

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

Mariana A. B. Morais ^{1,2}, Joan Coines ², Mariane N. Domingues ¹, Renan A. S. Pirolla¹, Celisa C. C. Tonoli³, Camila R. Santos¹, Jessica B. L. Correa¹, Fabio C. Gozzo ⁴, Carme Rovira ^{2,5}✉ & Mario T. Murakami ¹✉

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 xylooligosaccharides. 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.

¹Brazilian Biorenewables National Laboratory (LNBR), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas 13083-100, Brazil.

²Departament de Química Inorgànica i Orgànica & Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, Barcelona 08028, Spain. ³Brazilian Biosciences National Laboratory (LNBio), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas 13083-100, Brazil.

⁴Dalton Mass Spectrometry Laboratory, Institute of Chemistry, University of Campinas, Campinas 13083-970, Brazil. ⁵Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona 08010, Spain. ✉email: c.rovira@ub.edu; mario.murakami@lnbr.cnpe.br

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)^{1,2}. They constitute one of the largest GH families annotated in the CAZy database³ and are divided into 37 subfamilies that diverge in terms of structure and function⁴. 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 polysaccharide specificity, members from the same subfamily might exhibit distinct modes of action, substrate recognition, and quaternary structure^{5,6}.

A striking feature of GH43 enzymes is their bi-functionality, with several members exhibiting both β -xylosidase and α -arabinofuranosidase activities^{7–11}. Taking advantage of their intrinsic promiscuous scaffold, alternative functions have been discovered or introduced through rational redesign of the active-site cleft^{6,12–14}, which make members of this family good candidates for engineering relevant modes of action for industrial processes. To further increase the biotechnological potential of GH43 enzymes and to better understand their biological roles in microbial communities specialized in plant cell wall degradation, it is instrumental to elucidate their mechanisms of action and catalytic pathways at molecular detail, which have been so far not revealed.

GH43 enzymes cleave glycosidic bonds with inversion of the anomeric configuration. The reaction is catalyzed by two essential carboxylate-based residues (glutamic or aspartic acid). One of them (the general acid) protonates the scissile glycosidic oxygen atom, whereas the other essential residue (the general base) coordinates the nucleophile, a water molecule, to assist its deprotonation to complete the reaction (Fig. 1). Based on crystallographic data, some works suggested the importance of calcium for maintaining the protonation state of the catalytic general acid residue through the interaction with a water molecule^{15,16}. In fact, calcium ions have been found to be essential for the catalytic activity of GH43 enzymes (arabinanases, xylosidases, and arabinofuranosidases)^{14,15,17–19}. They are observed in several crystal structures in a hydrated cavity located near the -1 subsite, but not directly interacting with the substrate^{12,14–17,20–22}. 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

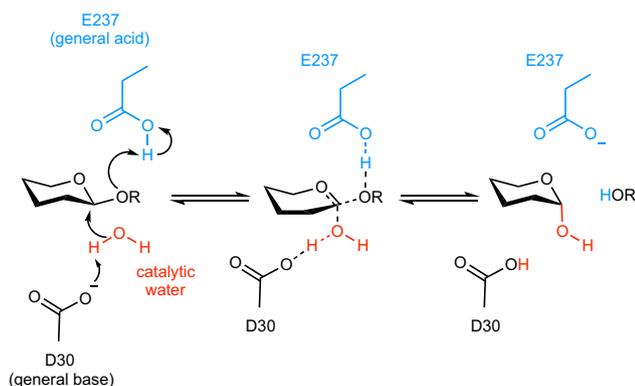


Fig. 1 General reaction mechanism of inverting GH43 enzymes. Residue numbers correspond to the enzyme under investigation.

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^{24,25}. In the case of GH43 enzymes, those acting on β -galactans have been proposed to follow a ${}^1S_3 \rightarrow [{}^4H_3]^{\ddagger} \rightarrow {}^4C_1$ itinerary^{24,26}, whereas a ${}^2S_0 \rightarrow [{}^2,5B]^{\ddagger} \rightarrow {}^5S_1$ itinerary was predicted for those acting on β -xylose derivatives²⁷. 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-oligoxyranase 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 configuration 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-oligoxyranase. The xylanolytic system of *Xanthomonas citri* pv. *citri* (Xac) is highly equipped with several CAZymes involved in xylan breakdown, including three GH10s, one GH51, one GH39, two predicted GH30, and six GH43s. Some of these enzymes exhibited atypical functional attributes such as the cleavage of the recalcitrant internal di-substitutions in arabinoxylans and the reducing-end exo-oligo activity^{28,29}. Herein, we investigated another enzyme from this system, encoded by the ORF Xac4258 (named here as XacGH43_1), which displayed both β -xylosidase and α -arabinofuranosidase activities, with preference for xylose-derived substrates, according to activity assays against several natural and synthetic substrates (Supplementary Table 1). The enzyme shows typical pH and temperature dependence observed for other CAZymes from this phytopathogen (Supplementary Fig. 1) and capillary zone electrophoresis (CZE) experiments revealed that XacGH43_1 releases xylose from all xylooligosaccharides (XOS) tested (from xylobiose (X2) to xylohexaose (X6)) (Fig. 2a). Its catalytic activity is positively influenced by the presence of Ca^{2+} ions, in comparison to the control and to other cations (Fig. 2b), being also responsive to increasing concentrations of CaCl_2 (Fig. 2b inset). To verify the preference of XacGH43_1 for longer or shorter XOS, kinetics assays were performed using quantitative mass spectrometry to monitor the xylose release from X2 and X6 (Fig. 2c, d). Kinetic parameters revealed \approx threefold higher catalytic efficiency of XacGH43_1 on X6 compared to X2, without the addition of CaCl_2 , and \approx eightfold higher when CaCl_2 was added to the reaction (Fig. 2c, d and Supplementary Table 2). The saturation over xylan was not reached, indicating that XacGH43_1 is able to cleave efficiently only XOS, despite being active on the polymeric substrate (Supplementary Fig. 1).

The calcium was found to be responsible for improving the substrate affinity (K_M) in one order of magnitude when tested

