Discovery of processive catalysis by an exo-hydrolase with a pocket-shaped active site

Substrates associate and products dissociate from enzyme catalytic sites rapidly, which hampers investigations of their trajectories. The high-resolution structure of the native *Hordeum* exo-hydrolase HvExoI isolated from seedlings reveals that non-covalently trapped glucose forms a stable enzyme-product complex. Here, we report that the alkyl β-D-glucoside and methyl 6-thio-β-D-gentiobioside substrate analogues perfused in crystalline HvExoI bind across the catalytic site after they displace glucose, while methyl 2-thio-β-D-sophoroside attaches nearby. Structural analyses and multi-scale molecular modelling of nanoscale reactant movements in HvExoI reveal that upon productive binding of incoming substrates, the glucose product modifies its binding patterns and evokes the formation of a transient lateral cavity, which serves as a conduit for glucose departure to allow for the next catalytic round. This path enables substrate-product assisted processive catalysis through multiple hydrolytic events without HvExoI losing contact with oligo- or polymeric substrates. We anticipate that such enzyme plasticity could be prevalent among exo-hydrolases.
Enzymes are biological catalysts that are fundamental to life. Enzymes afford enormous accelerations to chemical reactions compared to uncatalysed reactions. It has become increasingly recognised that the largest contribution to the enzyme catalytic power arises from the electrostatic environment of the active site.

Enzymes use protein architectures to precisely position a set of amino acid residues to catalyse interconversions of substrates into products. This structure-based view is supported by thermodynamic and kinetic models that treat enzymes in complex with substrates and transition states in succession. Only rarely these models consider substrate and product diffusion as a part of their catalytic mechanisms, as these processes proceed rapidly, although on occasions products are seen entrapped in active sites. Yet, the lack of descriptions of substrate associative and product dissociative pathways creates a knowledge gap.

A vital conclusion drawn from the structural studies with the native β-1,6-glucan glucohydrolase, isoenzyme ExoI (HvExoI) isolated from barley seedlings, was that the glucose (Glc) product released from β-1,6-glucoside substrates remains entrapped in the enzyme active site until an incoming substrate binds, presumably lowering the energy barrier to facilitate Glc displacement. This stage, the mechanism of Glc displacement and how it is linked to the catalytic cycle of HvExoI remained unanswered. Notably, no other native GH3 structures with entrapped Glc are available, because these recombinant enzymes not exposed to presumably high enough Glc concentration, do not entrap Glc during protein maturation and secretion. However, several GH3 enzyme complexes with in crystalllo-perfused Glc or Glc-derivatives are available. Additionally, the native GH78 α-1-rhamnosidase with a deep pocket-shaped active site also holds the entrapped Glc molecule, which was not perfused in crystals.

We have previously observed that mechanism-based inhibitors conduritol B epoxide and 2',4'-dinitrophenyl 2-deoxy-2-fluoro-β-D-glucoside dislodge Glc from the −1 subsite and form stable cyclitol esters via an α-anomeric linkage with Asp285. Similarly, transition-state ion-like gluco-phenylmidaole mimics displace Glc and bind in the −1 subsite with their Glc component distorted in the 4E envelope conformation. Non-hydrolysable S-glycoside analogues also remove Glc from the −1 subsite and bind across the active site. Because HvExoI hydrolyses various cyclitol esters via an unencumbered.

In this work, we examine product and substrate pathways along the catalytic cycle of a plant exo-hydrolase HvExoI, and how this enzyme utilises its plasticity and that of glycoside substrates. We use the high-resolution X-ray crystallography supported with enzyme kinetics, mass spectrometry, nuclear magnetic resonance (NMR) spectroscopy, multi-scale molecular modelling employing docking, molecular dynamics (MD) simulations, Genetic Algorithms with Unrestricted Descriptors for Intuitive Molecular Modelling (GaudiMM) and Protein Energy Landscape Exploration (PELE) calculations, to reveal the Glc product displacement route, and to extract how each hydrolytic event including Glc release is precisely coordinated with the incoming substrate association and hydrolysis. This leads us to generate a variant enzyme and probe it for Glc entrapment and the S-glycoside analogue binding. Along this journey with HvExoI, we observe a remarkable phenomenon, the term ‘substrate-product assisted processive catalysis’ and describe how product and substrate trajectories are coordinated via this mechanism during catalysis. We discover that through chemical signalling, the incoming substrate evokes the formation of an autonomous and transient lateral cavity that serves as a conduit for the Glc product displacement from the active site. Crucially, these co-operative reactant pathways, where polymeric substrates never dissociate from the enzyme, sustain consecutive relocations of substrates and products via dedicated routes, ensuring that this exo-hydrolase adopts processive catalysis.

**Results**

**GC/MS proves that Glc is bound to crystalline native HvExoI.**

We purified, crystallised and chemically analysed crystals to reconcile our structural observations with Glc entrapment in native HvExoI purified from barley seedlings. Several crystals removed from the mother liquor were extracted and fractionated by normal-phase high performance liquid chromatography (HPLC) with evaporative light scattering detection (Fig. 1a; top panel, solid line). The mother liquor from these crystals was also analysed, both without and augmented with authentic Glc (Fig. 1a; top panel, dashed and dotted lines, respectively). The materials from these preparations eluting between 6.0 and 7.3 min during HPLC separation were pooled, reduced and peracetylated. The total ion chromatograms of the material extracted from crystals and from the mother liquor augmented with Glc showed a peak with the retention time of 23.26 min, corresponding to glucitol hexa-acetate (Fig. 1b; 1st and 3rd middle panels). Mass fragmentation spectra of these materials (Fig. 1c; bottom panels) showed identical profiles of total ions, demonstrating that the chemical substance eluting at 23.26 min was Glc. We could not detect Glc in the mother liquor, from which crystals had been removed (Fig. 1b; middle panel), while authentic Glc added to the same mother liquor was readily identifiable.

