Two distinct catalytic pathways for GH43 xylanolytic enzymes unveiled by X-ray and QM/MM simulations

Mariana A. B. Morais\textsuperscript{1,2}, Joan Coines\textsuperscript{2}, Mariane N. Domingues\textsuperscript{1}, Renan A. S. Pirolla\textsuperscript{1}, Celisa C. C. Tonoli\textsuperscript{3}, Camila R. Santos\textsuperscript{1}, Jessica B. L. Correa\textsuperscript{1}, Fabio C. Gozzo\textsuperscript{4}, Carme Rovira\textsuperscript{2,5} & Mario T. Murakami\textsuperscript{1,5} 

Xylanolytic enzymes from glycoside hydrolase family 43 (GH43) are involved in the breakdown of hemicellulose, the second most abundant carbohydrate in plants. Here, we kinetically and mechanistically describe the non-reducing-end xylose-releasing exo-oligoxylanase activity and report the crystal structure of a native GH43 Michaelis complex with its substrate prior to hydrolysis. Two distinct calcium-stabilized conformations of the active site xylosyl unit are found, suggesting two alternative catalytic routes. These results are confirmed by QM/MM simulations that unveil the complete hydrolysis mechanism and identify two possible reaction pathways, involving different transition state conformations for the cleavage of xyloooligosaccharides. Such catalytic conformational promiscuity in glycosidases is related to the open architecture of the active site and thus might be extended to other exo-acting enzymes. These findings expand the current general model of catalytic mechanism of glycosidases, a main reaction in nature, and impact on our understanding about their interaction with substrates and inhibitors.
Hemicellulose, a main component of plant cell walls, is one of the most abundant and complex carbohydrates in nature, composed of several sugars including mostly β-xylose and α-arabinose. Microorganisms mastering plant cell wall degradation are provided with multiple genes with broad action on hemicellulose, commonly encoding several glycoside hydrolases from family 43 (GH43)\(^1\). They constitute one of the largest GH families annotated in the CAZy database\(^2\) and are divided into 37 subfamilies that diverge in terms of structure and function\(^3\). A significant fraction of these subfamilies remains partially characterized without structural data, complete functional profile or even fully unexplored. In addition, due to their atypical diverse taxallographic data, some works suggested the importance of catalytic mechanisms at molecular detail, which have been so far not elucidated. The calcium was found to be responsible for improving the catalytic activity of GH43 enzymes (arabinanases, xylosidases, and arabinoxyranosidases)\(^14\). They are observed in several crystal structures in a hydrated cavity located near the −1 subsite, but not directly interacting with the substrate\(^12\). However, the molecular basis behind the role of calcium in GH43 enzymes remains an intriguing and unanswered question.

Another open question in GH43 catalytic mechanisms concerns the conformation that the substrate follows during the hydrolysis reaction, in particular the conformational itinerary of the sugar located at the −1 enzyme subsite (the one bearing the scissile glycosidic bond)\(^23\). In glycosidases, such itinerary is a signature of enzyme family, and has been proven to be instrumental for inhibitor and activity-based probe design\(^24\). In the case of GH43 enzymes, those acting on β-galactans have been proposed to follow a \(\text{S}_1 \rightarrow \left[\text{H}_4 S\right] \rightarrow \text{C}_1\) itinerary\(^26\), whereas a \(\text{S}_1 \rightarrow \left[\text{B}_3 S\right] \rightarrow \text{C}_1\) itinerary was proposed for those acting on β-xylose derivatives\(^27\). Considering the large functional diversity within GH43 family, with specificity for distinct polysaccharides or sugar moieties, the conformational routes and catalytic mechanisms of GH43 enzymes might further diverge. However, the rarity of Michaelis complex structures available for these enzymes hampers the elucidation of their molecular mechanisms of action.

Here we report the non-reducing-end xylose-releasing exo-oligoxylanase activity in the GH43 family and provide the crystal structure of a native GH43 Michaelis complex prior to catalysis. The two saccharide configurations trapped in the active Michaelis complex showed to be catalytically viable through distinct conformational itineraries as demonstrated by QM/MM metadynamics. The placing of catalytic conformational promiscuity in glycosidases, a major class of enzymes in most living systems and of high biotechnological relevance, will impact on the current understanding of how these enzymes can interact with substrates and inhibitors. Moreover, we demonstrate the calcium role in the GH43 activity, which involves the stabilization of the productive conformation of the enzyme-substrate complex including the pre-activated −1 saccharide conformation and the catalytically competent state of the general base.

**Results**

**Discovery of a GH43 calcium-activated exo-oligoxylanase.** The xylanolytic system of *Xanthomonas citri* pv. *citrorum* (Xac) is highly equipped with several CAZymes involved in xylan breakdown, including three GH10s, one GH51, one GH39, two predicted arabinoxyranosidases, and a major class of enzymes in most living systems and of high biotechnological relevance, will impact on the current understanding of how these enzymes can interact with substrates and inhibitors. Moreover, we demonstrate the calcium role in the GH43 activity, which involves the stabilization of the productive conformation of the enzyme-substrate complex including the pre-activated −1 saccharide conformation and the catalytically competent state of the general base.
The three independent experiments (V0 + ΔV) show that the velocity of xylose hydrolysis. A xylose molecule in the oligomeric substrate (xylooligosaccharide) was obtained by means of short soaking followed by ash-freezing in a nitrogen stream. To understand the catalytic route for the exo-oligoxylanase XacGH43_1, a crystallographic snapshot of an active Michaelis substrate complex was found occupying the monosaccharide was found occupying the relaxed chair conformation (C1). The XacGH43_1 structure consists of a 5-bladed β-propeller catalytic core conserved in GH43 enzymes and does not display accessory domain, which is observed in some GH43 subfamilies. According to structural comparisons, XacGH43_1 is not interacting with the calcium ion directly. In particular, residues Asp150 (pKa modulator) and Asp30 (general base) are part of the second coordination sphere of the calcium ion.

Calcium stabilizes a pre-catalytic configuration of the active site. As described above, the crystal structure shows the presence of a Ca2+ ion near the −1 subsite. The Ca2+ ion is coordinated by the His288 N2 atom and by six water molecules in a pentagonal bipyramidal coordination geometry (Fig. 3 and Supplementary Fig. 3). The water molecules coordinating the calcium ion are connected to the carbonyl groups of residues Ser32, Asp150, Pro151, Ala99, Pro100, Asp30, His289, Gly238, and Gly309 residues from the very N- and C-termini (Supplementary Fig. 2).

These residues act as an extended and continuous platform for substrate anchoring, which may explain the higher affinity to X6 compared to X2. Therefore, based on these functional and structural observations, XacGH43_1 can be defined as a GH43 non-reducing-end xylose-releasing exo-oligoxylanase.

A crystallographic snapshot of an active Michaelis substrate complex. To understand the catalytic route for the exo-oligoxylanase activity in the GH43 family, the native enzyme and the complex with product (xylose) were crystallized. In addition, a complex with the substrate (xylooligosaccharide) was obtained by means of short soaking followed by flash-freezing in a nitrogen stream.