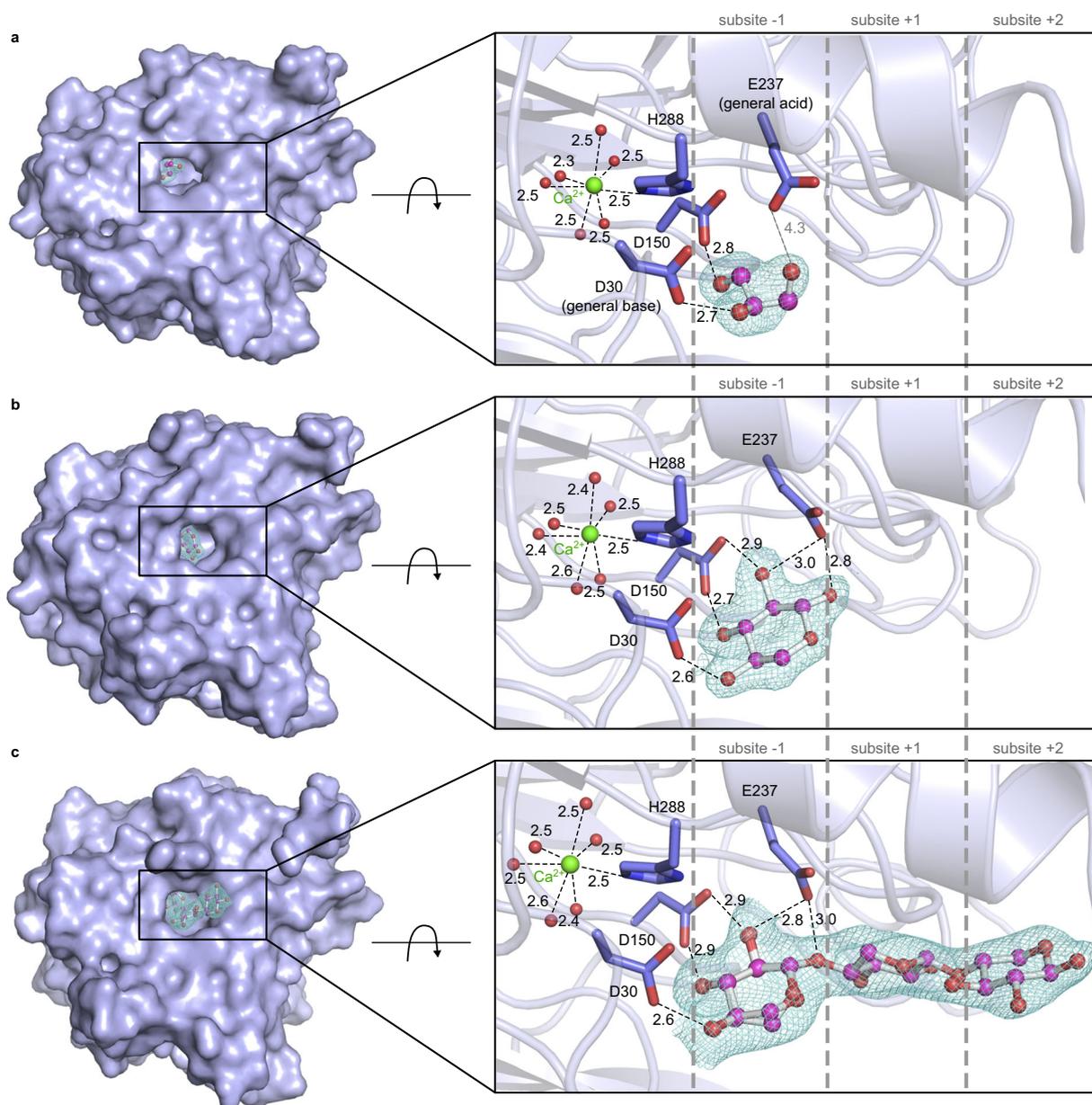


Fig. 3 Crystal structures of XacGH43_1. Structure of XacGH43_1 without sugar (a), where a glycerol molecule occupies the -1 subsite; in complex with xylose (b) and in complex with xylotriose (c). In all structures, the main interactions in the -1 subsite and with Ca^{2+} (green sphere) are shown with 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 1.0σ , after refinement. Subsites numbering scheme follows the previously proposed^{91,92}.

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 $\approx 5^\circ\text{C}$ 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 ($+\text{Ca}^{2+}$) in solution.

To assess 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 calcium¹⁵. 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

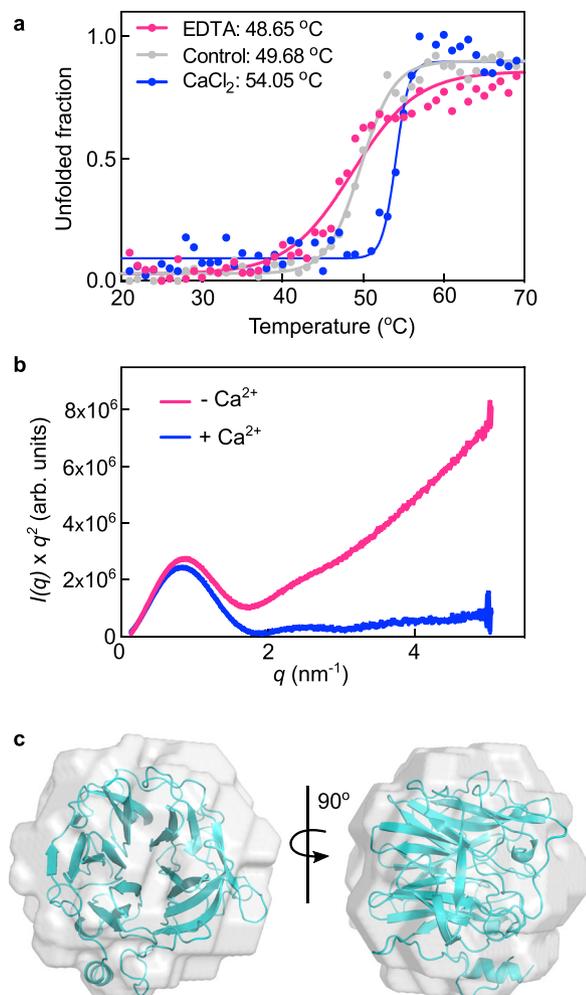


Fig. 4 Ca^{2+} increases XacGH43_1 stability. **a** Thermal denaturation profile of XacGH43_1, monitored at 228 nm by circular dichroism spectroscopy without additives (control—gray), with EDTA (pink) or CaCl_2 (blue). The melting temperatures (T_m) are indicated for each sample. **b** Kratky plot⁹³ 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.

conformation (${}^1\text{C}_4$), 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 $\text{O}_{\delta 2}$ atom and the xylosyl H2 that was instead attracted by Asp30, the general base, stabilizing the ${}^1\text{C}_4$ conformation (Fig. 5d) and breaking the interaction of Asp30 residue with the catalytic water. Furthermore, in the alternate rotameric conformation, the His288 $\text{N}_{\delta 1}$ 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 consistence 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.

Two putative catalytic pathways for the same substrate. Interestingly, the Michaelis complex X-ray structure shows that the -1 xylopyranosyl ring of the xylotriase substrate adopts two distinct conformations, a ${}^4\text{C}_1$ chair and a distorted ${}^2\text{S}_0$ 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 GHs^{34–36}.

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 ${}^4\text{C}_1$ chair and a conformation in between ${}^2\text{S}_0$ and ${}^{2,5}\text{B}$ (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 xylotriase Michaelis complex. Three collective variables (CVs), involving all covalent bonds to be cleaved by the enzyme, were used to drive the system from reactants (xylotriase) to products (xylose + xylobiose). The first collective variable (CV_1) accounts for proton transfer from 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 ${}^2\text{S}_0/{}^{2,5}\text{B}$ conformation is likely to be the most pre-activated for catalysis in view of its axial $\text{C}1\text{-O}1'$ 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 ${}^2\text{S}_0/{}^{2,5}\text{B}$ 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 ${}^2\text{S}_0$ and ${}^{2,5}\text{B}$. 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 proton donor (general acid, Glu237) forms a 1.59 ± 0.38 Å hydrogen bond with the xylosyl $\text{O}1'$ atom (Fig. 7b, c and Supplementary Table 3).

The reaction FEL obtained from the QM/MM metadynamics simulations, shown in Fig. 7d and Supplementary Fig. 9, is indicative of a concerted one-step reaction with two clear minima of low energy: the reactants (R), corresponding to the Michaelis complex and the products (P), separated by a single transition state (TS). 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 ${}^2\text{S}_0/{}^{2,5}\text{B}$ to ${}^{2,5}\text{B}$ (Fig. 7a), a conformation compatible with the