**Crystal structures reveal the Glc product in native HvExoI.**

The key observation that led us to explore the precise spatial disposition of the Glc product in the active site and the Glc displacement route in HvExoI, was based on our original observation that the Glc product remains entrapped in the enzyme’s active site that is isolated from the native source of young seedlings. We anticipated that the last remaining Glc molecule originating from oligo- and polysaccharide substrates, stays associated with the enzyme in plant tissues. To this end, previous crystal structures of native HvExoI at 2.20 Å and 2.70 Å (in-house X-ray source data) resolution showed a single Glc molecule bound in the −1 subsite. In the current work we refined native structures.
Fig. 1 Glc trapped in native HvExoI is bound non-covalently in the $^{4}\text{C}_{1}$ conformation. a HPLC chromatogram of material extracted from crystals (solid line), mother liquor (dashed) and mother liquor augmented with authentic Glc (dotted). Arrow points to the peak of authentic Glc in augmented mother liquor. Fractions (left right arrow) were pooled for GC/MS analysis. b Total ion chromatograms obtained by GC/MS of HPLC-eluted materials (6–7.3 min) containing alditol acetates. Materials from crystals (top), mother liquor (middle) and mother liquor augmented with Glc (bottom). Myo-inositol (20 ng) served as an internal standard (top and middle panels). Numbers near peaks indicate retention times. Substance eluted in 17.01 min peak corresponds to a contaminating plasticiser. Material extracted from crystals and eluting at 23.26 min (top panel) is identical to that of glucitol hexa-acetate at 23.26 min (bottom panel) on a Prevail Carbohydrate ES column (peaks indicated by asterisks). c Fragmentation mass spectra of glucitol hexa-acetate: top, material extracted from crystals; bottom glucitol hexa-acetate eluting at 23.26 min.
HvExoI at 1.65 Å resolution (synchrotron data), revealing that Glc is bound at 0.5 occupancy at each −1 and +1 subsites in the 4C1 chair conformation, suggesting that it may be mobile between both subsites (Fig. 2a; Supplementary Fig. 1a).

**Glc is absent in recombinant HvExoI but could be perfused in.** Although Glc has always been detected in the native enzyme isolated from barley seedlings6,8, it was uncertain if it could be observed in recombinant HvExoI produced in *Pichia pastoris*. The 1.45 Å structure of the recombinant enzyme, which is kinetically and structurally nearly identical to the native one24,25, failed to show the presence of Glc in the active site, where we identified glycerol and up to seven water molecules mimicking the positions of hydroxyl groups of Glc bound in native HvExoI (Fig. 3a; Supplementary Fig. 2a). However, after the crystals of recombinant HvExoI were perfused with Glc at a near saturating concentration, the 1.55 Å structure revealed two Glc molecules one each in the −1 and +1 subsites, adopting alternate 4C1 chair (occupancy 0.8) and 1S1 skew boat (occupancy 0.2) conformations with classical Cremer–Pople ring-puckering parameters26.

In the −1 subsite, we detected the network of 12–13 mono- and co-operative bi- and tri-dentate H-bonds27,28; of interest were those forming short H-bonds of 2.5–2.7 Å between O6β and O82 of Asp95, and the C6-OH and C4-OH groups of the Glc moiety. We attempted to position into the electron density other skew boat conformers of Glc (3S1, 1S1) with various occupancies, however, convergence in the refinement was reached with those of 4C1 and 1S1. We concluded that during maturation in *P. pastoris*, Glc cannot be entrapped and thus observed, because the intracellular concentration in this host is not high enough during enzyme maturation or secretion. To demonstrate that Glc was not displaced from the active site by the cryoprotectant, data were collected from recombinant crystals at ambient temperatures without applying vitrification solutions. Here, no Glc was observed in the active site, but was readily observable after perfusion of crystals with Glc.

**SPR analysis of Glc and substrate binding to HvExoI.** To determine the strength of Glc binding and to compare it with that of the substrate thio-analogue methyl 6-thio-β-gentiobioside (G6SG-OMe), we used the recombinant HvExoI enzyme. The steady-state affinity29 $K_D$ value of 0.008 × 10$^{-3}$ M for G6SG-OMe revealed that it bound tighter than the Glc hydrolytic product ($K_D$ of 0.16 × 10$^{-3}$ M) (Table 1; Supplementary Fig. 3).

**NMR spectroscopy of Glc binding to recombinant HvExoI.** To identify conformational states of bound Glc with the measured $K_D$ value of 0.16 × 10$^{-3}$ M (Table 1; Supplementary Fig. 3), we used $^1$H saturation transferred difference (STD)30 and transferred nuclear Overhauser effect spectroscopy (trNOESY)31 in solution NMR. STD spectra recorded with Glc in a 60-fold excess relative to HvExoI detected Glc bound non-competitively. As only STDs for protons of β-anomeric Glc (and not for α-form) were observed, while one would expected an α:β anomeric ratio of ∼2:3 in solution equilibrium, we concluded that the enzyme specifically binds β-anomeric Glc (Fig. 4a); this observation supports the retaining hydrolytic mechanism of HvExoI22. To identify the conformation of Glc, trNOESY experiments under a relatively low HvExoI versus Glc ratio revealed weak trNOEs between H1−H3 and H1−H5 protons of β-anomeric Glc (Fig. 4b), consistent with β-anomeric Glc bound in the low-energy 4C1 chair conformation. To corroborate these findings, thioceillose which cannot be hydrolysed by HvExoI7, was added in excess relative to HvExoI. Here, clear negative NOEs between specific thioceillose proton pairs were observed (Fig. 4c), while strong trNOEs of the
maps are contoured at 1 Å in the Stereo view of recombinant HvExoI with two Glc molecules in the sticks) at 0.7 occupancy, bound across the lines. Derived from the active site residues (carbons: grey sticks) are shown as dashed Water molecules are shown as red spheres. Separations of less than 3.50 Å three alternate conformations) are bound in the glycerol molecules (carbons: green sticks) (Gol 1 at occupancy 0.5; Gol 2 in conformation (carbons: cyan sticks) at 0.2 occupancy, bound in the Reccominant in ligand-free form.

Table 1 Binding (K_D) and inhibition (K_i) parameters for HvExoI by ligands and substrate analogues

| Ligand/inhibitor                  | K_D/K_i (M) | \( \Delta G \) (kJ mol\(^{-1}\)) | Chemical formula \(^a\) |
|----------------------------------|-------------|-----------------------------------|-------------------------|
| Glc\(^a\)                        | 0.16 × 10\(^{-3}\) | -21.7                             | C\(_6\)H\(_{12}\)O\(_6\) |
| G65G-OMe\(^b\)                   | 0.008 × 10\(^{-3}\) | -29.1                             | C\(_{32}\)H\(_{62}\)O\(_{26}\)S |
| 3dGlc\(^b\)                      | 9.8 × 10\(^{-3}\)  | -11.7                             | C\(_6\)H\(_{12}\)O\(_5\) |
| 4dGlc\(^b\)                      | 9.4 × 10\(^{-3}\)  | -11.8                             | C\(_6\)H\(_{12}\)O\(_5\) |
| Octyl-O-Glc\(^b\)                | 0.13 × 10\(^{-3}\) | -22.6                             | C\(_{14}\)H\(_{26}\)O\(_6\) |
| Octyl-S-Glc\(^b\)                | 1.1 × 10\(^{-3}\)  | -17.2                             | C\(_{14}\)H\(_{26}\)O\(_5\) |
| G25G-OMe\(^b\)                   | 2.55 × 10\(^{-3}\) | -15.1                             | C\(_{13}\)H\(_{22}\)O\(_{10}\)S |