The XacGH43_1 structure shows that the active site is clearly blocked at the −1 subsite, in a relaxed chair conformation (C1). The complex of XacGH43_1 with xylooligosaccharide is not interacting with the calcium ion directly. In particular, residues Asp150 (pKa modulator) and Asp30 (general base) are part of the second coordination sphere of the calcium ion and with the preferred substrate (X6). The catalytic turnover (kcat), even though, was not affected as the Km, showing that the calcium influence on the catalytic efficiency was due to the increase in substrate affinity. All these observations indicate that we have found a calcium-dependent xylose-releasing exo-oligoxylanase.

Fig. 2 Catalytic properties of XacGH43_1. a. Capillary electrophoresis analysis showing the cleavage pattern of xylobiose, xylotriose, xylotetraose, xylopentaose, and xylohexaose. The reactions were interrupted at 0 min (dark yellow) or 30 min (blue). The standard patterns of xylose (X1), xylobiose (X2), xylooligosaccharide (X3), xylotetraose (X4), xylopentaose (X5), and xylohexaose (X6) are shown as gray curves. All data were normalized (starting from 1). The catalytic turnover (kcat), velocity at X6, and affinity for X6 (Km) were not affected, while the Km for other molecules of the asymmetric unit, with similar conformation as in the xylose complex.

All structures show that the active site is clearly blocked at the negative subsites, having only the highly conserved −1 subsite, which supports the exo mode of action, with the recognition of the non-reducing ends of the substrate. Molecular docking performed with X6 revealed that xylosyl residues beyond those crystallographic observed, i.e., +3, +4, and +5, can establish additional interactions with hydrophobic (Trp285, Trp344, Leu5, and Leu10) and polar (Thr312, Ser59, Gly310, and Gly309) residues from the very N- and C-termini (Supplementary Fig. 2). These residues act as an extended and continuous platform for substrate anchoring, which may explain the higher affinity to X6 compared to X2. Therefore, based on these functional and structural observations, XacGH43_1 can be defined as a GH43 non-reducing-end xylose-releasing exo-oligoxylanase.

Calcium stabilizes a pre-catalytic configuration of the active site. As described above, the crystal structure shows the presence of a Ca2+ ion near the −1 subsite. The Ca2+ ion is coordinated by the His288 N2 atom and by six water molecules in a pentagonal bipyramidal coordination geometry (Fig. 3 and Supplementary Fig. 3). The water molecules coordinating the calcium ion are connected to the carbonyl groups of residues Ser32, Asp150, Pro151, Ala99, Pro100, Asp30, His289, Gly238, and Gly309 (Supplementary Fig. 3). The catalytic triad, inferred from structural comparisons with other members of subfamily GH43_1, is not interacting with the calcium ion directly. In particular, residues Asp150 (pKa modulator) and Asp30 (general base) are part of the second coordination sphere of the calcium ion and

with the preferred substrate (X6). The catalytic turnover (kcat), even though, was not affected as the Km, showing that the calcium influence on the catalytic efficiency was due to the increase in substrate affinity. All these observations indicate that we have found a calcium-dependent xylose-releasing exo-oligoxylanase.
might interact with His288, while Glu237 (the general acid) is far from the Ca\(^{2+}\) ion.

To further investigate how the Ca\(^{2+}\) ion can affect enzyme stability and conformation, circular dichroism (CD) spectroscopy and small angle X-ray scattering (SAXS) were carried out. According to CD analysis, under calcium saturation, the enzyme had an increase of \(\approx 5\) °C in the transition temperature, indicating a gain of stability in the presence of the cation (Fig. 4a and Supplementary Fig. 4). SAXS data showed that the Ca\(^{2+}\) ion was responsible for decreasing the protein flexibility (Fig. 4b) and the ab initio molecular envelope exhibited a low normalized spatial discrepancy with the crystallographic monomer in the presence of the cation (Fig. 4c, Supplementary Fig. 5), confirming the monomeric state of XacGH43_1 (+Ca\(^{2+}\)) in solution.

To evaluate the role of Ca\(^{2+}\) ion in the structure and dynamics of the enzyme active site, we performed molecular dynamics (MD) simulations considering both the WT enzyme (Ca\(^{2+}\)-bound) (Fig. 5a) and the enzyme in the absence of the Ca\(^{2+}\) ion (in this case, we replaced it by a water molecule). Unexpectedly, in the absence of Ca\(^{2+}\), a Na\(^{+}\) ion from the solvent spontaneously entered into the calcium-binding site in just a few nanoseconds (Fig. 5b and Supplementary Fig. 6). While Ca\(^{2+}\) ion was coordinated by six or seven water molecules during the entire simulation, Na\(^{+}\) ion was coordinated by only four or five water molecules (Fig. 5a–c). This is in agreement with the sodium coordination previously observed by Matsuzawa et al. \(^{15}\) for another calcium-activated GH43_1 enzyme in the absence of calcium.\(^{15}\) In our simulations, the lack of an ion caused the side chain of His288 to alternate between two different orientations (named as rotamers 1 and 2, Fig. 5b, c, e, f), unlike what was observed in the Ca\(^{2+}\)-bound enzyme (rotamer 1). Moreover, the saccharide at the \(-1\) subsite changed to an inverted chair form.

---

**Fig. 3 Crystal structures of XacGH43_1.** Structure of XacGH43_1 without sugar (a), where a glycerol molecule occupies the \(-1\) subsite; in complex with xylose (b) and in complex with xylotriose (c). In all structures, the main interactions in the \(-1\) subsite and with Ca\(^{2+}\) (green sphere) are shown with distances in angstroms. Water molecules coordinating the Ca\(^{2+}\) are also shown (red spheres). The 2F\(_{o}\)-F\(_{c}\) electron density maps of ligands are shown with contour level of 1.0 \(\sigma\), after refinement. Subsites numbering scheme follows the previously proposed.\(^{91,92}\)
conformation ($1C_1$), a high energy conformer for β-xylose32 (see Fig. 5b, c and Supplementary Fig. 7a), a conformation that was not visited during the simulation of the Ca$^{2+}$-bound enzyme (Supplementary Fig. 7b). The lack of a divalent cation likely weakened the interaction between the Asp150 O2 atom and the xylosyl H2 that was instead attracted by Asp30, the general base, stabilizing the $1C_1$ conformation (Fig. 5d) and breaking the interaction of Asp30 residue with the catalytic water. Furthermore, in the alternate rotameric conformation, the His288 N41 atom interacts with the general base (Asp30). All these changes disrupt active site catalytic configuration, weakening substrate binding when Ca$^{2+}$ is not present.

Previous experiments in the presence of Na$^+$ ions (even when not supplied in the reaction, Na$^+$ ions were present in the enzyme buffer; Fig. 3b) indicated a decrease in enzyme activity in comparison with Ca$^{2+}$ supplied samples, in accordance with the above findings. All these observations indicate a critical role of calcium-coordinated His288 in stabilizing a pre-activated conformation of the −1 xylosyl moiety and in maintaining a productive configuration of the enzyme catalytic machinery.