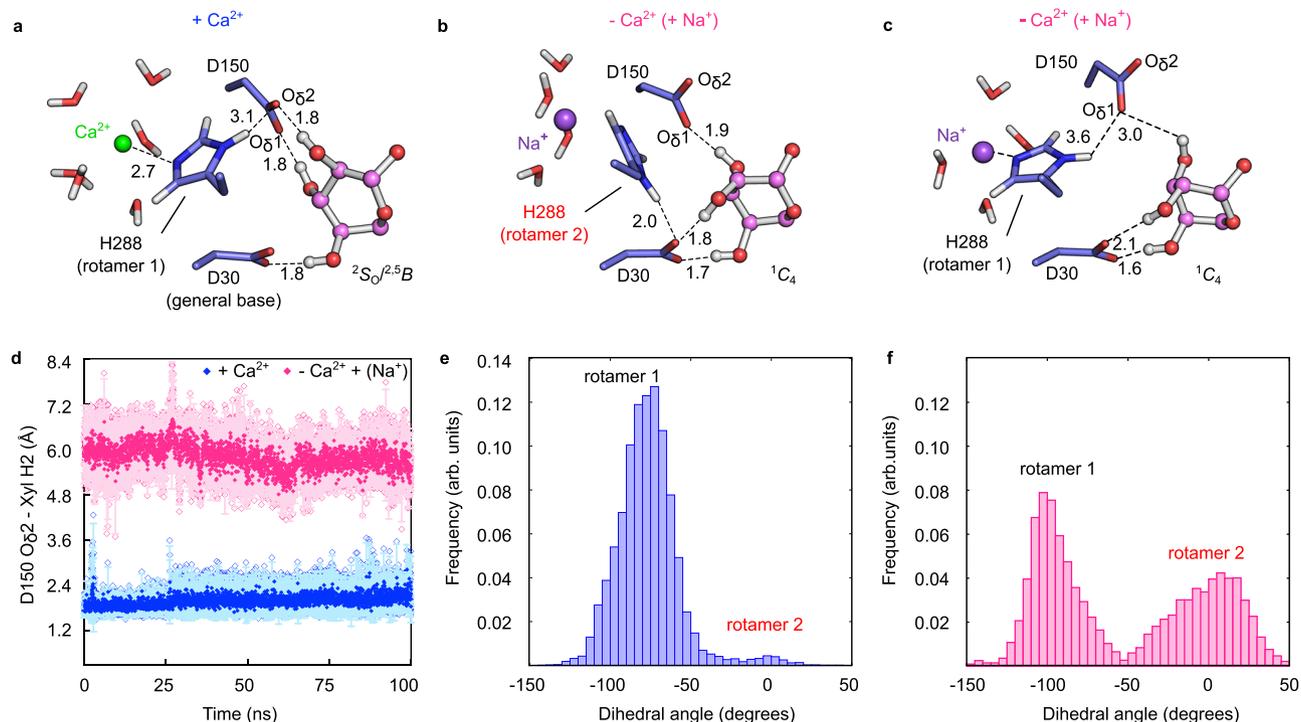


Fig. 5 Effects of the presence or absence of Ca^{2+} ion in XacGH43_1 active site. **a** Representation of the -1 subsite in the presence of Ca^{2+} in the metal-binding site. **b** Snapshot of the MD simulation in which the Ca^{2+} was initially replaced by a water molecule, followed by Na^+ (from the 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 $\text{O}_{\delta 2}$ 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.

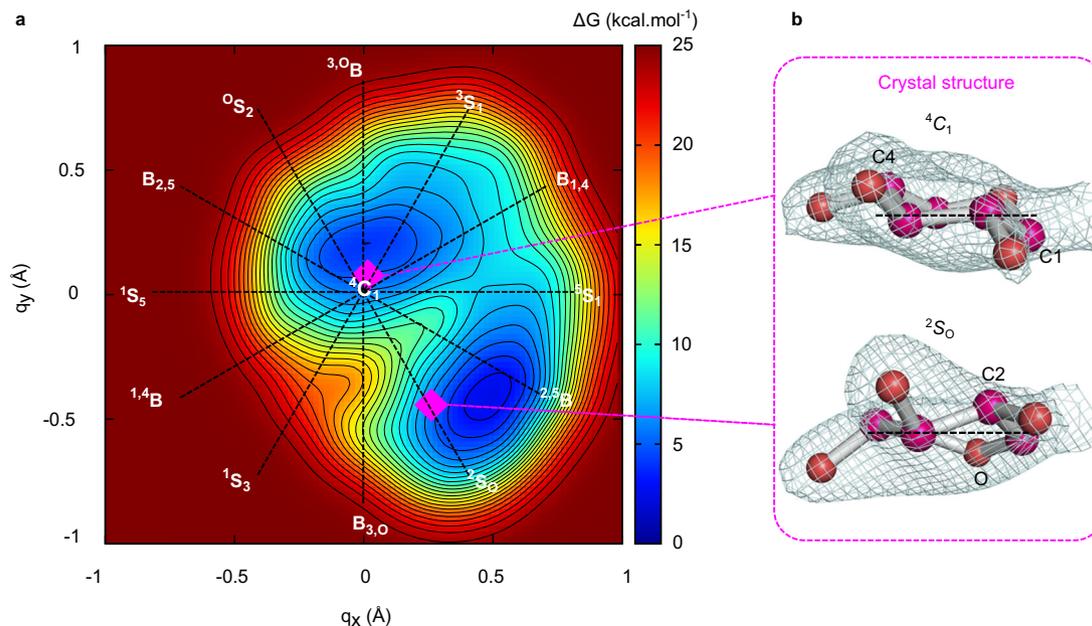


Fig. 6 Conformations adopted by the xylopyranosyl ring at the -1 subsite of XacGH43_1 active site. **a** FEL of xylopyranose in the -1 subsite of XacGH43_1 active-site obtained by ab initio metadynamics. The two conformations (found in the xylotriose crystal complex) are represented as purple squares. Isolines represent intervals of $1.0 \text{ kcal mol}^{-1}$. **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 1.7σ , after refinement. Source data are provided as a source data file.