\(^a\)Using Surface Plasmon Resonance with recombinant HvExoI
\(^b\)Using inhibition kinetics with the 4-nitrophenyl \( \beta \)-D-glucopyranoside substrate\(^3\) with native HvExoI
\(^c\)Calculated according to \( \Delta G = -RT \ln K_i \) or \( \Delta G = -RT \ln K_D \)

Conformational Free Energy Landscape of Glc bound to HvExoI. To further assess conformational states of bound Glc in the high-resolution structure of recombinant HvExoI in complex with Glc, quantum mechanics/molecular mechanics (QM/MM) metadynamics simulations were performed to assess Conformational Free Energy Landscape (FEL) maps of \( \beta \)-D-Glc bound in the active site of HvExoI (Fig. 5; Supplementary Fig. 2b). These analyses indicated that although Glc was engaged in extensive H-bond networks, the \( \beta \)-D-glucopyranose ring adopted \(^4\)C\(_1\) chair and \(^1\)S\(_3\)/B\(_{3,0}\) (−1 subsite) or \(^4\)C\(_1\) chair and B\(_{3,0}\) (+1 subsite) distorted skew boat/boat conformations (Fig. 5; left and middle panels). More precisely in the −1 subsite, H-bonds with Asp285 and Arg158 restrained the 2-OH group in the equatorial position, and stabilised the \(^4\)C\(_1\) conformation by 8 kcal/mol more than that of \(^1\)S\(_3\) or B\(_{3,0}\) with a conformational energy barrier change of 11 kcal/mol. Conversely, hydroxyl groups of Glc in the +1 subsite interacted with Asp95, Arg158, Lys206, His207, Asp285, and Glu491, stabilising the \(^4\)C\(_1\) Glc conformation by 2 kcal/mol more than that of B\(_{3,0}\) with a conformational energy barrier change of 6 kcal/mol. Notably, these findings concurred with conformational FEL of isolated Glc\(^3\), in which \(^4\)C\(_1\) and \(^1\)S\(_3)/B\(_{3,0}\) conformers were most stable (Fig. 5; right panel). Nonetheless, FEL of Glc bound in the −1 and +1 subsites exhibited exclusively these two minima due to a restricted conformational space within the enzyme active site (Figs. 2a, 3b, 5) as demonstrated in other glycosyl hydrolases\(^4\).

Inhibition kinetics of substrate analogues in native HvExoI. To shed light on the molecular basis of the Glc displacement route, we selected a series of deoxy-Glc (dGlc) and alkyl-glucoside derivatives, and substrate mimics that could potentially displace Glc\(^6\). To assess the strength of inhibition, we determined dissociation constants of enzyme-inhibitor complexes (K_i) for 3-deoxy-glucose (3dGlc), 4-deoxy-glucose (4dGlc), n-octyl \( \beta \)-D-glucopyranoside (octyl-O-Glc), n-octyl 1-thio-\( \beta \)-D-glucopyranoside (octyl-S-Glc), a hydrophilic polymer polyethylene glycol (PEG; putative fractional polysaccharide mimic\(^3\)) and the thioglycoside methyl 2-thio-\( \beta \)-sophoroside (G25G-OMe). All inhibitors showed competitive inhibition with K_i values between 9.8 × 10\(^{-3}\) M and 0.13 × 10\(^{-3}\) M (Table 1). The most effective was octyl-O-Glc with K_i of 0.13 × 10\(^{-3}\) M, while surprisingly G25G-OMe was weak with K_i of 2.55 × 10\(^{-3}\) M, while PEG was non-inhibitory, as expected.

Glc is not removed from native HvExoI by dGlc derivatives. To test the hypothetical Glc displacement route by using dGlc derivatives, crystals were perfused with saturating concentrations of

[Image 54x304 to 287x394]
3dGlc (Fig. 2b; Supplementary Fig. 1b) and 4dGlc. The structures of both complexes showed that dGlc derivatives could not displace Glc from the $-1$ subsite, but that Glc consolidated to the $-1$ subsite. The well-defined electron density map indicated that 3dGlc was orientated parallel to Trp286 and Trp434 in the $+1$ subsite, through $\pi$-system stacking contacts, forming H-bonds via C6-OH with Ne1 of Trp434 and Oe2 of Glu491, and via C4-OH with the N atom of Gly57 (Fig. 2b; Supplementary Fig. 1b). Notably, as 3dGlc rotated to a different position relative to that of Glc in native HvExoI (such that the C4-OH group of 3dGlc overlapped the position of C3-OH of Glc in the native structure), it could establish H-bonds with Gly57 and Glc in the $-1$ subsite.

Fig. 4 Recombinant HvExoI recognises $\beta$-D-Glc in the $4C_1$ conformation. a $^1$H STD NMR spectrum (top) and reference spectrum (below), acquired with recombinant HvExoI and Glc (40 $\mu$M and 2.4 mM, respectively) at 600 MHz and 283 K. Only protons of $\beta$-Glc (squares) showed STDs. Off and on resonance frequencies were set at 100 ppm and 0.65 ppm, respectively. The Gaussian shaped pulse (30 ms) was used for selective irradiation with total saturation time of 2 s. b trNOESY spectrum of recombinant HvExoI and Glc (57 $\mu$M and 285 $\mu$M), acquired at 800 MHz, 283 K and with 300 ms mixing time. Only NOEs observed were weak trNOEs of Glc between H1/H3 and H1/H5. c trNOESY spectrum of recombinant HvExoI with Glc and thiocellobiose (57 $\mu$M, 285 $\mu$M, and 171 $\mu$M, respectively), acquired at 800 MHz, 283 K and with 300 ms mixing time. All NOEs were negative. trNOEs defining the $4C_1$ conformation of non-reducing Glc of bound thiocellobiose are in boxes. Blue lines in (b) and (c) refer to residual noise signals.
Alkyl-glucosides and PEG remove Glc from the active site. Contrary to what we observed with dGlc derivatives, n-octyl β-D-glucosides with tighter $K_d$ values (Table 1) replaced Glc and bound across the active site (Fig. 2c; Supplementary Fig. 1c). While aliphatic chains of alkyl β-D-glucosides were threaded through the Trp286 and Trp434 aromatic clamp and did not interact with a surrounding environment, Glc moieties established a network of 10–11 mono- and co-operative bi- and tridentate H-bonds, like those observed in native (with trapped Glc) or recombinant Glc-perfused HvExoI (Figs. 2, 3). Similarly, after applying PEG at saturating concentrations to native HvExoI crystals, the 1.80 Å PEG-perfused structure revealed that Glc was replaced by five water molecules occupying the positions of the C1-OH to C6-OH groups of Glc in the –1 subsite (Fig. 2d; Supplementary Fig. 1d). The PEG molecule bound between Trp286 and Trp434 in two alternate conformations made a water-mediated H-bond with Gly57, which was reminiscent of the bond formed between C4-OH of 3°Glc and Gly57. The second PEG molecule near Trp434 did not contact the enzyme, but instead interacted with the first PEG via water-mediated H-bonds (Fig. 2d; Supplementary Fig. 1d).