**Fig. 4 Ca$^{2+}$ increases XacGH43_1 stability.** a Thermal denaturation profile of XacGH43_1, monitored at 228 nm by circular dichroism spectroscopy without additives (control—gray), with EDTA (pink) or CaCl$_2$ (blue). The melting temperatures ($T_m$) are indicated for each sample. b Kratky plot93 obtained for XacGH43_1 samples without (control—pink) or with CaCl$_2$ (blue). The dataset collected with CaCl$_2$ were used to generate an ab initio envelope. c XacGH43_1 crystal structure fitted into the SAXS envelope. Source data are provided as a source data file.

Two putative catalytic pathways for the same substrate. Interestingly, the Michaelis complex X-ray structure shows that the −1 xylopyranosyl ring of the xylotriose substrate adopts two distinct conformations, a $4C_1$ chair and a distorted 2,5-B skew-boat, each one having 50% occupancy. This suggests that both conformations could contribute to the enzyme reaction mechanism. To get further insight into this question, we quantified the conformational flexibility of the substrate by computing the conformational free energy landscape (FEL) of the β-xylosyl unit at the −1 enzyme subsite with respect to the Cremer–Pople puckering coordinates33, using QM/MM metadynamics. This approach has been used with success to predict the conformation of glycosides in the active site of several GHs34–36.

The computed conformational FEL of the β-xylosyl ring, reconstructed from the metadynamics simulation, shows that the protein scaffold significantly restricts the conformations available for the −1 sugar so that only two main regions of low energy are accessible, a $4C_1$ chair and a conformation in between 2S$_O$ and 2,5-B (Fig. 6a). Remarkably, the computed FEL is consistent with the two conformations observed in the crystal structure (Fig. 6b), both placed close to the minima of the conformational FEL. The small energetic difference between the two minima (≈1 kcal mol$^{-1}$, in favor of the distorted conformation) explains why two different conformations are experimentally observed.

To elucidate the catalytic mechanism by which XacGH43_1 exo-oligoxylanase hydrolyzes XOS, we performed QM/MM metadynamics simulations of the hydrolytic reaction, starting from the native xylotriose Michaelis complex. Three collective variables (CVs), involving all covalent bonds to be cleaved by the enzyme, were used to drive the system from reactants (xylotriose) to products (xylose + xylobiose). The first collective variable (CV1) accounts for proton transfer between Asp30 and the water molecule; CV$_2$ accounts for the nucleophilic attack of the catalytic water molecule; and CV$_3$ accounts for the transfer of the Glu237 proton to the glycosidic oxygen atom (Supplementary Fig. 8).

Of the two possible substrate conformations (Fig. 6), the distorted 2S$_O$/2,5-B conformation is likely to be the most pre-activated for catalysis in view of its axial C1-O1 bond orientation32. It is thus expected that the enzyme preferentially reacts via this conformation. Accordingly, the simulations were started from a snapshot of the QM/MM MD simulation in which the −1 sugar is in a 2S$_O$/2,5-B conformation.

Representative states along the minimum reaction free energy pathway are shown in Fig. 7a. Consistent with the conformational FEL (Fig. 6), the −1 sugar at the reactants state (R) adopts a conformation intermediate between 2S$_O$ and 2,5-B. At this state, the catalytic water for the inverting mechanism is positioned at 1.80 ± 0.06 Å from the general base (Asp30) and 3.64 ± 0.07 Å from the anomeric carbon (Fig. 7b, c and Supplementary Table 3). The reaction free energy barrier (14.1 kcal mol$^{-1}$) is in agreement with the value estimated from the measured reaction rate (≈15.8 kcal mol$^{-1}$).

The reaction begins with the elongation of the glycosidic bond, simultaneously with the transfer of the carboxylic hydrogen atom of the general acid residue to the glycosidic oxygen. From the reactants (R) to the TS, the −1 xylopyranosyl ring distorts from 2S$_O$/2,5-B to 2,5-B (Fig. 7a), a conformation compatible with the
Fig. 5 Effects of the presence or absence of Ca\(^{2+}\) ion in XacGH43_1 active site. a Representation of the \(-1\) subsite in the presence of Ca\(^{2+}\) in the metal-binding site. b Snapshot of the MD simulation in which the Ca\(^{2+}\) was initially replaced by a water molecule, followed by Na\(^{+}\) (from the solvent) entrance to the metal-binding site. Here the His288 adopts a different conformation (rotamer 2) and interacts with Asp30 (general base) and Asp30 interacts with the xylosyl H2. c A snapshot from the same simulation of b, with the Na\(^{+}\) being coordinated by His288. The interaction of Asp30-xylosyl H2 is maintained. Distances in a–c are represented in angstroms. d Distances between Asp150 O\(_2\) and xylosyl H2 atoms during the production of the three MD independent simulations with Ca\(^{2+}\) (blue) or with water (spontaneously replaced by Na\(^{+}\) during the course of the simulation) (pink). Averages are shown as solid symbols and the SD with lighter colors. Data points are shown as empty symbols. e His288 dihedral angle distribution of the rotamers 1 and 2 during the MD simulation with Ca\(^{2+}\). f Same results in the MD simulation without Ca\(^{2+}\) (spontaneously replaced by Na\(^{+}\) during the course of the simulation). Source data are provided as a source data file.

Fig. 6 Conformations adopted by the xylopyranosyl ring at the \(-1\) subsite of XacGH43_1 active site. a FEL of xylopyranose in the \(-1\) subsite of XacGH43_1 active-site obtained by ab initio metadynamics. The two conformations (found in the xylotriose crystal complex) are represented as purple squares. Isolines represent intervals of 1.0 kcal mol\(^{-1}\). b Representation of the xylosyl residue at the \(-1\) subsite in the xylotriose crystal complex, with 2F\(_o\)-F\(_c\) electron density map contoured at 1.7 \(\sigma\), after refinement. Source data are provided as a source data file.
requirement of an oxocarbenium-like TS\textsuperscript{23}. Likewise, the C1'-O5' bond of the −1 sugar shrinks with respect to its value at the reactants state (from 1.39 ± 0.04 to 1.29 ± 0.02 Å), indicative of the formation of a partial double bond between the C1' and O5' atoms. At the TS, the proton of the general acid residue (Glu237) is already transferred, the glycosidic bond is partially broken (2.72 ± 0.15 Å) (Fig.7b, c and Supplementary Table 3). Proton transfer from the water to the general base residue (Asp30) takes place after the TS, while the −1 sugar changes to a 5\textsuperscript{S1} conformation (P). Therefore, this conformational catalytic pathway of XacGH43\_1 can be described as 2\textsuperscript{S0}/2,5\textsuperscript{B} → [2,5\textsuperscript{B}]\textsuperscript{t} → 5\textsuperscript{S1}. Interestingly, the conformation at P is in agreement with the one recently observed by Matsuzawa et al\textsuperscript{115} in the product complex structure (PDB ID 5GLN) of a \(\beta\)-xylosidase/\(\alpha\)-L-arabinofuranosidase enzyme isolated from compost metagenome (CoXyl43), another member of the GH43\_1 subfamily.