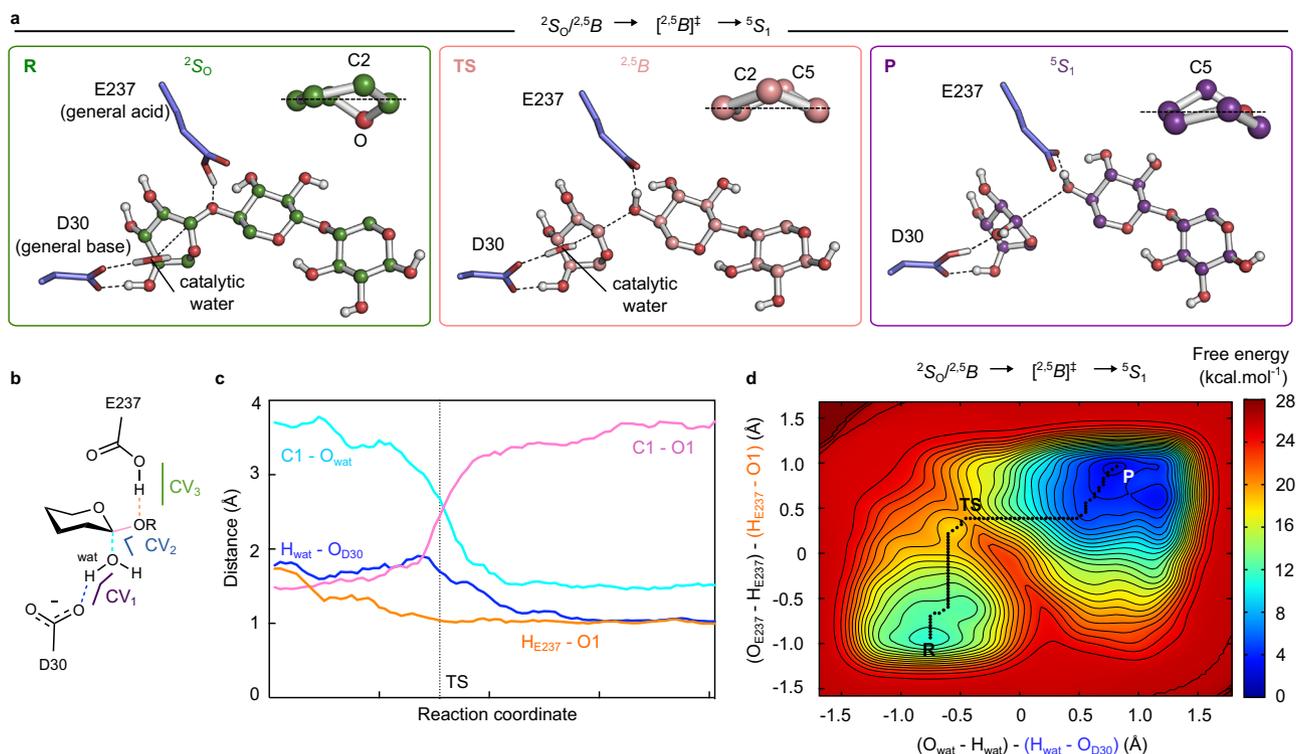


Fig. 7 The ${}^2S_0/{}^{2.5}B \rightarrow [{}^{2.5}B]^\ddagger \rightarrow {}^5S_1$ catalytic itinerary. **a** Representative snapshots of the reactant (R), transition state (TS) and product (P) states of the catalytic itinerary. A zoom of the -1 sugar conformation is represented at the top right of each state. **b** Schematic representation of catalytically relevant distances and collective variables (CV_1 , CV_2 , and CV_3) (wat = catalytic water). The CV_1 (purple) corresponds to the difference between $O_{\text{wat}}-H_{\text{wat}}$ and $H_{\text{wat}}-O_{\text{Asp30}}$ distances. The CV_2 (blue) was taken as the difference between $C1_{\text{xylosyl}}-O4_{\text{xylosyl}}$ and $O_{\text{wat}}-C1_{\text{xylosyl}}$ distances. The CV_3 (green) was taken as the difference between $O_{\text{Glu237}}-H_{\text{Glu237}}$ and $H_{\text{Glu237}}-O1_{\text{xylosyl}}$ distances. **c** Evolution of distances represented in **b** (running averages over five data points) along the reaction pathway. The distance between $H_{\text{E237}}-O1$ is represented in orange, the distance between $H_{\text{wat}}-O_{\text{D30}}$ in blue, $C1-O_{\text{wat}}$ in cyan and the glycosidic bond ($C1-O1$) in pink. **d** FEL of the catalytic reaction starting from ${}^2S_0/{}^{2.5}B$, projected into CV_1 ($(O_{\text{wat}}-H_{\text{wat}}) - (H_{\text{wat}}-O_{\text{Asp30}})$) and CV_3 ($(O_{\text{Glu237}}-H_{\text{Glu237}}) - (H_{\text{Glu237}}-O1_{\text{xylosyl}})$). The corresponding 3D FEL is shown in Supplementary Fig. 9. The minimum free energy pathway is indicated with a black dashed line on the two-dimensional FEL. Isolines are at intervals of 1 kcal mol^{-1} . Source data are provided as a source data file.

requirement of an oxocarbenium ion-like TS²³. 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 $1.29 \pm 0.02 \text{ \AA}$), 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 ($2.40 \pm 0.19 \text{ \AA}$) and the bond between the nucleophilic water and the anomeric carbon is partially formed ($2.72 \pm 0.15 \text{ \AA}$) (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 5S_1 conformation (P). Therefore, this conformational catalytic pathway of XacGH43_1 can be described as ${}^2S_0/{}^{2.5}B \rightarrow [{}^{2.5}B]^\ddagger \rightarrow {}^5S_1$. Interestingly, the conformation at P is in agreement with the one recently observed by Matsuzawa et al¹⁵ in the product complex structure (PDB ID 5GLN) of a β -xylosidase/ α -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 4C_1 conformation, i.e., the alternative minimum in the conformational FEL that was also observed in the crystallographic complex. This conformation, with an equatorially oriented leaving group, is expected to be either unreactive or much less reactive than the distorted ${}^2S_0/{}^{2.5}B$ conformation, in which the leaving group is axial and the xyloside ring is pre-activated for catalysis³². Nevertheless, the evolution of relevant reaction

distances along the resulting reaction pathway ${}^4C_1 \rightarrow [E_3]^\ddagger \rightarrow {}^4C_1$ (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 $17.5 \text{ kcal mol}^{-1}$, which is less favored than the one starting from a ${}^2S_0/{}^{2.5}B$ 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 (4C_1), although the TS conformation (E_3) is fully compatible with an oxocarbenium ion-like species²³. 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, ${}^2S_0/{}^{2.5}B \rightarrow [{}^{2.5}B]^\ddagger \rightarrow {}^5S_1$ and ${}^4C_1 \rightarrow [E_3]^\ddagger \rightarrow {}^4C_1$.

Interestingly, one of the two itineraries here described for XacGH43_1, ${}^2S_0/{}^{2.5}B \rightarrow [{}^{2.5}B]^\ddagger \rightarrow {}^5S_1$, was previously described for a GH43 enzyme from another subfamily (GH43_11 from *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 substrates²⁷. Even though alternative substrate conformations were not investigated by Barker and coworkers²⁷, the similarity of the active site cavity makes very likely that the conformational catalytic promiscuity obtained here are extensive to other GH43_11 members. These findings unfold an alternative

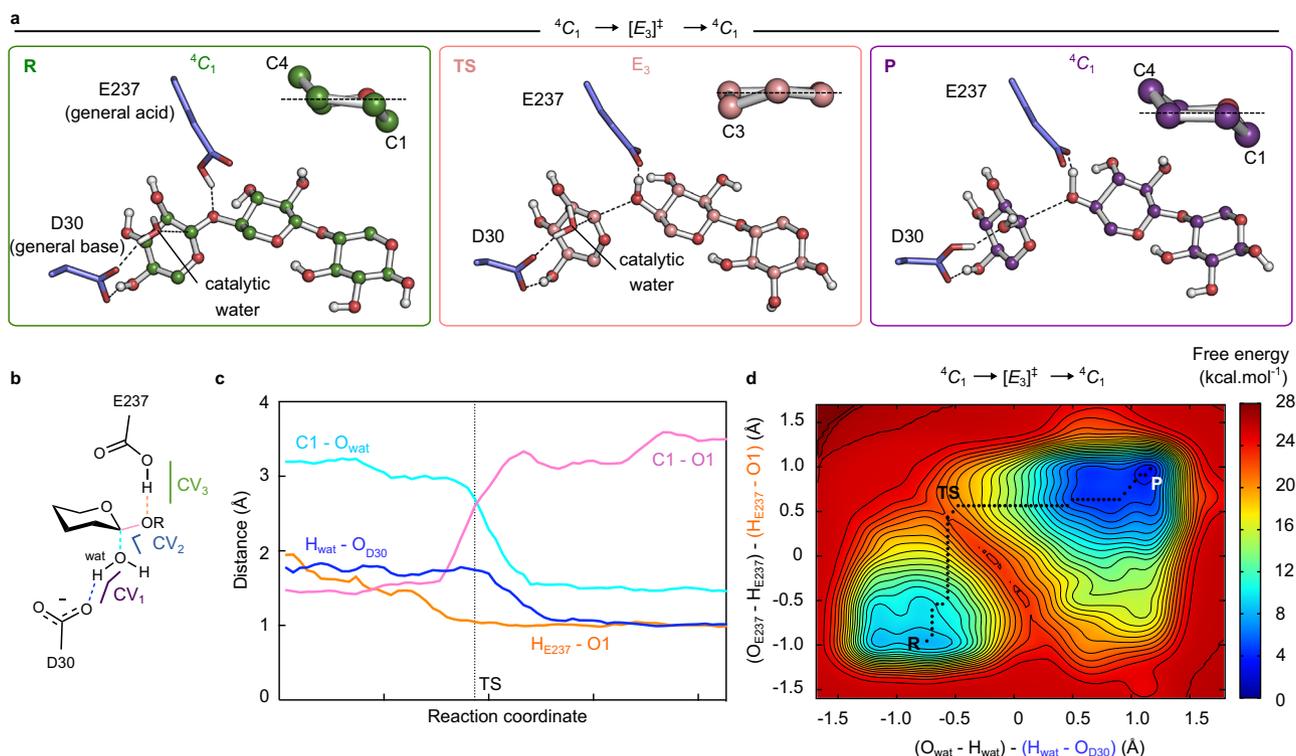


Fig. 8 The ${}^4C_1 \rightarrow [E_3]^\ddagger \rightarrow {}^4C_1$ catalytic itinerary. **a** Representative snapshots of the reactant (R) transition state (TS) and product (P) states of the catalytic itinerary. A zoom of the -1 sugar conformation is represented at the top right of each state. **b** Schematic representation of catalytically relevant distances and collective variables (CV_1 , CV_2 , and CV_3) (wat = catalytic water). The CV_1 (purple) corresponds to the difference between $O_{\text{wat}}-H_{\text{wat}}$ and $H_{\text{wat}}-O_{\text{Asp30}}$ distances. The CV_2 (blue) was taken as the difference between $C1_{\text{xylosyl}}-O4_{\text{xylosyl}}$ and $O_{\text{wat}}-C1_{\text{xylosyl}}$ distances. The CV_3 (green) was taken as the difference between $O_{\text{Glu237}}-H_{\text{Glu237}}$ and $H_{\text{Glu237}}-O1_{\text{xylosyl}}$ distances. **c** Evolution of these distances represented in **b** (running averages over five data points) along the reaction pathway. The distance between $H_{E237}-O1$ is represented in orange, the distance between $H_{\text{wat}}-O_{D30}$ in blue, $C1-O_{\text{wat}}$ in cyan and the glycosidic bond ($C1-O1$) in pink. **d** FEL of the catalytic reaction starting from 4C_1 , projected into CV_1 ($(O_{\text{wat}}-H_{\text{wat}}) - (H_{\text{wat}}-O_{\text{Asp30}})$) and CV_3 ($(O_{\text{Glu237}}-H_{\text{Glu237}}) - (H_{\text{Glu237}}-O1_{\text{xylosyl}})$). The corresponding 3D FEL is shown in Supplementary Fig. 10. The minimum free energy pathway is indicated with a black dashed line on the two-dimensional FEL. Isolines are at intervals of 1 kcal mol^{-1} . Source data are provided as a source data file.