G6SG-OMe and G2SG-OMe bind to HvExoI in two distinct modes. To understand how (1,6)- and (1,2)-linked substrate thioanalogues displace the Glc product from the active site, we pursued recombinant crystals of HvExoI and observed that the sugars adopted different poses (Fig. 3c, d; Supplementary Fig. 2c, d). It is worth noting that G6SG-OMe binds to the enzyme with $K_d$ of $0.008 \times 10^{-3}$ M, whereas G2SG-OMe binds about 320-fold weaker ($K_d$ of $2.55 \times 10^{-3}$ M) (Table 1). We observed that in the 1.57 Å structure of HvExoI with G6SG-OMe, the ligand formed well-defined electron densities for both Glc moieties across the active site (Fig. 3c; Supplementary Fig. 2c), and that the 1.68 Å structure with G2SG-OMe (refined at 0.7 occupancy at both subsites) had a less-defined density for the ligand, which bound to the enzyme in a markedly different position (Fig. 3d; Supplementary Fig. 2d). For G6SG-OMe, we observed short H-bonds in the –1 subsite, which resembled those observed in the native (Fig. 2a; trapped Glc) or recombinant (Fig. 3a; perfused with Glc) enzymes. In contrast, G2SG-OMe could not slide past the +1 subsite in the pocket, likely due to a thio-glycosidic bond rigidity, and hence it projected with the reducing-end Glc moiety into the solvent at the putative +2 subsite, while the non-reducing-end was held in the adjacent +1 subsite. For the –1 subsite, we assigned glycerol and PEG molecules that formed H-bonds with the G2SG-OMe non-reducing end that also established the H-bond with Gly57; the reducing moiety of G2SG-OMe interacted via C4-OH with Oe2 of Glu287 and via C3-OH with Nε1 of Trp286 (Fig. 3d; Supplementary Fig. 2d).

In the G6SG-OMe structure we noted significant elevations of B-factor values in two loops (Thr214-Glu228 and Glu491-Asn498) near the catalytic site entry (Supplementary Note 2); these B-factor elevations were not seen in the G2SG-OMe structure.

HvExoI-G2SG-OMe complex is an intermediate for Glc exit path. We considered the HvExoI in complex with G2SG-OMe (PDB 6MD6) (Fig. 3d; Supplementary Fig. 2d) to embody the attributes of an intermediate enzyme-substrate complex, and the disposition of G2SG-OMe that of an incoming substrate. Hence, we used this complex for investigations of the hypothetical Glc displacement route via multi-scale molecular modelling employing docking$^{19}$ and MD simulations, followed by GaudiMM$^{20}$ and PELE$^{21}$ pathway calculations. Respective reciprocal docking of the Glc product and the β-D-glucopyranosyl-(1,2)-D-glucose (G2OG) or β-D-glucopyranosyl-(1,3)-D-glucose (G3OG) substrates into HvExoI:G2SG-OMe or HvExoI:Glc complexes, combined with MD simulations indicated that the existence of ternary HvExoI:Glc:G2OG or HvExoI:Glc:G3OG complexes (designated 1–6) was plausible, where Glc was bound in the –1 subsite, and G2OG or G3OG were attached at +1 and putative +2 subsites (Fig. 6a; Supplementary Figs. 5, 6, 8, Supplementary Notes 3–4, Supplementary Data 1). This suggested that if the incoming substrate binds at the +1 to +2 subsites, while Glc is still trapped in the –1 subsite, an alternative Glc exit path, other than through the +1 subsite needs to be considered. We also searched for substrate binding modes different to those at the +1 to +2 subsites, using docking of the G2OG or G3OG substrates in structures based on MD simulations of the native HvExoI:Glc complex (Supplementary Figs. 5–8; Supplementary Note 4; Supplementary Data 1–2). However, no stable binding sites were found other than those at +1 and +2 subsites. In addition, docking and MD simulations...
unequivocally revealed that the HvExoI:Glc:G2OG and HvExoI:Glc:G3OG complexes (Fig. 6a) converged and were stable, and could be used for investigations of Glc displacement. MD simulations also revealed that stable binding of substrates required Trp434 adopting the parallel orientation to that of Trp286, where C-H/π interactions mediated binding. Notably, binding of the G2OG or G3OG substrates triggered the conformational change of Tyr253, modifying the buried lateral cavity adjacent to the −1 and +1 subsites. We observed that when only one Glc was bound at the −1 subsite, Tyr253 remained H-bonded to the carbonyl oxygen of the Trp286 backbone for the full extent of simulation. The sidechains of the catalytic nucleophile Asp285 and acid/base Glu491, together with Arg158 (Fig. 6a; dashed lines) physically separated this cavity from the −1 subsite. In complexes 1, 3 and 4, the Tyr253 side chain moved away from this lateral cavity enlarging it, while in complex 2,
this side chain rotated to the bottom of the cavity making it shallower (Supplementary Figs. 6, 8; Supplementary Tables 3–4).

To find out if binding of substrates by HvExoI lacking the bound Glc product would lead to tighter or weaker binding than that with Glc included, we carried out further calculations and compared binding affinities. Docking calculations of G2OG, G3OG and β-β-glucopyranosyl-(1-6)-β-glucose (G6OG) disaccharide substrates at the −1 and +1 subsites, predicted binding with higher affinities (Goldscore scoring function values of 66 for G2OG, 76 for G3OG and 74 for G6OG), when the Glc product was absent in the −1 subsite. However, when the Glc product was included in the −1 subsite, docking of G2OG, G3OG and G6OG in the +1 and putative +2 subsites predicted binding with lower affinities (Goldscore scoring function values of 60 for G2OG, 57 for G3OG and 61 for G6OG). These data indicated that bound Glc lowered binding energies for incoming substrates as they had no access to the higher affinity −1 subsite.