Another QM/MM metadynamics simulation was performed, following the same approach, starting from the −1 xylopyranosyl ring in the 5\textsuperscript{C1} conformation, i.e., the alternative minimum in the conformational FEL that was also observed in the crystallographic complex. This conformation is formed with an equatorially oriented leaving group, is expected to be either unreactive or much less reactive than the distorted 2\textsuperscript{S0}/2,5\textsuperscript{B} conformation, in which the leaving group is axial and the xyloside ring is pre-activated for catalysis\textsuperscript{22}. Nevertheless, the evolution of relevant reaction distances along the resulting reaction pathway 4\textsuperscript{C1} → [E\textsuperscript{t}][4] → 4\textsuperscript{C1} (Fig. 8a) was found to be very similar to the previous itinerary (Fig. 8b, c and Supplementary Table 3). Moreover, the computed reaction FEL (Fig. 8d and Supplementary Fig. 10) is an indicative of a viable reaction pathway with a calculated free energy barrier of 17.5 kcal mol\textsuperscript{−1}, which is less favored than the one starting from a 2\textsuperscript{S0}/2,5\textsuperscript{B} conformation but still compatible with experimental data. The reaction follows an unconventional itinerary for xylanolytic enzymes (and glycosidases in general) in which both Michaelis and product complexes have the same conformation (4\textsuperscript{C1}), although the TS conformation (E3) is fully compatible with an oxocarbenium ion-like species\textsuperscript{23}. Therefore, both experiment and simulation show that the substrate can adopt two different conformations in the active site of XacGH43\_1, each one leading to a distinct catalytic itinerary, 2\textsuperscript{S0}/2,5\textsuperscript{B} → [2,5\textsuperscript{B}]\textsuperscript{t} → 5\textsuperscript{S1} and 4\textsuperscript{C1} → [E\textsuperscript{t}][4] → 4\textsuperscript{C1}.

Interestingly, one of the two itineraries here described for XacGH43\_1, 2\textsuperscript{S0}/2,5\textsuperscript{B} → [2,5\textsuperscript{B}]\textsuperscript{t} → 5\textsuperscript{S1}, was previously described for a GH43 enzyme from another subfamily (GH43\_11 from \textit{Geobacillus stearothermophilus} T-6), which shares very low identity with XacGH43\_1, indicating a conservation of the itinerary among GH43 subfamilies acting on xylose-derived substrates\textsuperscript{27}. Even though alternative substrate conformations were not investigated by Barker and coworkers\textsuperscript{27}, the similarity of the active site cavity makes very likely that the conformational catalytic promiscuity obtained here are extensible to other GH43\_11 members. These findings unfold an alternative...
mechanistic aspect of glycosidases, in which two different catalytic itineraries might be viable for a given substrate.

Discussion

The xylose-releasing exo-oligoxylanase activity has been kinetically and mechanistically demonstrated in the GH43 family. Such activity was characterized so far only in families GH8 and GH16 and in both cases the recognition mechanism involves the reducing ends, differing from that here reported for the GH43 family (non-reducing ends). Most GH43_1 enzymes were kinetically characterized in previous studies only against synthetic substrates or not systematically using different lengths of XOS. The xylose-releasing exo-oligoxylanase activity has been kinetically and mechanistically demonstrated in the GH43 family. Such catalytic itineraries might be viable for a given substrate.

A snapshot of the Michaelis complex of the WT enzyme with a substrate (xylotriose) was obtained at high resolution. Such complexes are rarely available in GHs and the only precedent active Michaelis complex of a GH enzyme trapped in the crystal structure before hydrolysis was reported for a GH1 member. The −1 xylosyl residue in the XacGH43_1 complex showed two alternative conformations, 4C1 and 2S0, which were further confirmed by QM/MM metadynamics simulations being almost iso-energetic.

QM/MM metadynamics simulations of the inverting reaction mechanism starting from the 2S0 distorted conformation resulted in the catalytic itinerary 2S0 → 2.5B → 2.5B′ → 3S1, which is in agreement with both the experimental free energy barrier and the trapped product of a cleaved X3 in the GH43_1 structure (PDB ID 5GLN). In fact, this itinerary is similar to the one previously reported for another subfamily of GH43 enzymes (2.5B′/2S0/B14 → 2.5B′ → 2.5B′/2S0) that shares very low sequence identity with XacGH43_1, indicating that, despite differences in sequence and activity, this hydrolytic reaction mechanism is conserved in other GH43 subfamilies acting on XOS.

Although two alternative conformations of the −1 sugar were not expected to be catalytically competent, our results show that...
the reaction starting from the 4C1 conformation is also viable, with a free energy barrier just ~3 kcal mol⁻¹ higher than the one computed for the 3S,4S/2R,2B conformation. The reaction proceeds in this case via a different TS conformation (Et), delineating a “cyclic” itinerary, C1 → [E₂] → C1. Therefore, although less favored, this reaction pathway cannot be excluded from contributing to the experimental reaction rate.

The fact that a chair conformation leads to a viable itinerary could be surprising a priori, since it has been shown that all β-GHs operate via a catalytic itinerary that begins from a specific distorted conformation of the sugar⁻³⁻⁻. However, GH43_1 members are exo-acting enzymes, thus the leaving group can easily adapt to changes in the glycosidic bond during catalysis, in which an oxocarbonium ion-like is formed, something that it is more difficult in endo-acting GHs. In addition, these enzymes can bind distinct sugars in different stereochemistry, which may contribute to make them more likely to accept non-distorted XOS without interrupting specific time intervals by adding a saturated sodium tetraborate solution, followed by heating in 65 °C for 5 min. The hydrolysis of polymeric substrates was monitored by estimating the reducing sugar released, according to the 3,5-dinitrosalicylic acid method⁻²⁻. Initial activity tests were carried out with 25 mM (polygalacturonic acid) or 2.5 mM (synthetic substrate), 40 mM mCvaine buffer⁻²¹ pH 7.0 and 25 μg of enzyme (6.25 μM). Reactions with nPn-derivatives were stopped after 1 h and reactions with polymeric substrates after 4 h. Spectrophotometric data were collected in an Infinite 200 PRO microplate reader (TECAN Group Ltd., Männedorf, Switzerland) using the i-Scint solution (1:10:4.0) (TECAN).

The optimum pH was evaluated in 40 mM citrate/phosphate/glycine buffer, ranging from 2.5 to 9.5 at 40 °C using 1 mM nPN-β-xylanopyranoside (nPN-β-Xyl) and 1 μg (250 nM) of enzyme, with 10 min of reaction. Optimum temperature was evaluated at same conditions above at pH 7.0, from 5 to 60 °C. For the assays with different sugars and chelating agents, 5 mM of the respective additive was used in reaction in 40 mM Hepes pH 7.0 at 40 °C with reaction time of 10 min. The same parameters were maintained, with exception of the salt concentration, for testing different CaCl₂ concentrations. Kinetics experiments using nPN-β-Xyl as substrate were performed at 40 °C with 0.5 μg (125 nM) of enzyme for 10 min in 40 mM Hepes pH 7.0 with 6 mM CaCl₂. For enzyme assays against different concentrations of xylan, it was used 25 μg (6.25 μM) of enzyme in 40 mM Hepes pH 7.0 supplemented with 6 mM CaCl₂. The temperature was set to 40 °C and a reaction time of 120 min.