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^{37,38} and GH10²⁹ 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 from X2 to X6³⁹. Thus, the exo-oligoxylanase activity could be possibly extended to other GH43_1 members.

As previously observed for other GH43 enzymes^{14,15,18,19,22}, Ca^{2+} ions activate XacGH43_1. We verified that this cation is responsible for enhancing substrate affinity and for increasing protein stability, which has not been observed in other calcium-activated GH43 enzymes¹⁴. Our MD simulations show that the Ca^{2+} ion is necessary for maintaining the active site in a favorable configuration for catalysis and that a Na^+ ion does not totally recover it. This is in accordance with the Na^+ ion impairing/decreasing activity in other GH43 enzymes in the absence of calcium^{14,15}. Our MD simulations support that the Ca^{2+} ion contributes to the binding of the substrate in three ways: (1) it stabilizes the proper conformation of the substrate at the -1 subsite; (2) it facilitates the interaction between the -1 sugar

and the pKa modulator Asp150; and (3) it controls the rotameric conformation of the catalytically relevant His288, preventing its interaction with the general base. This explains why mutations in the conserved active-site His impair the catalysis in calcium-activated GH43 members^{22,40}.

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 member³⁰. The -1 xylosyl residue in the XacGH43_1 complex showed two alternative conformations, 4C_1 and 2S_0 , 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 ${}^2S_0/{}^2,5B$ distorted conformation resulted in the catalytic itinerary ${}^2S_0/{}^2,5B \rightarrow [{}^2,5B]^\ddagger \rightarrow {}^5S_1$, which is in agreement with both the experimental free energy barrier and the trapped product of a cleaved X3 in the GH43_1 structure¹⁵ (PDB ID 5GLN). In fact, this itinerary is similar to the one previously reported for another subfamily of GH43 enzymes (${}^2,5B/{}^5S_1/B_{1,4} \rightarrow [{}^2,5B]^\ddagger \rightarrow {}^2,5B/{}^5S_1$)²⁷ 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 4C_1 conformation is also viable, with a free energy barrier just ≈ 3 kcal mol $^{-1}$ higher than the one computed for the ${}^2S_0/{}^2.5B$ conformation. The reaction proceeds in this case via a different TS conformation (E_3), delineating a “cyclic” itinerary, ${}^4C_1 \rightarrow [E_3]^\ddagger \rightarrow {}^4C_1$. 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 sugar²⁴. 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 oxocarbenium 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 imposing a notable decrease in enzyme performance. While this work was under review, another mechanistic study on a retaining exo-enzyme (the GH59 member β -galactocerebrosidase)⁴¹ also pointed out to alternative and viable conformations of the -1 sugar and leaving group. This reinforces the concept of catalytic conformational promiscuity proposed here and indicates that it is not strictly linked to the substrate type and might occur in other exo-acting glycosidases independently of the stereochemical outcome (retaining or inverting).

In summary, beyond the molecular elucidation of calcium role and the high-resolution structure of an active Michaelis complex prior to catalysis, this work demonstrates that some glycosidases are more flexible than others in binding their substrate for catalysis. Conformational promiscuity may be relevant for the design of TS-like inhibitors and activity-based probes for GHs, opening other venues in glycobiochemistry.

Methods

Cloning, protein production, and purification. The sequence encoding XacGH43_1 (NCBI accession code AAM39093) was amplified from genomic DNA of Xac using the following nucleotides: 5'-CATATGTCGGATGAACTGCAAGCC-3' and 5'-CTCGAG TCACCAAGCGCATCCGC-3' and cloned into pET28a expression vector after digestion with *NdeI* and *XhoI* restriction enzymes. The recombinant XacGH43_1 was expressed in *E. coli* BL21 (DE3) cells. The culture was grown at 37 °C in Luria-Bertani medium 1% (w.v $^{-1}$) tryptone, 0.5% (w.v $^{-1}$) yeast extract, 1% (v.v $^{-1}$ sodium chloride) containing kanamycin (50 μ g mL $^{-1}$) at 200 rpm until the O.D._{600 nm} reached values of ≈ 0.6 and incubated with 0.5 mM isopropylthio-D-galactopyranoside (Sigma-Aldrich, St. Louis, MO) for 16 h at 18 °C and 180 rpm. The cells were resuspended in lysis buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 5 mM imidazole, and 4 mM PMSF) 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 $^{2+}$ 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 imidazole (up to 0.5 M) at a flow rate of 1 mL min $^{-1}$. The eluted fractions were analyzed by SDS-PAGE⁴², 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 150 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 Zetasizer, 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 μ M in 20 mM Hepes pH 7.0 were used without additives (control) and with addition of 6 mM CaCl $_2$ 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-Boltzman 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 μ M) in 20 mM Hepes pH 7.5 with or without the addition of 6 mM CaCl $_2$. Measurements were performed using a 1-mm mica cell at the SAXS1 beamline from the Brazilian Synchrotron Light Laboratory (LNLS, Campinas, Brazil). Scattering data were recorded using a PILATUS 300 K (Dectris, Baden-Dattwil, Switzerland) and integrated using Fit2D⁴³. Data were processed using the ATSAS package⁴⁴. The program GNOM⁴⁵ was used to evaluate the pair-distance distribution function $p(r)$. The ab initio molecular envelope was calculated for the sample with calcium using DAMMIN⁴⁶. Averaged models were generated from several runs using DAMAVER⁴⁷. The theoretical scattering curve of XacGH43_1 crystallographic coordinates was calculated using CRY SOL⁴⁸. The XacGH43_1 crystal structure was fitted into the SAXS molecular envelope using the program SUPCOMB⁴⁹.

Enzyme assays. XacGH43_1 activity was tested against several natural and 4-nitrophenyl derived substrates (Supplementary Table 1). The 4-paranitrophenyl (pNP) derived substrates were purchased from Sigma-Aldrich (St. Louis, MO) and polymeric substrates were purchased from Megazyme (Wicklow, IE). Reactions with pNP-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 evaluated by estimating the reducing sugar released, according to the 3,5-dinitrosalicylic acid method⁵⁰. Initial activity tests were carried out with 2.5 mg mL $^{-1}$ (polysaccharides) or 2.5 mM (synthetic) of substrate, 40 mM mM McIlvaine buffer⁵¹ pH 7.0 and 25 μ g of enzyme (6.25 μ M). Reactions with pNP-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-Control software (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 pNP- β -xylopyranoside (pNP- β -Xyl) and 1 μ g (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 cations 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 CaCl $_2$ concentrations. Kinetics experiments using pNP- β -Xyl as substrate were performed at 40 °C with 0.5 μ g (125 nM) of enzyme for 10 min in 40 mM Hepes pH 7.0 with 6 mM CaCl $_2$. For enzyme assays against increasing concentrations of xylan, it was used 25 μ g (6.25 μ M) of enzyme in 40 mM Hepes pH 7.0 supplemented with 6 mM CaCl $_2$. The temperature was set to 40 °C and a reaction time of 120 min.