Discovery of substrate-product assisted processivity in HvExoI.

Initially, we used the GPathFinder extension of the GaudiMM platform20 to reveal potential Glc exit routes from the active site into the bulk solvent, whereby we considered steric clashes. The initial set of simulations was performed with the HvExoI:Glc: G2OG ternary complex 1, based on the coordinates of HvExoI in complex with G2SO-OMe and Glc docked in the −1 subsite. No exit path was detected in this case, as Glc never overcame steric clashes to vacate the −1 subsite (Fig. 6b; top panel; Supplementary Data 3). However, further calculations with the protein backbone displaced by 2.2 Å along the lowest-energy normal mode strikingly identified the Glc displacement path. As a bonus, these calculations also revealed a cork-like motion between domains 1 and 2, which was correlated with the open and closed states of the active site (Fig. 6c). The same ligand path was also obtained with the MD-derived HvExoI:Glc:G3OG ternary complex 3 (Fig. 6b; bottom panel), highlighting the importance of incorporating local and collective protein motions in these calculations.

To explore the Glc displacement route in detail, the ligand migration calculations of the PELE approach21 were used, whereby structures generated at each step along the path (by sequential ligand and protein geometric perturbations followed by energy minimisation using the OPLS-AA force field) were evaluated and accepted or rejected according to a Metropolis criterion at a given temperature. Using this approach, the HvExoI:Glc:G2OG and HvExoI:Glc:G3OG complexes 1–3 converged, and revealed that Glc egressed from the −1 subsite to the adjacent lateral cavity that was enlarged after Tyr253 rotated. This cavity was defined by an ensemble of 14 residues: Trp156, Arg158, His207, Phe208-backbone, Asp211, Asn219, Glu220, Ser252, Tyr253-backbone, Ser254, Asp285, Arg291, Glu491 and Thr492 (Fig. 6; Supplementary Fig. 9). This cavity was partially separated from the solvent by the Asn219, Glu220 and Arg291 sidechains, with the two latter residues forming a salt bridge with Or1 and Or2 of Glu220 to NH1 of Arg291 at separations of 3.10 Å/2.65 Å, 2.77 Å/3.56 Å and 4.65 Å/2.84 Å for complexes 1, 2, and 3, respectively. The abundance of data points in the PELE plot at separations between 8–11 Å (Fig. 6d) indicated that there was a local energy minimum for Glc binding in this cavity (Fig. 6d; subpanel g). For complex 2, PELE calculations suggested a more transient passage of Glc through the lateral cavity, most likely due to a shallower profile resulting from the specific Tyr253 conformation, compared to other complexes. Nevertheless, in all cases, if Glc was to advance to the lateral cavity from the −1 subsite, it must traverse the space between Asp285, Glu491 and Arg158, which likely represent a toll-like barrier (Fig. 6a, dashed lines; Fig. 6e); this Glc passage corresponded to fewer data points at separations between 5–7 Å (Fig. 6d; Supplementary Data 4). With Glc bound in the −1 subsite, the shortest separations between the Arg158-Asp285, Arg158-Glu491 and Glu491-Asp285 sidechains were those at 5.10 Å/5.42 Å/5.17, 3.09 Å/3.16 Å/4.42 Å and 5.39 Å/6.32 Å/6.65 Å in respective complexes 1, 2 and 3. Notably, conformational changes of Arg158, Asp285 and Glu491 sidechains facilitated Glc movement (Fig. 6e, f), and altered the separations specified above to 7.00 Å/5.33 Å/6.21 Å, 3.00 Å/2.97 Å/6.82 Å and 6.96 Å/7.23 Å/7.71 Å, when Glc bypassed the toll-like barrier. Once in the lateral cavity, Glc was free to exit from the cavity into the bulk solvent via a transient and autonomous aperture. This transient opening was formed through rotations and backbone fluctuations of Glu220, Arg291, Thr492 and Lys493, and surrounding residues in the vicinity of the bound β-β-glucoside molecule (Fig. 6h). As Glc migrated across the lateral cavity (Fig. 6e, f), it established H-bonds39 with the protein residues and incoming substrates to maintain energetic favorability (Supplementary Tables 2–4). This suggested that the hydrophilic environment of the toll-like barrier and the lateral cavity may have evolved for this exit route and may be evolutionarily conserved in GH3 enzymes (Supplementary Fig. 9). Finally, and most importantly, the Glc displacement route raised the possibility that this trajectory may facilitate processive catalysis in HvExoI and other GH3 exo-hydrolases with pocket-shaped active sites.

Proof of concept using the R158A/E161A variant.

Ligand migration calculations using PELE (Fig. 6) and the conservation patterns of the toll-like barrier (Supplementary Fig. 9) led us designing the double non-conservative R158A/E161A variant to critically asses the roles of these residues in binding Glc and the
G6S-GOMe substrate thio-analogue (Fig. 7; Supplementary Fig. 10). The 1.65 Å and 2.21 Å structures without and with perfused Glc showed two glycerol molecules, one each in the −1 and +1 subsites (Fig. 7a, b; Supplementary Fig. 10a, b), and the absence of Glc despite perfusing crystals at saturating concentrations. On the other hand, the 2.30 Å structure of R158A/E161A perfused with G6S-GOMe strikingly revealed this sugar bound across the active site, but in a different pose compared to that of the wild-type (WT) enzyme (cf. Figures 3c, 7c; Supplementary Figs. 2c and 10c). In the variant structure the position of the Glc moiety in the −1 subsite matched that of WT, but the disposition of the reducing-end Glc moiety was flipped in its position between the indole moieties of Trp286 and Trp434, such that intra-ring oxygen of the β-D-glucopyranosan moeity pointed to Trp434. Obviously, this G6S-GOMe pose was due to Ala replacements for Arg158 and Glu161.

Discussion

Substrates and products in enzyme active sites bind and unbind at fast rates that challenge investigations of their trajectories, leading to the lack of a deep understanding of these hallmarks of enzyme catalysis. Here, we capitalise on the Glc product entrapment observation in native HvExoI isolated from seedlings, which renders it an archetype model to examine product/substrate dissociations/associations in a pocket-shaped active site. Most approaches towards the descriptions of reactant movements are based on MD simulations of substrate and product binding and unbinding, including random collisions or diffusion. Other studies reveal substrate/product migration pathways that accompany conformational changes of protein backbones or secondary structures.