To analyzecleavage patterns by CE, reactions with different XOS from X2 to X6 were carried at pH 7.0 at 40 °C, using 1 μg (0.32 μM) of enzyme and 5 mM of substrate. Aliquots of each reaction were taken at 0 or 30 min, heated to 95 °C for 15 min and dried. For labeling, samples were incubated with 150 nM cyanoborohydride and 6.5 mM APTS for 90 min at 60 °C. Samples were resuspended in running buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 5 mM imidazole, and 4 mM PMSE) and incubated on ice with lysozyme (0.1 mg mL⁻¹) 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-Ca and pre-equilibrated with 20 mM sodium phosphate, pH 7.4, 500 mM NaCl and 5 mM imidazole at a flow rate of 1 mL min⁻¹. XacGH43_1 was eluted using a non-linear gradient of imidazoles (up to 0.5M) at a flow rate of 1 mL min⁻¹. The eluted fractions were analyzed by SDS-PAGE, pooled, concentrated and submitted to size-exclusion chromatography in a HiLoad 16/600 Superdex 200 column (GE Healthcare), pre-equilibrated with 20 mM sodium phosphate, pH 7.4 and 130 mM NaCl at a flow rate of 1.0 mL min⁻¹. Samples from size-exclusion chromatography were analyzed by dynamic light-scattering (DLS) in a Malvern Zetasizer Nano series Nano-ZS (model ZEN3600) instrument (Malvern Zetasiser, Worcestershire, UK). DLS data were collected and analyzed with Zetasizer (7.12) software to evaluate sample homogeneity before pooling and concentration for crystallization trials.

Circular dichroism spectroscopy. Far UV CD spectra were recorded on a Jasco J-815 spectropolarimeter (JASCO International Co., Tokyo, Japan) using a 1-mm quartz cuvette and Spectra Manager II software (Jasco). Samples at 7.5 μM in 20 mM Hepes pH 7.0 were used without additives (control) and with addition of 6 mM CaCl₂ or 5 mM EDTA. Spectra were collected at 20 °C, with a response time of 4 s nm⁻¹. Buffers spectra were also collected and subtracted from the respective sample. Thermal unfolding experiments were monitored at 228 nm. Samples were heated from 20 to 100 °C, with a heating rate of 1 °C min⁻¹. Melting temperatures were obtained by sigmoidal-Bolzmann fit of denaturation curves from 20 to 85 °C.

Small angle X-ray scattering (SAXS). SAXS experiments were collected at protein concentration of 12 mg mL⁻¹ (300 μM) in 20 mM Hepes pH 7.5 with or without addition of 6 mM CaCl₂. Measurements were performed using 1.1 μm copper X-ray microbeam at the SAXS beamline from the Brazilian Synchrotron Light Laboratory (LNLS, Campinas, Brazil). Scattering data were recorded using a PILATUS 300 K (Dectris, Baden-Dättwil, Switzerland) and integrated using Fit2D. Data were processed using the ATSAS package. The program GNOM was used to evaluate the angular distribution of the particles and Guinier analysis of the size was calculated for the sample with calcium using DAMMIN. Averaged models were generated from several runs using DAMAVER. The theoretical scattering curve of XacGH43_1 crystallographic coordinates was calculated using CRYSOL. The XacGH43_1 crystal structure was fitted into the SAXS molecular envelope using the program SUPCOMB.

Enzyme assays. XacGH43_1 activity was tested against several natural and 4-nitrophenyl derived substrates (Supplementary Table 1). The 4-paranitrophenyl (nPN) derived substrates were purchased from Sigma-Aldrich (St. Louis, MO) and polymeric substrates were purchased from Megazyme (Wicklow, IE). Reactions with nPN-derived substrates were initiated by the addition of the enzyme and interrupted after specific time intervals by adding a saturated sodium tetraborate solution, followed by heating in 65 °C for 5 min. The hydrolysis of polymeric substrates was monitored by estimating the reducing sugar released, according to the 3,5-dinitrosalicylic acid method. Initial activity tests were carried out with 25 mM (polygalacturonic acid) or 2.5 mM (polygalactosamine) of substrate, 40 mM mCvaine buffer pH 7.0 and 25 μg of enzyme (6.25 μM). Reactions with nPN-derived substrates were stopped after 1 h and reactions with polymeric substrates after 4 h. Spectrophotometric data were collected in an Infinite 200 PRO microplate reader (TECAN Group Ltd., Männedorf, Switzerland) using the i-Scint solution (1:10:4.0) (TECAN).

The optimum pH was evaluated in 40 mM citrate/phosphate/glycine buffer, ranging from 2.5 to 9.5 at 40 °C using 1 mM nPN-β-xylanopyranoside (nPN-β-Xyl) and 1 μg (250 nM) of enzyme, with 10 min of reaction. Optimum temperature was evaluated at same conditions above at pH 7.0, from 5 to 60 °C. For the assays with different sugars and chelating agents, 5 mM of the respective additive was used in reaction in 40 mM Hepes pH 7.0 at 40 °C with reaction time of 10 min. The same parameters were maintained, with exception of the salt concentration, for testing different CaCl₂ concentrations. Kinetics experiments using nPN-β-Xyl as substrate were performed at 40 °C with 0.5 μg (125 nM) of enzyme for 10 min in 40 mM Hepes pH 7.0 with 6 mM CaCl₂ and enzyme concentration of 125 nM. The reaction time and enzyme concentration were selected in a linear region of the hydrolysis reaction. For reactions with X6 in the presence of CaCl₂, it was used 2.5 μg mL⁻¹ (62.5 μM) XacGH43_1, whereas 5 μg mL⁻¹ (125 nM) of the enzyme was employed in reactions without the addition of CaCl₂. For reactions with X2, we used 15 μg mL⁻¹ (375 nM) in all the reactions. The reactions were set with increasing concentrations of oligosaccharides (X6 and X2) in 30 mM Hepes buffer pH 7.0, with or without 6 mM CaCl₂. After 5 min of reaction, 40 μL methanol were added to stop the enzyme activity. A 15 μL aliquot of the stopped reaction was added with 2 μL manualsulfate (1 mM) and 183 μL water and injected into the micellar liquid chromatography column. The reaction time and enzyme concentration were selected in a linear region of the hydrolysis reaction. For reactions with X6 in the presence of CaCl₂, it was used 2.5 μg mL⁻¹ (62.5 μM) XacGH43_1, whereas 5 μg mL⁻¹ (125 nM) of the enzyme was employed in reactions without the addition of CaCl₂. For reactions with X2, we used 15 μg mL⁻¹ (375 nM) in all the reactions. The reactions were set with increasing concentrations of the reaction products, plotting the intensity data of the reaction product divided by the internal standard intensity (Iₖ/Iₛ) versus the xylose concentration (Supplementary Fig. 11). Thus, for enzyme kinetics reactions, the values obtained from intensity Iₖ/Iₛ were converted to xylose concentration. The kinetic parameters (kcat and KM) were calculated using OriginPro (OriginLab Corporation, Northampton, MA). All quantitative enzyme assays were expressed as mean ± SD from three independent experiments.
X-ray crystallography. XacGH43_1 at 28 mg mL\(^{-1}\) (0.7 mM) crystallized by vapor diffusion method in sitting and hanging drops containing 0.2 M ammonium sulfate, (pH 7.0) and 0.02 M sodium cacodylate pH 6.5 and 10% (v/v) 1,4-dioxane. To obtain space groups, the crystal was transferred to a solution containing the mother liquor added to 10 mM of sugar solution and incubated at room temperature before being flash-frozen in the nitrogen stream at 100 K and exposed to X-ray radiation.

