxAbfD3* in complex with 3<sup>2</sup>- $\alpha$ -L-arabinofuranosyl-xylotriose (XA<sup>3</sup>X) was performed with PDBEOLD [61] and the coordinates of the ligand positioned into the *XacAbf51* active site were transferred for the PDB file containing one trimer of *XacAbf51*. An Araf substitution was added to XA<sup>3</sup>X to generate XA<sup>2+3</sup>X and simulation systems using explicit solvent were created for energy-minimized trimeric structures of *XacAbf51* in complex with XA<sup>2+3</sup>X. Energy minimization and MD simulations were carried out using YAMBER3 force field with the program YASARA [62]. Long-range Coulomb interactions were included with a cutoff of 7.86 Å. The simulation box was defined at 15 Å around all atoms of the structure. Protonation was performed at pH 7. Cell neutralization was reached filling the box with water molecules ( $d = 0.997$  g mL<sup>-1</sup>) and Na/Cl counter ions (0.9% m v<sup>-1</sup>) coupled with a short MD simulation for solvent relaxation. MD simulations were performed during 100 ns at 298 K, using a multiple time step of 2.0 fs for inter-molecular forces, 1.2 fs for intra-molecular forces, periodic boundary conditions and unconstrained bonds and angles. Root mean square deviations (RMSDs) were calculated for the whole system and Euclidean distances between enzyme and substrate atoms were measured through the trajectory in the three active sites of the trimer and the average value is presented in function of the simulation time.

**Table 2 Data collection and refinement statistics of *XacAbf51* crystal structure**

Data collection	
Space group	<i>P</i> 2 <sub>1</sub>
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	91.34, 163.23, 114.42
$\beta$ (°)	107.57
Resolution (Å)	48.74–1.91 (2.03–1.91)
<i>R</i> <sub>meas</sub>	0.113 (0.639)
<i>I</i> / $\sigma$ <i>I</i>	9.74 (2.06)
Completeness (%)	95.1 (80.3)
Redundancy	3.4 (2.8)
CC 1/2	99.6 (72.9)
Refinement	
Resolution (Å)	48.74–1.91
No. of reflections	234,093
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.149/0.167
Protein residues	2964 (6 chains)
Ligands	6 GOL
Water	2384
<i>B</i> -factors (Å <sup>2</sup> )	
Average	24.64
Protein	24.00
Ligand	23.05
Water	30.75
R.m.s. deviations	
Bond lengths (Å)	0.0116
Bond angles (°)	1.5092
Ramachandran <sup>a</sup>	
Favored (%)	97
Allowed (%)	3
Outliers (%)	0
MolProbity score	1.00
PDB code	6D25

Values in parentheses are for highest resolution shell

<sup>a</sup> Ramachandran data were calculated using MolProbity

#### Phylogenetic analyses

The sequences of characterized GH51 enzymes present in the CAZY database, excluding redundant sequences (sequences from the same species with >95% sequence identity) and synthetic constructs, were manually edited to include only the fragment corresponding to the ( $\beta/\alpha$ )<sub>8</sub> barrel, as predicted by the webserver SUPERFAMILY [63]. The edited sequences were aligned using the software MUSCLE, available at the EMBL-EBI webserver (<https://www.ebi.ac.uk/Tools/msa/muscle/>) [64]. The multiple sequence alignment was provided for the MEGA7 software to perform evolutionary analyses [65]. Initial tree(s) for the heuristic search were obtained automatically by applying

Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites [five categories (+G, parameter = 1.9905)]. The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 1.20% sites). The analysis involved 72 amino acid sequences. All positions with less than 80% site coverage were eliminated. That is, fewer than 20% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 292 positions in the final dataset. The confidence of tree topology was assessed using the Bootstrap analysis based on 1000 bootstrap replications [66].