When we first described the HvExoI structure, we were unable to explain why the trapped Glc product remained bound in the active site and has not diffused away, and what was the implication of this unassuming observation. The obvious explanation for not seeing naturally bound products or co-factors in structures of other GH enzymes is that these proteins are generated in recombinant hosts, where intracellular concentrations of potential enzyme reactants are not high enough during protein maturation. To this end, solving crystal structures of enzymes purified from native sources offers an additional information that could be beneficial for the understanding of catalytic cycles.

When we first detected the Glc product entrapment in native HvExoI, we hypothesised that the role of Glc may be linked to the pre-organised state of the active site to maintain efficient catalysis required for the growth of a plant embryo. One way of preserving this pre-organised state would be to keep entropic costs of catalysed reactions low, where an entrapped product from a previous hydrolytic cycle could lower overall entropic demands for binding of incoming substrates. The Glc product retention in the active site may also disfavour binding of incorrect substrates through product dissociation rates that would govern the selectivity of substrate binding. Others may argue that although most enzymes have evolved to use conformational adjustments to favour tight binding of correct substrates and product release, some enzymes may have not acquired this asset. This assumption was confirmed by perfusing dGlc derivatives that akin of reaction products could not remove Glc, but substrate analogues and mimics could.

As mentioned above, despite the plethora of structural data for HvExoI, we could not explain, how Glc egresses from the 13 Å-deep pocket-shaped active site, although we detected cross-talk between residues entrapping Glc in the −1 subsite, and those in the +1 subsite delineated by Trp286 and Trp434. Function of Trp residues in retaining Glc or substrates is unsurprising, as these residues have extended heterocyclic indole ring systems with amphipathic characteristics, high de-localised electron densities and permanent dipole moments due the intra-ring nitrogen(s), that form H-bonds. In all 22 structures of HvExoI we so far resolved (Supplementary Table 5), Trp286 and Trp434 at the +1 subsite showed no structural heterogeneity, and at separations of about 4 Å from ligands were compliant with the provision of two sets of C-H/π interactions depending on the stereochemistry of bound ligands. Crystalline HvExoI was also perfused with the (3)β-D-Glc-S-(1,3)-β-D-Glc-(1)14-22 polymeric substrate, but we did not observe potential binding sites beyond the −1 and +1 subsites.

The time-space averaged vision of crystal structures presented here was supported by the multi-scale molecular modelling, based on docking, MD simulations, QM/MM, GaudiMM and PELE, that provided the first clear view of succession of events during Glc displacement. This approach also hinted that trapped Glc may play a surprising role in catalysis and led to the discovery of processivity by this exo-hydrolase. In the light of this discovery we coin the term ‘substrate-product assisted processive catalysis’, due to the key role that a substrate and a product play in the evocation of the catalytic pathway.
We demonstrate that the substrate binding and product displacement routes proceed through stages that are carefully orchestrated in succession (Fig. 8). We suggest that a disaccharide (Fig. 8a) binds near the catalytic site at the +1 and putative +2 subsites most likely through random collisions, from where it progresses to the −1 and +1 subsites. Following binding into a productive mode (step 1), the disaccharide (empty blue and filled grey squares) is poised for hydrolysis. After the non-reducing Glc is cleaved off and aglycon diffuses away, the Glc product (red square) remains non-covalently trapped and oscillates (double arrow) between the −1 subsite and +1 subsite (step 2). Glc (yellow square) is consolidated to the −1 subsite after an incoming substrate binds (step 3), which later advances to the catalytic site. Glc (red square) modifies its binding patterns and exits (large arrow) via an autonomous and transient lateral cavity (cylinder in dashed lines) formed near the catalytic site (step 4); a next hydrolytic cycle can begin.

b) Mechanism of Glc displacement with a polysaccharide to facilitate substrate-product assisted processive catalysis. After non-reducing (blue) and penultimate (filled square) residues bind in a productive mode at the −1 and +1 subsites (step 1), the non-reducing Glc residue (yellow square) is cleaved off, with the remainder of the substrate attached (step 2). Glc (red square) modifies its binding patterns and is released (large arrow) via a lateral cavity (cylinder in dashed lines, step 3). Here, the next hydrolytic cycle continues with the same polysaccharide, where the polysaccharide, short of one Glc (filled squares) advances into the catalytic site after uninterrupted binding. Non-reducing (blue) and hydrolysed (cyan-yellow-red in a, yellow-red in b) Glc is shown.

c) Structural basis of substrate-product assisted processive catalysis. Left: The substrate (orange spheres) slides into the active site (protein surface: domain 1-light grey, domain 2-dark grey). After hydrolysis, the Glc product (yellow spheres) exits from the −1 subsite through the lateral cavity into the bulk solvent. Right: detail of the active site. Circle line with arrows indicates directionality during processive catalysis. The image was generated using the coordinates of HvExoI:G6SG-OMe and HvExoI:G2SG-OMe complexes (Fig. 3c-d; orange spheres), and the Glc displacement snapshots based on PELE calculations (Fig. 6e-h; yellow spheres).
vital role of the barrier in Glc entrapment. The sequence of events involving incoming substrate binding and Glc placement are captured in Supplementary Movie 1. We propose a similar substrate/product progression route (Fig. 8b) takes place with a polysaccharide (empty blue and filled grey squares), often with differently linked β-D-glucosyl residues19,20, through processive catalysis. Here, (step 1) after Glc (blue square) is cleaved off from the polysaccharide, (step 2) the Glc product (yellow square) remains bound at the -1 subsite, while the remainder of the polysaccharide (grey squares) stay attached at the +1 and putative +2 subsites. As the Glc product (red square) exits (arrow) through the lateral cavity (step 3, shown as a barrel in dashed lines), the polysaccharide can advance to the -1 and +1 subsites, such that an uninterrupted hydrolysis of the polymer continues until it is completely hydrolysed. In both instances, after Glc egresses (Fig. 8a, b), the next hydrolytic cycle that is facilitated by the dedicated substrate binding and product displacement routes, can continue. Consequently, this coordinated system with oligo- and polymeric substrates, resembles a loop of a conveyor belt that directionally shuffles substrates and ensures efficient catalysis (Fig. 8c). Although in all computations, we modelled Glc in 4C$_2$ conformation, we do not rule out that the Glc distortion to a higher energy of the S$_2$ conformer, as it occurs in crystal structures, may contribute to displacement of the Glc product or its progression.

To our knowledge, this type of processive catalysis17–49 has not been described in any exo-acting hydrolase with a pocket-shaped active site like that of HvExoI. In this work, we define processive catalysis in broader terms, rather than those based on hydrolytic efficiencies, where oligo- or polymeric substrates, resembles a loop of a conveyor belt that directionally shuffles substrates and ensures efficient catalysis (Fig. 8c). Although in all computations, we modelled Glc in 4C$_2$ conformation, we do not rule out that the Glc distortion to a higher energy of the S$_2$ conformer, as it occurs in crystal structures, may contribute to displacement of the Glc product or its progression.