To analyze cleavage patterns by CZE, reactions with different XOS from X2 to X6 were carried at pH 7.0 at 40 °C, using 1 μ g (0.32 μ M) 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 1 M sodium cyanoborohydride and 6.5 mM APTS for 90 min at 60 °C. Samples were resuspended in running buffer (40 mM potassium phosphate pH 2.5) and analyzed in a P/ACE MDQ instrument (Beckman Coulter, Brea, CA) equipped with a laser-induced fluorescence detection module. A capillary of 50 μ m internal diameter and 12 cm effective length was used. Electrophoretic conditions were set to 20 kV mA with reverse polarity at 25 °C. The electrophoretic behavior of labeled products was compared to standards. CZE data were collected using the software 32 Karat 8.0 (Beckman Coulter).

The kinetic experiments of XOS were performed on a Waters Synapt HDMS (Waters Corp., Milford, CT) at V mode and ESI(+) with a spray voltage maintained at 3.0 kV and heated to 130 °C in the source⁵². Data were collected using the software MassLynx (4.1) (Waters). The samples were injected in scan mode (m/z 150–1100) with direct infusion at a flow rate of 20 μ L min $^{-1}$ and 1 Hz acquisition. An internal standard with ionization similar to analytes (mannotetraose—M4) was used to increase the accuracy of the method. To ensure that the reaction velocity is equal to the initial velocity, linearity tests were carried with different enzyme concentrations, for X6 and X2 (at 1 mM) in the presence and absence of CaCl $_2$ with different reaction times (0, 2.5, 5, 10, 20, and 30 min). The reaction time and enzyme concentration were selected in a linear region of the hydrolysis reaction. For reactions with X6 in the presence of CaCl $_2$, it was used 2.5 μ g mL $^{-1}$ (62.5 nM) XacGH43_1, whereas 5 μ g mL $^{-1}$ (125 nM) of the enzyme was employed in reactions without the addition of CaCl $_2$. For reactions with X2, we used 15 μ g 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 at pH 7.0, with or without 6 mM CaCl $_2$. After 5 min of reaction, 40 μ L methanol were added to stop the enzyme activity. A 15 μ L aliquot of the stopped reaction was added with 2 μ L mannotetraose (1 mM) and 183 μ L water and injected into the mass spectrometer. A calibration curve was obtained to determine the concentrations of the reaction products, plotting the intensity data of the reaction product divided by the internal standard intensity (I_{X1}/I_{M4}) versus the xylose concentration (Supplementary Fig. 11). Thus, for enzyme kinetics reactions, the values obtained from intensity I_{X1}/I_{M4} were converted to xylose concentration. The kinetic parameters (k_{cat} and K_M) were calculated using OriginPro (OriginLab Corporation, Northampton, MA). All quantitative enzyme assays were expressed as mean \pm 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, 30% (w.v⁻¹) polyethylene glycol 8000, 0.1 M sodium cacodylate pH 6.5 and 10% (v.v⁻¹) glycerol. To obtain sugar complexes, 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-Dattwil, Switzerland) and MXCube software⁵³. Diffraction data were scaled and reduced using XDS⁵⁴. The structure of XacGH43_1 was solved by molecular replacement using the program MOLREP⁵⁵ and coordinates of the β-xylosidase RS223BX¹⁷ (PDB ID 4MLG) as search model. Two molecules were found in the asymmetric unit and the generated model was refined using phenix.refine⁵⁶ and REFMAC5⁵⁷, with visual inspections and manual building using COOT⁵⁸. Final model was validated using MolProbity⁵⁹ and CheckMyMetal server⁶⁰. 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 software⁶¹. To prepare both the protein and the ligand (XacGH43_1 and X6, respectively), it was employed the Autodock Tools graphical interface⁶². 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 ligand size (a single grid box of 60.0 × 90.0 × 60.0 Å³, centered at $x = 0.5$ Å, $y = 6.5$ Å, and $z = 9.0$ Å). While the enzyme was described as rigid, the ligand was able to rotate each glycosidic bond. Sugar rings were considered as a rigid ⁴C₁ chair conformation, consistent with the X-ray and MD results obtained in this work. The exhaustiveness of search parameter was set at 10, the energy range at 10 kcal mol⁻¹, and the desired number of poses up to 20. This calculation was repeated 10 times starting with different random seeds in order to generate a total number of 200 binding poses.

Classical molecular dynamics simulations. The initial structure of XacGH43_1 for the MD simulations was taken from the X-ray structure reported in this work (complex XacGH43 + X3), considering the ²S₀ conformation of the active site xylosyl unit (nevertheless, during the simulation, the sugar sampled several conformations, as shown in Supplementary Fig. 7b). The missing loops in the crystal structure were constructed with Modeller⁶³. All crystallographic water molecules were retained and extra water molecules were added to form a 15 Å water box around the protein surface. For neutralizing the system charge, 13 sodium ions were added to the crystal structure with calcium (15 for the system where the calcium was replaced by a water molecule). The protonation state of all residues was assigned using PROPKA⁶⁴, considering pH 7.0. The protonation of His residues was further assessed according their chemical environment. Specifically, His residues 50, 71, 179, 246, 248, 262, 294, and 328 were considered neutral with their proton located at N_δ; His residues 34, 60, 73, 86, 161, 185, 203, 222, 231, 289, 303 and 313 were protonated at N_ε and His25 was double protonated. The general acid (Glu237) was protonated (pK_a ≈ 10 according to PROPKA⁶⁴, consistent with its role as proton donor during the reaction).

MD simulations were performed using Amber18⁶⁵. The protein was modeled with the FF14SB force-field⁶⁶, the X3 with the GLYCAM06 force-field⁶⁷, and water molecules were described with the TIP3P force-field⁶⁸. For MD, the systems were initially minimized, keeping both the protein and substrate fixed. Then, the entire systems have been allowed to relax. Afterwards, the system was heated gradually to 100, 200, and 300 K in the NVT ensemble at intervals of 50 ps. Spatial restraints to the protein and ligands were applied during the first heating interval, while all restraints were released after reaching 100 K. Subsequently, the density was converged up to water density at 300 K during 100 ps in the NPT ensemble, followed by a short MD of 100 ps in the NVT ensemble with a time step of 1 fs. The time step was then increased to 2 fs, employing the SHAKE algorithm⁶⁹ and the simulation was extended to 100 ns. We launched a total of three independent MD runs distributing random initial velocities during the first heating step. Analyses of the trajectories were carried out using standard tools of Amber and VMD⁷⁰. RMSDs evolutions of the 100 ns production runs are represented in the Supplementary Fig. 12. For analysis of sugar geometrical parameters, we used a documented script⁷¹.

QM/MM molecular dynamics simulations. One snapshot from the classical MD-equilibrated systems was taken for subsequent QM/MM calculations at the DFT level of theory. This snapshot had the xylosyl sugar in a ²S₀/²⁻⁵B conformation. The QM/MM calculations were performed using CPMD 3.15.1 (IBM Corp. and Max Planck Institute, Stuttgart, Germany) applying the method developed by Laio et al.⁷², that combines Car–Parrinello MD with force-field MD methodology. For the simulation of the conformational FEL, the QM region was composed by the xylosyl at the -1 subsite and half of the sugar at the +1 subsite, whose C2 and C4 were saturated with capping hydrogens. All other atoms of the

systems were included in MM region. Kohn–Sham orbitals were expanded in a plane-wave basis set with a kinetic energy cutoff of 70 Ry. Norm-conserving Troullier–Martins ab initio pseudopotentials⁷³ were used for all elements and calculations were performed using the Perdew, Burke, and Ernzerhoff generalized gradient-corrected approximation⁷⁴. This functional form has been proven to give a good performance in the description of hydrogen bonds^{75,76} and sugar conformations⁷⁷ and was already used with success in previous works on GHs and glycosyltransferases^{76,78,79}. 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.

Metadynamics of conformational free energy landscape (FEL). Metadynamics^{80,81} was used for obtaining the conformational FEL of the -1 sugar in the active site of XacGH43_1. The Cartesian projections of the Cremer–Pople of the puckering coordinates³³ were used as CVs (q_x and q_y)^{32,82}. The well-tempered metadynamics approach⁸³ was used within the metadynamics driver provided by the Plumed2 plugin⁸⁴. The height and the width of the Gaussian terms were set to 1.0 kcal mol⁻¹ and 0.1 Å, respectively, for q_x and q_y, with a deposition time of 200 MD steps (24 fs) and a bias factor of 10. The simulation was stopped once no quantitative changes in the FEL were observed, specially concerning energy differences among the two main minima. A total number of 2000 Gaussians were deposited. FEL representation was plotted with the program GnuPlot⁸⁵.