#### Abbreviations

Abf:  $\alpha$ -L-arabinofuranosidase; Araf: arabinofuranosyl; AUC: analytical ultracentrifugation; AXOS: arabinoxylooligosaccharides; *Bl*: *Bifidobacterium longum*; CD: circular dichroism; DSC: differential scanning calorimetry; DLS: dynamic light scattering; *Gs*: *Geobacillus stearothermophilus*; GH: glycoside hydrolase; *Hi*: *Humicola insolens*; MD: molecular dynamics; pNP: 4-nitrophenyl; NR: non-reducing end; R: reducing end;  $R_h$ : hydrodynamic radius; *Rt*: *Ruminiclostridium thermocellum*; SAXS: small angle X-ray scattering; *Tx*: *Thermobacillus xylanilyticus*;  $T_m$ : melting temperature; *Tm*: *Thermotoga maritima*; *Tp*: *Thermotoga petrophila*; *Xac*: *Xanthomonas axonopodis* pv. *citri*; *Xylp*: xylopyranosyl.

#### Authors' contributions

CRS and MTM designed the study. CRS, POG and MTM wrote the paper. POG made the figures and the phylogenetic analyses. VPMM and FB constructed vectors for expression and established the heterologous expression protocol. CCCT purified and crystallized the protein. MABM determined and refined the structure. RVH and PSLO performed the molecular dynamics simulations. FHMS and MND characterized the enzyme activity by colorimetric methods. LMF produced the *T. reesei* RUT-C30 enzyme cocktail. MND performed the enzyme complementation assays. RASP and FCG carried out the mass spectrometry analyses. LMZ characterized the biophysical properties of the protein. All authors analyzed the results of the manuscript. All authors read and approved the final manuscript.

#### Author details

<sup>1</sup> Brazilian Bioethanol Science and Technology Laboratory (CTBE), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, Sao Paulo 13083-970, Brazil. <sup>2</sup> Brazilian Biosciences National Laboratory (LNBio), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, Sao Paulo 13083-970, Brazil. <sup>3</sup> Dalton Mass Spectrometry Laboratory, Institute of Chemistry, University of Campinas, Campinas, Sao Paulo 13083-861, Brazil.

#### Acknowledgements

We are thankful to Brazilian Synchrotron Light Laboratory (LNLS) and Brazilian Biosciences National Laboratory (LNBio) for the provision of time on the SAXS2 beamline, and both crystallization (Robolab) and spectroscopy (LEC) facilities. We are also thankful to Stanford Synchrotron Radiation Lightsource (SSRL) for the provision of time on the BL12-2 beamline.

#### Competing interests

The authors declare that they have no competing interests.

#### Availability of data and materials

The atomic coordinates for the crystal structure of *XacAbf51* were deposited at the Research Collaboratory for Structural Bioinformatics Protein Databank (<http://www.rcsb.org/>) under PDB ID 6D25.

#### Consent for publication

Not applicable.

#### Ethics approval and consent to participate

Not applicable.

#### Funding

This research was supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) [Grants #2013/13309-0, #2014/07135-1, #2014/17264-3 and #2015/26982-0], Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 25 April 2018 Accepted: 23 July 2018