The discovery of substrate-product assisted processive catalysis in HvExoI prompts investigations of the evolutionary origin of this mechanism. The availability of the presented experimental and computational tools alongside a rich source of information for the GH3 family enzymes that originate from all phyla5,13 (currently around 23,000 entries), will now allow the study of evolution of this catalytic mechanism. Preliminary analyses of 500 sequences related to HvExoI revealed that the residues underlying this mechanism are conserved in land plants, but absent in red and green algae, suggesting that this mechanism is rather ancient and has evolved in land plants about 470 million years ago.

We conclude that through the descriptions of associative and dissociative reactant trajectories that we explored using our interdisciplinary approach, we are now in a better position to understand how reactants advance in active sites, and how to improve enzyme catalytic rates, stability, product inhibition and drug discovery.
Second, methyl 3,4,6-tri-O-acetyl-2-S-(2,3,4,6-tetra-O-acetyl-β-D-glucopyranosyl)-2-thio-β-D-glucopyranoside (methyl 2-thio-β-D-glucopyranoside) (chemical structure in Supplementary Table 1): diethylamine (1 mL) under Argon was added to a solution of 2,3,4,6-tetra-O-acetyl-1-thio-β-D-glucopyranoside (91 mg, 0.22 mmol), to which crude methyl 3,4,6-tri-O-acetyl-2-O-trifuoroureasulfonyl-β-D-nanopyranoside (89 mg, 0.19 mmol) in anhydrous DMF (10 mL) was added. The mixture was stirred overnight at room temperature and then concentrated. A solution of the residue in CH3CN was washed in water, dried over anhydrous Na2SO4 and concentrated. Flash chromatography on silica gel (ethyl acetate/petroleum ether, 1:1) afforded methyl 2-thio-β-D-glucopyranoside (26 mg, 21%). Analyses data agreed with those reported previously66. ESI MS: m/z = 689 [M Na]+; 1H NMR (CDCl3) δ 5.12 (t, 1 H, J = 9.3 Hz), 5.04 (t, 1 H, J = 9.4 Hz), 4.97 (m, 3 H), 4.73 (d, 1 H, J = 10.2 Hz), 4.36 (d, 1 H, J = 8.5 Hz), 4.30 (dd, 1 H, J = 4.6, J = 12.3 Hz), 4.26 (dd, 1 H, J = 4.8, J = 12.5 Hz), 4.15 (m, 2 H), 3.76 (2dd, 2 H), 3.56 (s, 3 H), 3.06 (dd, 1 H, J = 8.7 Hz, J = 10.7 Hz), 2.08–1.99 (7 H, 21 H). 13C NMR (CDCl3) δ 170.5–169.22 (CO), 103.79, 83.79, 76.51, 74.02, 71.31, 71.71, 71.48, 69.51, 62.85, 62.25, 62.15, 57.40, 56.10, 53.27, 51.23.

Steady-state affinity binding with recombinant HvExoI. Interaction between recombinant HvExoI and analytes Glc (molecular mass 180.2 g), G6SG-OMe (372.4 g and PEG (n = 5–10); (380–420 g) were performed using Surface Plasmon Resonance (SPR) at 25 °C using the Biacore X100 V2.0.1 instrument (GE Healthcare, USA) equipped with the plus package. Purified recombinant HvExoI at 99% homogeneity (detected by SDS-PAGE) was covalently attached to a CM5 chip (GE Healthcare) in 10 mM sodium acetate buffer, pH 5.0. The carboxymethyl groups present on the dextran layer of the CM5 chip were activated with 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) and N-ethylmaleimide. The enzyme/analyte complexation reactions were blocked by a 7 min injection of 1 M ethanolamine hydrochloride-NaOH, pH 8.5. A ligand density of ~15 000 RU of recombinant HvExoI was achieved. Due to the fast rate constants of interactions, it was impossible to obtain individual rate constants kₙ and k₋. Hence equilibrium analyses were performed for all protein-analyte interactions. For each interaction serial dilutions of analytes covering the 5–1000fold excess range for the enzyme were used. All responses were double referenced using a blank-immobilised chip at a rate of 30 µl/min for 90 s, after which the formed enzyme/analyte complex was washed in water, dried over anhydrous Na2SO4 and concentrated. Flash chromatography on silica gel (ethyl acetate/petroleum ether, 1:1) afforded methyl 2-thio-β-D-glucopyranoside (26 mg, 0.04 mmol) in MeOH (5 mL) and stirred at ambient temperature for 2 h. The mixture was neutralised with the Amberlite IR 120 H⁺ resin, filtered and concentrated. After lyophilisation, purified methyl 2-thio-β-D-glucopyranoside was isolated with 97% yield (14.5 mg).

Expression and crystallisation of recombinant HvExoI. The optimised HvExoI cDNA (GenBank Accession GU441353) was subcloned into the pPZCAZBHN vector, from which the protein was expressed in Pichia pastoris strain SMD1168H (Invitrogen, Carlsbad, USA). The R189H mutation (gating between 1 and 2) was site-directed mutagenesis and expressed in Pichia®. Recombinant HvExoI enzymes were purified using SP-Sepharose column chromatography and immoblised metal affinity chromatography24,25. Purified recombinant wild-type (WT) and variant HvExoI were crystallised via a hanging-drop vapour diffusion method using 24-mercaptoethanol and 100% saturated native protein seeds. The fully-grown crystals of recombinant HvExoI reached up to 500 x 250 x 375 µm sizes after 5 to 14 days and were used for diffraction25.