X-ray diffraction data were collected at the MX2/LNLS beamline (Campinas, São Paulo, Brazil) with a 1.459 Å wavelength, using a PILATUS2M detector (Dectris, Baden-Dättwil, Switzerland) and MXCube software. Diffraction data were scaled and reduced using XDSD. The structure of XacGH43_1 was solved by molecular replacement using the program MOLREP and coordinates of the β-xylanase RS2328R (PDB ID 4MLG) as search model. Two molecules were found in the asymmetric unit and the generated model was refined with phenix.refine\(^6\) and REFMAC5, with visual inspections and manual building using COOT. Final model was validated using MolProbity\(^5\) and CheckMyMetal server. Figures containing crystallographic coordinates were generated using PyMol (Schrödinger, LLC, New York). Data processing and refinement statistics are summarized in Supplementary Table 4.

**Molecular docking.** Molecular docking calculations were carried out using Autodock Vina software.\(^6\) To prepare both the protein and the ligand (XacGH43_1 and X6, respectively), it was employed the Autodock Tools graphical interface.\(^5\) The enzyme atomic coordinates were taken from the XacGH43_1 crystal structure, whereas the X6 ligand from PDB entry 4HK8. The surface area evaluated corresponds to the active site pocket and its extended vicinity compatible with the crystal structure, whereas the X6 ligand from PDB entry 4HK8. The surface area for spatial restraints to the protein and ligands were applied during the 100, 200, and 300 K in the NVT ensemble at intervals of 50 ps. Spatial restraints to the protein and substrate were initially minimized, keeping both the protein and substrate fixed. Then, the entire system was minimized, keeping both the protein and substrate fixed. Then, the entire system was relaxed. Afterwards, the system was re-equilibrated with 6 ps of QM/MM MD at 300 K. A representative snapshot was used as starting points for the corresponding QM/MM metadynamics calculations of the conformational FEL.

### Metadynamics simulations of the chemical reaction.

To model the chemical reaction, the QM region was extended to include, beside the sugar atoms, the catalytic residues sidechains and also the catalytic water (Supplementary Fig. 13). Hydrogens were added to the 2 uses of pseudopotentials (P57). We used for all elements and calculations were performed using the Perdew, Burke, and Ernzerhof generalized gradient-corrected approximation\(^4\). This functional form has been proven to give a good performance in the description of hydrogen bonds\(^5\) and sugar conformations\(^6\) and was already used with success in previous works on GHs and glycosyltransferases.\(^7\) A fictitious electron mass of 700 a.u. and a time step of 0.12 fs was used. The structure was optimized using QM/MM MD with annealing of the ionic velocities, until the maximal component of the nuclear gradient was lower than 10\(^{-4}\) a.u. Afterwards, the system was re-equilibrated with 6 ps of QM/MM MD at 300 K. A representative snapshot was used as starting points for the corresponding QM/MM metadynamics calculations of the conformational FEL.

The metadynamics approach\(^8\) was used within the metadynamics driver provided by the Plumed plugin.\(^9\) The height and the width of the Gaussian terms were set to 1.0 kcal mol\(^{-1}\) and 0.1 Å, respectively, for q1 and q2, with a temperature of 200 MD steps (20 fs) and a bias factor of 10. The simulation was stopped once no qualitative changes in the FEL were observed; the simulation time corresponds to 70 and 66 ps, respectively. Transition states were further refined by isocommittor analyses, according to the following protocol. A region around the point of maximum energy of the reaction pathway\(^10\) in the reaction FEL was defined as a starting point for isocommittor analyses,\(^9\), which were performed on configurations/structures corresponding to the conformational frames falling on this region (a total of 23 configurations were initially tested, considering both TS configurations). Twenty independent unbiased dynamical trajectories with random initial velocities were launched for selected configurations, which were stopped once the system reached either the reactants or the products state. The putative TS for each reaction was taken as the configuration that gives a reactants/products ratio close to 50% (Supplementary Fig. 14).

To further confirm the values of the computed energy barriers, we launched metadynamics simulations with different Gaussian heights (0.50 and 1.00 kcal mol\(^{-1}\)) for reaction starting from C4 and C5. The simulations were performed with the protocol proposed in the literature\(^8\), corresponding to 70 and 66 ps, respectively. Transition states were further refined by isocommittor analyses, according to the following protocol. A region around the point of maximum energy of the reaction pathway\(^10\) in the reaction FEL was defined as a starting point for isocommittor analyses,\(^9\), which were performed on configurations/structures corresponding to the conformational frames falling on this region (a total of 23 configurations were initially tested, considering both TS configurations). Twenty independent unbiased dynamical trajectories with random initial velocities were launched for selected configurations, which were stopped once the system reached either the reactants or the products state. The putative TS for each reaction was taken as the configuration that gives a reactants/products ratio close to 50% (Supplementary Fig. 14).

The well-tempered metadynamics approach\(^8\) was used within the metadynamics driver provided by the Plumed plugin.\(^9\) The height and the width of the Gaussian terms were set to 1.0 kcal mol\(^{-1}\) and 0.1 Å, respectively, for q1 and q2, with a temperature of 200 MD steps (20 fs) and a bias factor of 10. The simulation was stopped once no qualitative changes in the FEL were observed; the simulation time corresponds to 70 and 66 ps, respectively. Transition states were further refined by isocommittor analyses, according to the following protocol. A region around the point of maximum energy of the reaction pathway\(^10\) in the reaction FEL was defined as a starting point for isocommittor analyses,\(^9\), which were performed on configurations/structures corresponding to the conformational frames falling on this region (a total of 23 configurations were initially tested, considering both TS configurations). Twenty independent unbiased dynamical trajectories with random initial velocities were launched for selected configurations, which were stopped once the system reached either the reactants or the products state. The putative TS for each reaction was taken as the configuration that gives a reactants/products ratio close to 50% (Supplementary Fig. 14).