Published online: 11 August 2018

#### References

- Dervilly-Pinel G, Tran V, Saulnier L. Investigation of the distribution of arabinose residues on the xylan backbone of water-soluble arabinoxylans from wheat flour. *Carbohydr Polym*. 2004;55:171–7.
- Scheller HV, Ulvskov P. Hemicelluloses. *Annu Rev Plant Biol*. 2010;61:263–89.
- de Souza AP, Leite DCC, Pattathil S, Hahn MG, Buckeridge MS. Composition and structure of sugarcane cell wall polysaccharides: implications for second-generation bioethanol production. *BioEnergy Res*. 2013;6:564–79.
- Van Dongen FEM, Van Eylen D, Kabel MA. Characterization of substituents in xylans from corn cobs and stover. *Carbohydr Polym*. 2011;86:722–31.
- Riviere A, Moens F, Selak M, Maes D, Weckx S, De Vuyst L. The ability of bifidobacteria to degrade arabinoxylan oligosaccharide constituents and derived oligosaccharides is strain dependent. *Appl Environ Microbiol*. 2014;80:204–17.
- Broekaert WF, Courtin CM, Verbeke K, Van de Wiele T, Verstraete W, Delcour JA. Prebiotic and other health-related effects of cereal-derived arabinoxylans, arabinoxylan-oligosaccharides, and xylooligosaccharides. *Crit Rev Food Sci Nutr*. 2011;51:178–94.
- Henrissat B, Davies G. Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol*. 1997;7:637–44.
- Pell G, Taylor EJ, Gloster TM, Turkenburg JP, Fontes CM, Ferreira LM, Nagy T, Clark SJ, Davies GJ, Gilbert HJ. The mechanisms by which family 10 glycoside hydrolases bind decorated substrates. *J Biol Chem*. 2004;279:9597–605.
- Vandermaerliere E, Bourgois TM, Rombouts S, Van Campenhout S, Volckaert G, Strelkov SV, Delcour JA, Rabijns A, Courtin CM. Crystallographic analysis shows substrate binding at the –3 to +1 active-site subsites and at the surface of glycoside hydrolase family 11 endo-1,4-beta-xylanases. *Biochem J*. 2008;410:71–9.
- Alvira P, Negro MJ, Ballesteros M. Effect of endoxylanase and alpha-L-arabinofuranosidase supplementation on the enzymatic hydrolysis of steam exploded wheat straw. *Bioresour Technol*. 2011;102:4552–8.
- Wilkens C, Andersen S, Dumon C, Berrin JG, Svensson B. GH62 arabinofuranosidases: structure, function and applications. *Biotechnol Adv*. 2017;35:792–804.
- Bourgois TM, Van Craeyveld V, Van Campenhout S, Courtin CM, Delcour JA, Robben J, Volckaert G. Recombinant expression and characterization of XynD from *Bacillus subtilis* subsp. *subtilis* ATCC 6051: a GH 43 arabinoxylan arabinofuranohydrolase. *Appl Microbiol Biotechnol*. 2007;75:1309–17.
- Lagaert S, Pollet A, Delcour JA, Lavigne R, Courtin CM, Volckaert G. Substrate specificity of three recombinant alpha-L-arabinofuranosidases from *Bifidobacterium adolescentis* and their divergent action on arabinoxylan and arabinoxylan oligosaccharides. *Biochem Biophys Res Commun*. 2010;402:644–50.
- McKee LS, Pena MJ, Rogowski A, Jackson A, Lewis RJ, York WS, Krogh KB, Vikso-Nielsen A, Skjot M, Gilbert HJ, et al. Introducing endo-xylanase activity into an exo-acting arabinofuranosidase that targets side chains. *Proc Natl Acad Sci USA*. 2012;109:6537–42.