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). Monovalent pseudopotentials, located at the C_β atoms of Glu237 and Asp30 and at C3 and C5 of the +1 xylosyl, were used to saturate the QM region⁸⁶. Two metadynamics simulations were performed, starting from each of the two minima of the conformational FEL of the xylosyl unit, corresponding to the ²S₀/²⁻⁵B and ⁴C₁ conformations, respectively.

To drive the chemical reaction, we used three CVs that consider all covalent bonds that are to be broken or formed during the reactive process. The first CV (CV₁) was taken as the difference between H_{wat}-O_{wat} and O_{Asp30}-H_{wat} distances (wat = catalytic water). The second CV (CV₂) was taken as the difference between C1_{xylosyl}-O1_{xylosyl} and O_{wat}-C1_{xylosyl} distances, while the third CV (CV₃) was taken as the difference between H_{Glu237}-O_{Glu237} and O1_{xylosyl}-H_{Glu237}. The height and the width of the Gaussian terms were set to 1.0 kcal mol⁻¹, and 0.15 Å, respectively, for the three CVs, according to their oscillations observed in a free dynamics. The deposition time was set at 350 MD steps (42 fs). The simulations were stopped after all minima of the FEL were sampled, which involved 1673 and 1563 Gaussians, for reactions starting from the ²S₀/²⁻⁵B and from ⁴C₁ substrate conformations, respectively. In terms of metadynamics simulation time, this 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 pathway⁸⁷ in the reaction FEL was defined as a starting point for isocommittor analyses^{88,89}, which were performed on configurations/structures corresponding to metadynamics 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 reaches either the reactants or the products state. The putative TS for each reaction pathway 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 ⁴C₁ and 0.75 and 1.00 kcal mol⁻¹, for reaction starting from ²S₀) following the protocol proposed in the literature⁹⁰, successfully applied to glycosidases⁷⁶. Each simulation was initiated from the point in which ~50–60% of the reactant basin of our previous simulation was filled. The molecular mechanism remained the same independently of the Gaussian height used and the reaction starting from ⁴C₁ consistently showed slightly higher energy barrier values, for all Gaussian heights tested. These simulations resulted in free energy barriers of 12.7 and 15.4 for simulations starting from ²S₀ and 18.6 and 16.4 for simulations starting from ⁴C₁.

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 xylotriose 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.

Received: 9 July 2020; Accepted: 10 December 2020;

Published online: 14 January 2021

References

- Kohler, A. et al. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat. Genet.* **47**, 410–415 (2015).
- 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).
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, 490–495 (2014).
- Mewis, K., Lenfant, N., Lombard, V. & Henrissat, B. Dividing the large glycoside hydrolase family 43 into subfamilies: a motivation for detailed enzyme characterization. *Appl. Environ. Microbiol.* **82**, 1686–1692 (2016).
- Brüx, C. et al. The Structure of an inverting GH43 β -Xylosidase from *Geobacillus stearothermophilus* with its substrate reveals the role of the three catalytic residues. *J. Mol. Biol.* **359**, 97–109 (2006).
- Zanphorlin, L. M. et al. Structure-guided design combined with evolutionary diversity led to the discovery of the xylose-releasing exo-xylanase activity in the glycoside hydrolase family 43. *Biotechnol. Bioeng.* **116**, 734–744 (2019).
- Xu, B. et al. Characterization of a novel salt-, xylose- and alkali-tolerant GH43 bifunctional β -xylosidase/ α -L-arabinofuranosidase from the gut bacterial genome. *J. Biosci. Bioeng.* **128**, 429–437 (2019).
- Yang, X. et al. Two xylose-tolerant GH43 bifunctional β -xylosidase/ α -arabinosidases and one GH11 xylanase from *Humicola insolens* and their synergy in the degradation of xylan. *Food Chem.* **148**, 381–387 (2014).
- Wongratpanya, K. et al. Multifunctional properties of glycoside hydrolase family 43 from *Paenibacillus curdlanolyticus* strain B-6 including exo- β -xylosidase, endo-xylanase, and α -L-arabinofuranosidase activities. *BioResources* **10**, 2492–2505 (2015).
- Teeravittanakit, T. et al. Novel trifunctional xylanolytic enzyme Axy43A from *Paenibacillus curdlanolyticus* strain B-6 exhibiting endo-xylanase, β -D-xylosidase, and arabinoxylan arabinofuranohydrolase activities. *Appl. Environ. Microbiol.* **82**, 6942–6951 (2016).
- Huy, N. D., Thayumanavan, P., Kwon, T. H. & Park, S. M. Characterization of a recombinant bifunctional xylosidase/arabinofuranosidase from *Phanerochaete chrysosporium*. *J. Biosci. Bioeng.* **116**, 152–159 (2013).
- Proctor, M. R. et al. Tailored catalysts for plant cell-wall degradation: Redesigning the exo/endo preference of *Cellvibrio japonicus* arabinanase 43A. *Proc. Natl Acad. Sci. USA* **102**, 2697–2702 (2005).
- McKee, L. S. et al. Introducing endo-xylanase activity into an exo-acting arabinofuranosidase that targets side chains. *Proc. Natl Acad. Sci. USA* **109**, 6537–6542 (2012).
- Santos, C. R. et al. Mechanistic strategies for catalysis adopted by evolutionary distinct family 43 arabinanases. *J. Biol. Chem.* **289**, 7362–7373 (2014).
- Matsuzawa, T., Kaneko, S., Kishine, N., Fujimoto, Z. & Yaoi, K. Crystal structure of metagenomic β -xylosidase/ α -L-arabinofuranosidase activated by calcium. *J. Biochem.* **162**, 173–181 (2017).
- 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).
- Jordan, D. B. et al. X-ray crystal structure of divalent metal-activated β -xylosidase, RS223BX. *Appl. Biochem. Biotechnol.* **177**, 637–648 (2015).
- 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).
- Ahmed, S. et al. A novel α -L-arabinofuranosidase of family 43 glycoside hydrolase (Ct43Araf) from *Clostridium thermocellum*. *PLoS One* **8**, e73575 (2013).
- Cartmell, A. et al. The structure and function of an arabinan-specific α -1,2-arabinofuranosidase identified from screening the activities of bacterial GH43 glycoside hydrolases. *J. Biol. Chem.* **286**, 15483–15495 (2011).
- Alhassid, A. et al. Crystal structure of an inverting GH 43 1,5- α -L-arabinanase from *Geobacillus stearothermophilus* complexed with its substrate. *Biochem. J.* **422**, 73–82 (2009).
- 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 J.* **277**, 4562–4574 (2010).
- Davies, G. J., Planas, A. & Rovira, C. Conformational analyses of the reaction coordinate of glycosidases. *Acc. Chem. Res.* **45**, 308–316 (2012).
- Speciale, G., Thompson, A. J., Davies, G. J. & Williams, S. J. Dissecting conformational contributions to glycosidase catalysis and inhibition. *Curr. Opin. Struct. Biol.* **28**, 1–13 (2014).
- Wu, L. et al. An overview of activity-based probes for glycosidases. *Curr. Opin. Chem. Biol.* **53**, 25–36 (2019).
- Jiang, D. et al. Crystal structure of 1,3Gal43A, an exo- β -1,3-galactanase from *Clostridium thermocellum*. *J. Struct. Biol.* **180**, 447–457 (2012).
- Barker, I. J., Petersen, L. & Reilly, P. J. Mechanism of xylobiose hydrolysis by GH43 β -xylosidase. *J. Phys. Chem. B* **114**, 15389–15393 (2010).
- Dos Santos, C. R. et al. The mechanism by which a distinguishing arabinofuranosidase can cope with internal di-substitutions in arabinoxylans. *Biotechnol. Biofuels* **11**, 1–19 (2018).
- Santos, C. R. et al. Molecular mechanisms associated with xylan degradation by xanthomonas plant pathogens. *J. Biol. Chem.* **289**, 32186–32200 (2014).
- Isorna, P. et al. Crystal Structures of Paenibacillus polymyxa β -Glucosidase B complexes reveal the molecular basis of substrate specificity and give new insights into the catalytic machinery of family I glycosidases. *J. Mol. Biol.* **371**, 1204–1218 (2007).
- Nurizzo, D. et al. *Cellvibrio japonicus* α -L-arabinanase 43a has a novel five-blade β -propeller fold. *Nat. Struct. Biol.* **9**, 665–668 (2002).
- Iglesias-Fernández, J., Raich, L., Ardèvol, A. & Rovira, C. The complete conformational free energy landscape of β -xylose reveals a two-fold catalytic itinerary for β -xylanases. *Chem. Sci.* **6**, 1167–1177 (2015).
- Cremer, D. & Pople, J. A. A general definition of ring puckering coordinates. *J. Am. Chem. Soc.* **97**, 1354–1358 (1975).
- Biarnés, X., Ardèvol, A., Iglesias-Fernández, J., Planas, A. & Rovira, C. Catalytic itinerary in 1,3-1,4- β -glucanase unraveled by QM/MM metadynamics. Charge is not yet fully developed at the oxocarbenium ion-like transition state. *J. Am. Chem. Soc.* **133**, 20301–20309 (2011).
- Thompson, A. J. et al. The reaction coordinate of a bacterial GH47 α -mannosidase: a combined quantum mechanical and structural approach. *Angew. Chem. Int. Ed.* **51**, 10997–11001 (2012).
- Coines, J., Alfonso-Prieto, M., Biarnés, X., Planas, A. & Rovira, C. Oxazoline or oxazolium ion? The protonation state and conformation of the reaction intermediate of chitinase enzymes revisited. *Chem. Eur. J.* **24**, 19258–19265 (2018).
- Honda, Y. & Kitaoka, M. A family 8 glycoside hydrolase from *Bacillus halodurans* C-125 (BH2105) is a reducing end xylose-releasing exo-oligoxylanase. *J. Biol. Chem.* **279**, 55097–55103 (2004).
- Fushinobu, S. et al. Structural basis for the specificity of the reducing end xylose-releasing exo-oligoxylanase from *Bacillus halodurans* C-125. *J. Biol. Chem.* **280**, 17180–17186 (2005).
- Jordan, D. B., Stoller, J. R., Lee, C. C., Chan, V. J. & Wagschal, K. Biochemical characterization of a GH43 β -Xylosidase from *Bacteroides ovatus*. *Appl. Biochem. Biotechnol.* **182**, 250–260 (2017).
- McVey, C. E. et al. The importance of the Abn2 calcium cluster in the endo-1,5-arabinanase activity from *Bacillus subtilis*. *J. Biol. Inorg. Chem.* **19**, 505–513 (2014).
- Nin-hill, A. & Rovira, C. The catalytic reaction mechanism of β -galactocerebrosidase, the enzyme deficient in krabbe disease. *ACS Catal.* **10**, 12091–12097 (2020).
- Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nat. Publ. Gr.* **227**, 680–685 (1970).
- Hammersley, A. P. et al. Calibration and application of an X-ray image intensifier/charge-coupled device detector for monochromatic macromolecular crystallography. *J. Synchrotron Radiat.* **4**, 67–77 (1997).
- Franke, D. et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).
- Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).
- Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886 (1999).
- Volkov, V. V. & Svergun, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Cryst.* **36**, 860–864 (2003).
- Svergun, D., Barberato, C. & Koch, M. H. CRYSOLE - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
- Kozin, M. B. & Svergun, D. I. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41 (2001).
- Miller, G. L. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.* **31**, 426–428 (1959).
- McIlvaine, T. C. A buffer solution for colorimetric comparison. *J. Biol. Chem.* **49**, 183–186 (1921).
- Domingues, M. N. et al. Structural basis of exo- β -mannanase activity in the GH2 family. *J. Biol. Chem.* **293**, 13636–13649 (2018).
- Gabadiño, J. et al. MxCuBE: a synchrotron beamline control environment customized for macromolecular crystallography experiments. *J. Synchrotron Radiat.* **17**, 700–707 (2010).
- Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).