To further confirm the values of the computed energy barriers, we launched metadynamics simulations with different Gaussian heights (0.50 and 1.00 kcal mol\(^{-1}\)) for reaction starting from C4 and C5. The simulations were performed with the protocol proposed in the literature\(^8\), corresponding to 70 and 66 ps, respectively. Transition states were further refined by isocommittor analyses, according to the following protocol. A region around the point of maximum energy of the reaction pathway\(^10\) in the reaction FEL was defined as a starting point for isocommittor analyses,\(^9\), which were performed on configurations/structures corresponding to the conformational frames falling on this region (a total of 23 configurations were initially tested, considering both TS configurations). Twenty independent unbiased dynamical trajectories with random initial velocities were launched for selected configurations, which were stopped once the system reached either the reactants or the products state. The putative TS for each reaction was taken as the configuration that gives a reactants/products ratio close to 50% (Supplementary Fig. 14).

To further confirm the values of the computed energy barriers, we launched metadynamics simulations with different Gaussian heights (0.50 and 1.00 kcal mol\(^{-1}\)) for reaction starting from C4 and C5. The simulations were performed with the protocol proposed in the literature\(^8\), corresponding to 70 and 66 ps, respectively. Transition states were further refined by isocommittor analyses, according to the following protocol. A region around the point of maximum energy of the reaction pathway\(^10\) in the reaction FEL was defined as a starting point for isocommittor analyses,\(^9\), which were performed on configurations/structures corresponding to the conformational frames falling on this region (a total of 23 configurations were initially tested, considering both TS configurations). Twenty independent unbiased dynamical trajectories with random initial velocities were launched for selected configurations, which were stopped once the system reached either the reactants or the products state. The putative TS for each reaction was taken as the configuration that gives a reactants/products ratio close to 50% (Supplementary Fig. 14).

To further confirm the values of the computed energy barriers, we launched metadynamics simulations with different Gaussian heights (0.50 and 1.00 kcal mol\(^{-1}\)) for reaction starting from C4 and C5. The simulations were performed with the protocol proposed in the literature\(^8\), corresponding to 70 and 66 ps, respectively. Transition states were further refined by isocommittor analyses, according to the following protocol. A region around the point of maximum energy of the reaction pathway\(^10\) in the reaction FEL was defined as a starting point for isocommittor analyses,\(^9\), which were performed on configurations/structures corresponding to the conformational frames falling on this region (a total of 23 configurations were initially tested, considering both TS configurations). Twenty independent unbiased dynamical trajectories with random initial velocities were launched for selected configurations, which were stopped once the system reached either the reactants or the products state. The putative TS for each reaction was taken as the configuration that gives a reactants/products ratio close to 50% (Supplementary Fig. 14).

### Reporting summary.

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

### Data availability.

Coordinates and structure factors of XacGH43_1, XacGH43_1 with xylose and XacGH43_1 with xylotrose have been deposited in the Protein Data Bank with accession codes 6XN0, 6XN1 and 6XN2 respectively. Source data are provided with this paper. Additional data supporting the findings of this study are available from the corresponding authors on reasonable request.
References

1. Kohler, A. et al. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. Nat. Genet. 47, 4115 (2015).

2. DeBoy, R. T. et al. Insights into plant cell wall degradation from the genome sequence of the soil bacterium Cellvibrio japonicus. J. Bacteriol. 190, 5455–5463 (2008).

3. Lombard, V., Golconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, 490–494 (2014).

4. Mewis, K., Lenfant, N., Lombard, V. & Henrissat, B. Dividing the large genome. J. Biocat. Bioeng. 116, 734–749 (2014).

5. Xu, B. et al. Characterization of a novel salt-, xylose- and alkali-tolerant GH43 enzyme characterization. Environ. Microbiol. 82, 1686–1692 (2016).

6. Zanphorlin, L. M. et al. Structure-guided design combined with evolutionary insights into the catalytic machinery of family 1 glycosidases. J. Mol. Biol. 428, 429–437 (2019).

7. Yang, X. et al. Two xylose-tolerant GH43 bifunctional β-xylosidase/a-L-arabinofuranosidase activities from the gut bacterial genome. J. Biocat. Bioeng. 116, 120–128 (2015).

8. De Sanctis, D., Inácio, J. M., Lindley, P. F., De Sá-Nogueira, I. & Bento, I. New evidence for the role of calcium in the glycosidase reaction of GH43 arabinanases. FEBS Lett. 577, 4562–4574 (2010).

9. Davies, G. J., Planas, A. & Rovira, C. Conformational analyses of the reaction coordinate of glycosidases. Acc. Chem. Res. 45, 308–316 (2012).

10. DeSantis, I., Inácio, J. M., Lindley, P. F., De Sá-Nogueira, I. & Bento, I. New evidence for the role of calcium in the catalytic reaction of GH43 arabinanases. FEBS J. 277, 73–82 (2009).

11. DeSantis, I., Inácio, J. M., Lindley, P. F., De Sá-Nogueira, I. & Bento, I. New evidence for the role of calcium in the catalytic reaction of GH43 arabinanases. FEBS J. 277, 4562–4574 (2010).

12. Davies, G. J., Planas, A. & Rovira, C. Conformational contributions to glycosidase catalysis and inhibition. Curr. Opin. Struct. Biol. 28, 1–13 (2014).

13. Franke, D. et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. J. Appl. Crystallogr. 51, 1209–1225 (2018).

14. Honda, Y. & Kitaoka, M. A family 8 glycoside hydrolase from Bacillus halodurans C-125 (BH2105) is a reducing end xylose-releasing exo-oligosaccharide hydrolase. J. Bacteriol. 189, 1647–1650 (2007).

15. Huy, N. D., Thayumanavan, P., Kwon, T. H. & Park, S. M. Characterization of a recombinant bifunctional xylosidase/arabinofuranosidase from Phanerochaete chrysosporium. J. Biocat. Bioeng. 116, 152–159 (2013).

16. Proctor, M. R. et al. Tailored catalysis for plant cell wall degradation: Revealing the exo/end preference of Cellvibrio japonicus arabinanase 43A. Proc. Natl Acad. Sci. USA 102, 2697–2702 (2005).

17. Till, M. et al. Structural analysis of the GH43 enzyme Xsa43E from Butyrivibrio proteoclasticus. Acta Crystallogr. Sect. F Struct. Biol. Commun. 70, 1193–1198 (2014).

18. Jordao, D. B. et al. X-ray crystal structure of divalent metal-activated β-xylosidase, RS223BX. Appl. Biochem. Biotechnol. 177, 637–648 (2015).

19. Lee, C. C., Braker, J. D., Grigorescu, A. A., Wagschal, K. & Jordan, D. B. Divalent metal activation of a GH43 β-xylosidase. Enzyme Microb. Technol. 52, 84–90 (2013).

20. Ahmed, S. et al. A novel α-L-arabinofuranosidase from family 43 glycoside hydrolase (C43Araf) from Clostridium thermocellum. PLoS One 8, e73575 (2013).

21. Cartmell, A. et al. The structure and function of an arabinan-specific α-1,2-arabinofuranosidase identified from the genomes of bacterial GH43 glycoside hydrolases. J. Biol. Chem. 286, 15483–15495 (2011).

22. Alhassid, A. et al. Crystal structure of an inverting GH 43 1,5-a-L-arabinanase from Geobacillus stearothermophilus complexed with its substrate. Biochem. J. 422, 73–82 (2009).