15. van den Broek LA, Lloyd RM, Beldman G, Verdoes JC, McCleary BV, Voragen AG. Cloning and characterization of arabinoxylan arabinofuranohydrolase-D3 (AXHd3) from *Bifidobacterium adolescentis* DSM20083. *Appl Microbiol Biotechnol*. 2005;67:641–7.
16. Sorensen HR, Jorgensen CT, Hansen CH, Jorgensen CI, Pedersen S, Meyer AS. A novel GH43 alpha-L-arabinofuranosidase from *Humicola insolens*: mode of action and synergy with GH51 alpha-L-arabinofuranosidases on wheat arabinoxylan. *Appl Microbiol Biotechnol*. 2006;73:850–61.
17. Ferre H, Broberg A, Duus JO, Thomsen KK. A novel type of arabinoxylan arabinofuranohydrolase isolated from germinated barley analysis of substrate preference and specificity by nano-probe NMR. *Eur J Biochem FEBS*. 2000;267:6633–41.
18. Borsenberger V, Dornez E, Desrousseaux ML, Massou S, Tenkanen M, Courtin CM, Dumon C, O'Donohue MJ, Faure R. A (1)H NMR study of the specificity of alpha-L-arabinofuranosidases on natural and unnatural substrates. *Biochem Biophys Acta*. 2014;1840:3106–14.
19. Koutaniemi S, Tenkanen M. Action of three GH51 and one GH54 alpha-arabinofuranosidases on internally and terminally located arabinofuranosyl branches. *J Biotechnol*. 2016;229:22–30.
20. Beylot MH, McKie VA, Voragen AG, Doeswijk-Voragen CH, Gilbert HJ. The *Pseudomonas cellulosa* glycoside hydrolase family 51 arabinofuranosidase exhibits wide substrate specificity. *Biochem J*. 2001;358:607–14.
21. Borsenberger V, Dornez E, Desrousseaux M-L, Massou S, Tenkanen M, Courtin CM, Dumon C, O'Donohue MJ, Fauré R. A 1H NMR study of the specificity of alpha-L-arabinofuranosidases on natural and unnatural substrates. *Biochim Biophys Acta Gen Subj*. 2014;1840:3106–14.
22. Lee RC, Burton RA, Hrmova M, Fincher GB. Barley arabinoxylan arabinofuranohydrolases: purification, characterization and determination of primary structures from cDNA clones. *Biochem J*. 2001;356:181–9.
23. Santos CR, Hoffmam ZB, de Matos Martins VP, Zanphorlin LM, de Paula Assis LH, Honorato RV, Lopes de Oliveira PS, Ruller R, Murakami MT. Molecular mechanisms associated with xylan degradation by *Xanthomonas* plant pathogens. *J Biol Chem*. 2014;289:32186–200.
24. Debeche T, Cummings N, Connerton I, Debeire P, O'Donohue MJ. Genetic and biochemical characterization of a highly thermostable alpha-L-arabinofuranosidase from *Thermobacillus xylanilyticus*. *Appl Environ Microbiol*. 2000;66:1734–6.
25. Debeche T, Bliard C, Debeire P, O'Donohue MJ. Probing the catalytically essential residues of the alpha-L-arabinofuranosidase from *Thermobacillus xylanilyticus*. *Protein Eng*. 2002;15:21–8.
26. Shallom D, Belakhov V, Solomon D, Shoham G, Baasov T, Shoham Y. Detailed kinetic analysis and identification of the nucleophile in alpha-L-arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. *J Biol Chem*. 2002;277:43667–73.
27. Shallom D, Belakhov V, Solomon D, Gilead-Gropper S, Baasov T, Shoham G, Shoham Y. The identification of the acid-base catalyst of alpha-L-arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. *FEBS Lett*. 2002;514:163–7.
28. Paes G, Skov LK, O'Donohue MJ, Remond C, Kastrup JS, Gajhede M, Mirza O. The structure of the complex between a branched pentasaccharide and *Thermobacillus xylanilyticus* GH-51 arabinofuranosidase reveals xylan-binding determinants and induced fit. *Biochemistry*. 2008;47:7441–51.
29. Krissinel E. Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Res*. 2015;43:W314–9.
30. Winker S, Woese CR. A definition of the domains archaea, bacteria and eucarya in terms of small subunit ribosomal RNA characteristics. *Syst Appl Microbiol*. 1991;14:305–10.
31. Matsuo N, Kaneko S, Kuno A, Kobayashi H, Kusakabe I. Purification, characterization and gene cloning of two alpha-L-arabinofuranosidases from streptomyces chartreusis GS901. *Biochem J*. 2000;346(Pt 1):9–15.
32. Sakamoto T, Inui M, Yasui K, Hosokawa S, Ihara H. Substrate specificity and gene expression of two *Penicillium chrysogenum* alpha-L-arabinofuranosidases (AFQ1 and AFS1) belonging to glycoside hydrolase families 51 and 54. *Appl Microbiol Biotechnol*. 2013;97:1121–30.
33. Inacio JM, Correia IL, de Sa-Nogueira I. Two distinct arabinofuranosidases contribute to arabinoo-oligosaccharide degradation in *Bacillus subtilis*. *Microbiology*. 2008;154:2719–29.
34. Kosugi A, Murashima K, Doi RH. Characterization of two noncellulosomal subunits, ArfA and BgaA, from *Clostridium cellulovorans* that cooperate with the cellulosome in plant cell wall degradation. *J Bacteriol*. 2002;184:6859–65.
35. Gilead S, Shoham Y. Purification and characterization of alpha-L-arabinofuranosidase from *Bacillus stearothermophilus* T-6. *Appl Environ Microbiol*. 1995;61:170–4.