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. *Acta Crystallogr. Sect. D Struct. Biol.* **73**, 223–233 (2017).
61. Trott, O. & Olson, A. J. Autodock vina: improving the speed and accuracy of docking. *J. Comput. Chem.* **31**, 455–461 (2010).
62. Morris, G. M. et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).
63. Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinform.* **54**, 5.6.1–5.6.37 (2016).
64. Li, H., Robertson, A. D. & Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins Struct. Funct. Genet.* **61**, 704–721 (2005).
65. Case, D. A. et al. Amber 18 (University of California, San Francisco, 2018).
66. Maier, J. A. et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
67. Kirschner, K. N. et al. GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.* **29**, 622–655 (2008).
68. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
69. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
70. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
71. Biarnés, X. F., Rovira, C. & Sanz Carrasco, F. *Mecanismes de preactivació de substrat en 1,3-1,4-beta-glucanasa. Modelització mitjançant dinàmica molecular de primers principis. TDX (Tesis Doctorals en Xarxa)* (Universitat de Barcelona, Departament de Química Física and Centre de Recerca en Química Teòrica, 2008).
72. Laio, A., VandeVondele, J. & Rothlisberger, U. A hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations. *J. Chem. Phys.* **116**, 6941–6947 (2002).
73. Troullier, N. & Martins, J. L. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B* **43**, 1993–2006 (1991).
74. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
75. 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).
76. Raich, L. et al. A trapped covalent intermediate of a glycoside hydrolase on the pathway to transglycosylation. insights from experiments and quantum mechanics/molecular mechanics simulations. *J. Am. Chem. Soc.* **138**, 3325–3332 (2016).
77. 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).
78. Ardèvol, A. & Rovira, C. Reaction mechanisms in carbohydrate-active enzymes: glycoside hydrolases and glycosyltransferases. insights from ab initio quantum mechanics/molecular mechanics dynamic simulations. *J. Am. Chem. Soc.* **137**, 7528–7547 (2015).
79. Bilyard, M. K. et al. Palladium-mediated enzyme activation suggests multiphase initiation of glycogenesis. *Nature* **563**, 235–240 (2018).
80. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA* **99**, 12562–12566 (2002).
81. 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).
82. Biarnés, X. et al. The conformational free energy landscape of β -D-glucopyranose. Implications for substrate preactivation in β -glucoside hydrolases. *J. Am. Chem. Soc.* **129**, 10686–10693 (2007).
83. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100**, 1–4 (2008).
84. 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).
85. Williams, T. & Kelley, C. *gnuplot 5.0: An Interactive Plotting Program.* **7**, 259 (2017).
86. Zhang, Y., Lee, T. S. & Yang, W. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J. Chem. Phys.* **110**, 46–54 (1999).
87. Marcos-Alcalde, I., Setoain, J., Mendieta-Moreno, J. I., Mendieta, J. & Gómez-Puertas, P. MEPSA: minimum energy pathway analysis for energy landscapes. *Bioinformatics* **31**, 3853–3855 (2015).
88. Dellago, C., Bollhuis, P. G., Csajka, F. S. & Chandler, D. Transition path sampling and the calculation of rate constants. *J. Chem. Phys.* **108**, 1964–1977 (1998).
89. Peters, B. P(TP|q) peak maximization: necessary but not sufficient for reaction coordinate accuracy. *Chem. Phys. Lett.* **494**, 100–103 (2010).
90. 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**, 14148–14160 (2008).
91. Biely, P., Kratky, Z. & Vrsanska, M. Substrate-binding site of endo-1,4- β -xylanase of the yeast *Cryptococcus albidus*. *Eur. J. Biochem.* **119**, 559–564 (1981).
92. Davies, G. J., Wilson, K. S. & Henrissat, B. Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J.* **321**, 557–559 (1997).
93. Glatter, O. & Kratky, 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 Maria de Maeztu (MDM-2017-0767 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-2015-072055. 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 (LNBR) 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 the valuable scientific discussions.

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, 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 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.



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

© The Author(s) 2021