23. DeSantis, I., Inácio, J. M., Lindley, P. F., De Sá-Nogueira, I. & Bento, I. New evidence for the role of calcium in the catalytic reaction of GH43 arabinanases. FEBS J. 277, 4562–4574 (2010).

24. Mewis, K., Lenfant, N., Lombard, V. & Henrissat, B. Dividing the large genome. J. Biocat. Bioeng. 116, 734–749 (2014).

25. Wu, L. et al. An overview of activity-based probes for glycosidases. Curr. Opin. Chem. Biol. 53, 25–36 (2020).
55. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 22–25 (2010).

56. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. Sect. D Biol. Crystallogr. 68, 352–367 (2012).

57. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. Sect. D Biol. Crystallogr. 53, 240–255 (1997).

58. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 486–501 (2010).

59. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci. 27, 293–315 (2018).

60. Zheng, H. et al. CheckMyMetal: a macromolecular metal-binding validation tool. J. Mol. Biol. 427, 1521–1532 (2015).

61. Troullier, N. & Martins, J. L. Efficient pseudopotentials for plane-wave calculations. Phys. Rev. B 43, 1993–2006 (1991).

62. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. Phys. Rev. Lett. 77, 3865–3868 (1996).

63. Ireta, J., Neugebauer, J. & Scheffler, M. On the accuracy of DFT for describing hydrogen bonds: Dependence on the bond directionality. J. Phys. Chem. A 108, 5692–5698 (2004).

64. Raich, L. et al. A trapped coherent intermediate of a glycose hydrolyase on the pathway to transglycosylation. Insights from experiments and quantum mechanics/molecular mechanics simulations. J. Am. Chem. Soc. 138, 3325–3332 (2016).

65. Marianski, M., Supady, A., Ingram, T., Schneider, M. & Baldauf, C. Assessing the accuracy of across-the-scale methods for predicting carbohydrate conformational energies for the examples of glucose and α-maltose. J. Chem. Theory Comput. 12, 6157–6168 (2016).

66. Ardèvol, A. & Rovira, C. Reaction mechanisms in carbohydrate-active enzymes: glycoside hydrolyses and glycosyltransferases. Insights from ab initio quantum mechanics/molecular mechanics dynamic simulations. J. Am. Chem. Soc. 137, 7528–7547 (2015).

67. Bilyard, M. K. et al. Palladium-mediated enzyme activation suggests multiple initiation of glycosynthesis. Nature 563, 235–240 (2018).

68. Loa, A. & Parrinello, M. Escaping free-energy minima. Proc. Natl. Acad. Sci. USA 99, 12562–12566 (2002).

69. Valsson, O., Tiwary, P. & Parrinello, M. Enhancing important fluctuations: rare events and metadynamics from a conceptual viewpoint. Annu. Rev. Phys. Chem. 67, 159–184 (2016).

70. Biarnés, X. et al. The conformational free energy landscape of β-D-glucopyranose. Implications for substrate preactivation in β-glucoside hydrolyses. J. Am. Chem. Soc. 129, 10686–10693 (2007).

71. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. Phys. Rev. Lett. 100, 1–4 (2008).

72. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: new feathers for an old bird. Comput. Phys. Commun. 185, 604–613 (2014).

73. Williams, T. & Kelley, C. gnuplot 5.0: An Interactive Plotting Program. 7, 259 (2017).

74. Zhang, Y., Lee, T. S. & Yang, W. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. J. Chem. Phys. 110, 1146–1150 (1999).

75. Marcos-Alcalde, L., Setoain, J., Mendieta-Moreno, J. I., Mendieta, J. & Gómez-Puertas, M. PEPSA: minimum energy pathway analysis for energy landscapes. Bioinformatics 31, 3853–3855 (2015).

76. Deliago, C., Bolhuis, P. G., Csajka, F. S. & Chandler, D. Transition path sampling and the calculation of rate constants. J. Chem. Phys. 108, 1964–1977 (1998).

77. Peters, B. P(TP)q peak maximization: necessary but not sufficient for reaction coordinate accuracy. Chem. Phys. Lett. 494, 100–103 (2010).

78. Nair, N. N., Schreiner, E. & Marx, D. Peptide synthesis in aqueous environments: the role of extreme conditions on amino acid activation. J. Am. Chem. Soc. 130, 14111–14120 (2008).

79. Biely, P., Krakty, Z. & Vrsanska, M. Substrate-binding site of endo-1,4-β-xylanase of the yeast Cryptococcus albidus. Eur. J. Biochem. 119, 559–564 (1981).

80. Davies, G. J., Wilson, K. S. & Henriassat, B. Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochem. J. 321, 557–559 (1997).

81. Glätter, O. & Kraliky, O. Small Angle X-ray Scattering (Academic Press Inc., 1982).

Acknowledgements

This research was funded by São Paulo Research Foundation (FAPESP, #2015/26982-0 to M.T.M.), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNpq #306135/2016-7 to M.T.M.), the Spanish Ministry of Science and Innovation (MICINN, AEI/FEDER, UE) (CTQ2017-85-496 to C.R.), the Spanish Structures of Excellence María de Maeztu (MDM-2017-14189 to C.R.), and the Agency for Management of University and Research Grants (AGAUR) (SGR2017-1189 to C.R.). M.A.B.M. received FAPESP postdoctoral fellowships (#2016/19995-0 and #2018/22138-8) and J.C. received a MICINN predoctoral fellowship FPI-BES-2015-027051. We thank the Brazilian Synchrotron Light Laboratory (LNLS) for the provision of time on the SAXS1 and MX2 beamlines, the Brazilian Biosciences National Laboratory (LNBio) for access to the crystallization (Robolab) facility and the Brazilian Biorenewables National Laboratory (NBRI) for access to the Characterization of Macromolecules (MAC) facility and all their staff for the support. We thank the Red Española de Supercomputación (RES) and the Barcelona Supercomputing Center (BSC) for provision of time in MareNostrum4 and CTE machines (QSB-2020-2-0006), and the staff for technical assistance. We also acknowledge Vanesa P. M. Martins for the help in XacGH43_1 cloning and all members of the groups led by Rovira (Barcelona, Spain) and Murakami (Campinas, Brazil) for their contributions to the peer review of this work. Peer review reports are available.

Author contributions

M.A.B.M. performed experiments, solved the structures, and wrote the manuscript. M.A. B.M. and J.C. designed all simulations, analyzed and wrote the results. M.N.D. and J.B.L. C. performed activity assays and analyses. R.A.S.P. and F.C.G. performed mass spectrometry kinetics assays and analyses. C.R.S. and C.C.C.T. performed cloning, expression, and purification, and crystallization trials. M.T.M. and C.R. coordinated the project, analyzed results, and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-20620-3.

Correspondence and requests for materials should be addressed to C.R. or M.T.M.

Peer review information Nature Communications thanks Mauro Boero, Jan-Hendrik Hehemann, and Alessandro Silva Nascimento for their contributions to the peer review of this work. Peer review reports are available.

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

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