36. Moon YH, Iakviak M, Bauer S, Mackie RI, Cann IK. Biochemical analyses of multiple endoxylanases from the rumen bacterium *Ruminococcus albus* 8 and their synergistic activities with accessory hemicellulose-degrading enzymes. *Appl Environ Microbiol*. 2011;77:5157–69.
37. Fritz M, Ravanal MC, Braet C, Eyzaguirre J. A family 51 alpha-L-arabinofuranosidase from *Penicillium purpurogenum*: purification, properties and amino acid sequence. *Mycol Res*. 2008;112:933–42.
38. Mechelke M, Herlet J, Benz JP, Schwarz WH, Zverlov VV, Liebl W, Kornberger P. HPAEC-PAD for oligosaccharide analysis-novel insights into analyte sensitivity and response stability. *Anal Bioanal Chem*. 2017;409:7169–81.
39. Janis J, Pulkkinen P, Rouvinen J, Vainiotalo P. Determination of steady-state kinetic parameters for a xylanase-catalyzed hydrolysis of neutral underivatized xylooligosaccharides by mass spectrometry. *Anal Biochem*. 2007;365:165–73.
40. Gao D, Uppugundla N, Chundawat SPS, Yu X, Hermanson S, Gowda K, Brumm P, Mead D, Balan V, Dale BE. Hemicellulases and auxiliary enzymes for improved conversion of lignocellulosic biomass to monosaccharides. *Biotechnol Biofuels*. 2011;4:5.
41. Jourdirier E, Baudry L, Poggi-Parodi D, Vicq Y, Koszul R, Margeot A, Marbouty M, Bidard F. Proximity ligation scaffolding and comparison of two *Trichoderma reesei* strains genomes. *Biotechnol Biofuels*. 2017;10:151.
42. Maehara T, Fujimoto Z, Ichinose H, Michikawa M, Harazono K, Kaneko S. Crystal structure and characterization of the glycoside hydrolase family 62 alpha-L-arabinofuranosidase from *Streptomyces coelicolor*. *J Biol Chem*. 2014;289:7962–72.
43. Souza TA, Santos CR, Souza AR, Oldiges DP, Ruller R, Prade RA, Squina FM, Murakami MT. Structure of a novel thermostable GH51 alpha-L-arabinofuranosidase from *Thermotoga petrophila* RKU-1. *Protein Sci*. 2011;20:1632–7.
44. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res*. 2005;33:W299–302.
45. Miller GL. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Chem*. 1959;31:426–8.
46. Ge X, Sirich TL, Beyer MK, Desaire H, Leary JA. A strategy for the determination of enzyme kinetics using electrospray ionization with an ion trap mass spectrometer. *Anal Chem*. 2001;73:5078–82.
47. Chen FT, Evangelista RA. Analysis of mono- and oligosaccharide isomers derivatized with 9-aminopyrene-1,4,6-trisulfonate by capillary electrophoresis with laser-induced fluorescence. *Anal Biochem*. 1995;230:273–80.
48. Lowry OH, Rosebrough NJ, Farr AL, Randall RJ. Protein measurement with the Folin phenol reagent. *J Biol Chem*. 1951;193:265–75.
49. Schuck P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys J*. 2000;78:1606–19.
50. Hammersley AP, Brown K, Burmeister W, Claustre L, Gonzalez A, McSweeney S, Mitchell E, Moy JP, Svensson SO, Thompson AW. Calibration and application of an X-ray image intensifier/charge-coupled device detector for monochromatic macromolecular crystallography. *J Synchrotron Radiat*. 1997;4:67–77.
51. Svergun D. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr*. 1992;25:495–503.
52. Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J*. 1999;76:2879–86.
53. Volkov VV, Svergun DI. Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr*. 2003;36:860–4.
54. Svergun D, Barberato C, Koch MHJ. CRYSOLO—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr*. 1995;28:768–73.
55. Kozin MB, Svergun DI. Automated matching of high- and low-resolution structural models. *J Appl Crystallogr*. 2001;34:33–41.
56. Kabsch W. Xds. *Acta Crystallogr D Biol Crystallogr*. 2010;66:125–32.
57. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. 2010;66:486–501.



58. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr*. 2012;68:352–67.
59. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*. 1997;53:240–55.
60. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. 2010;66:12–21.
61. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr*. 2004;60:2256–68.
62. Krieger E, Koraimann G, Vriend G. Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins*. 2002;47:393–402.
63. Gough J, Karplus K, Hughey R, Chothia C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol*. 2001;313:903–19.
64. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R. Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res*. 2013;41:W597–600.
65. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33:1870–4.
66. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 1985;39:783–91.
67. Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol*. 2008;25:1307–20.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