Crystal structure determination. Enzyme crystals with lengths of up to 400 µm in the longest dimensions were transferred in 100 mM HEPES-NaOH buffer, pH 7.0, containing 1.2% (v/v) PEG (n = 5–10) and 1.7 M ammonium dihydrogen phosphate (solution A), where 3-deoxy-glucone (3dGlc), 4-deoxy-glucone (4dGlc), n-octyl β-D-glucopyranoside (octyl-O-Glc), n-octyl 1-thio-β-D-glucopyranoside (octyl-S-Glc), methyl 2-thio-β-D-gentioside (GS2G-Ome), methyl 6-thio-β-D-gentiobiose (GS6G-Ome) or Glc were supplied during ligand perfusions. The final crystallisations of ligands were between 5–20 mM and perfused throughout 5–720 h. Alternatively, the crystals were deposited in a solution of the concentration of PEG increased to the 35% (v/v). After various perfusion times at 4 ± 2°C crystals were cryo-protected with 15% (v/v) glycerol in the solution A and mounted on a synchrotron goniometer in a stream of N2 gas at 100 K (Oxford Instruments, UK). X-ray diffraction data (native, and with ligands Glc and 4dGlc) were collected on the undulator beamline BioCARS 14-ID-B at the National Institute of Chemical Physics and Biophysics, Tallinn, Estonia, using a MARCCD-165 detector. X-ray diffraction data from crystals WT and variants were collected at the MX1 and MX2 beamslines of the Australian Synchrotron (Australia) at 100 K (Oxford Instruments) or at an UltraTM cryocamera (291 K) with a 400 K cold stream and a 0.1-mm-long collimated mirror at 65% (v/v) CH3CN and separated by high performance liquid chromatography (HPLC) coupled with the Evaporative Light Scattering Detector (ELSD model 800, Alltech Associates Inc.). The samples were measured using the Diamond XFlash Detector (ELSD model 800, Alltech Associates Inc.). The samples were measured using the Diamond XFlash Detector (ELSD model 800, Alltech Associates Inc.). The samples were measured using the Diamond XFlash Detector (ELSD model 800, Alltech Associates Inc.). The samples were measured using the Diamond XFlash Detector (ELSD model 800, Alltech Associates Inc.). The samples were measured using the Diamond XFlash Detector (ELSD model 800, Alltech Associates Inc.). The samples were measured using the Diamond XFlash Detector (ELSD mode...
Carbohydrate ES column, 5 µm, 250 × 4.6 mm (Alltech Associates Inc.). The column temperature during separations was 21 °C. The column eluent was split in 6.5 (detector) to 1.5 % (reference detector) ratio. The detector drift tube was at 40 °C. The nitrogen inlet pressure was 1.5 bars and the signal GAIN setting was set at 4. The HPLC-eluted fractions (6–7.3 min) were reduced with 1 M Na2BD in 2 M NH2OH for 18 h at 4°C and acetylated with acetic anhydride. Hydroxy-inositol (20 ng) was added as an internal standard. Acetylated alditoils were analysed on a low pressure 25 m × 0.25 mm (0.25 mm I.D.) Chrompack Capillary Column CP-S5 CB (Varian Inc., CA, USA) using He as a gas flow with a Hewlett-Packard 6890 Series GC System and a Hewlett-Packard 5937 mass selective detector (Agilent Technologies)72.

NMR spectroscopy of Glc bound to HvExo. NMR spectra were acquired at 283 K on the Bruker AVANCE III 600 MHz and 800 MHz spectrometers (Bruker, MA, USA). For saturation Transfer Difference (STD) experiments, a standard solution of recombinant HvExo deglycosylated by endoglycosidase H72 in milliQ water (143 mM, 20 mM sodium acetate buffer, pH 5.25, 159 mM NaCl) was diluted to 40 mM in the same buffer in D2O, and 60 equivalents (2.4 mM) of Glc from a stock solution in D2O were added. For transferred Nuclear Overhauser Effect Spectroscopy (trNOESY) exhaustive exchange was used. Prior to Free Induction Decay acquisition, an excitation sculpting double echo pulse with a T1 filter (XY16 with 500 ms mixing times to minimise dephasing from JH coupling evolution) protein signal suppression was inserted. STD spectra for free Glc were acquired under the same conditions to show that no direct ligand saturation occurred. trNOESY spectra were acquired with 2048 scans, using 30 ms Gaussian shaped pulses for selective protein saturation at 0.65 ppm; the reference spectrum employed off-resonant saturation at 100 ppm. Different saturation times (0.5, 1, 2, 3 and 5 s) were used. Prior to Free Induction Decay acquisition, an excitation sculpting double echo pulse with a T1 filter, and a T2 filter (XY16 with 500 ms mixing times) with 500 ms mixing times to minimise cross relaxation were employed. WATERGATE scheme using 3–9 19 binomial pulses. 1H signal assignments of ligands were derived from two-dimensional Total Correlation Spectroscopy (TOSCY), NOESY and 13C-Heteronuclear Single Quantum Coherence (HSQC) spectra. The NOESY spectrum of free thiocellobiose was acquired at 313 K with 700 s mixing time to obtain positive NOEs. A NOESY spectrum acquired under the conditions used for the trNOESY spectrum (800 MHz, 283 K, 300 ms mixing time) yielded no observable NOEs, indicating NOE zero-crossing for thiocellobiose (for τc = 1.12/ω0).

**Classical MD simulations.** The structure of native HvExo in complex with Glc (defined as β-D-glucopyranosyl) (PDB 3WHL) with occupancy 0.5 for –1 and +1 subsites, suggests that the two subsites are alternatively occupied. We set up two coenzymes complexes, with Glc bound in the –1 subsite, whereas β-D-glucoside bound in the +1 subsite was occupied by water molecules, and vice versa. The protonation state of the titratable residues was selected based on their carbon dioxide binding environment. Namely, all Arg, Lys, Asp and Glu residues were in ionised form, except for the acid-base residue Glu491, which was protonated (i.e. non-ionised) to represent the catalytic acid/base form required in the first step of the reaction. The protonation state was determined by simulations at pH 7.0 were assigned with PROPKA3.18 through PDB2PQR version 2.0.089. In this complex, Glu491 was modelled in a non-ionised form; this is the required protonation state for the acid/base catalyst in substrate/product complexes of retaining GH3 enzymes that catalyse hydrolysis by the generally accepted double-displacement reaction mechanism.2 The entire system was solvated in a water box of 103.0 × 88.0 × 101.0 Å3 volume, which ensured that there was a 5 Å thick layer of water molecules in each direction of the box. The system was ground-truthed to neutralise the charge of the system forming a fully solvated model of 83,428 atoms. The system was energy-minimised to remove steric clashes and heated through 4 ns equilibration (integration steps of 2 fs) until reaching the temperature of 300 K at 1 Bar. This process also allowed the avoidance of any constraints on the protein and ligand atoms that were gradually released until the whole system was fully equilibrated. Equilibration of the system was stopped when root-mean-square deviation (RMSD) values of the main chain reached the stable value lower than 1.60 Å. Two independent simulations of 100 ns each were run under NPT conditions, whereas temperature and pressure were controlled through Langevin dynamics and the Nosé-Hoover algorithm combined with the Langevin piston method.29,80,80,80. Periodic boundary conditions were applied, and the SHAKE algorithm was used to adjust O-H separations of water molecules. The cut-off value of 12 Å was used for non-bonded interactions. All calculations were carried out with NAMD2.97 using the CHARMM2 force field.60 for the protein and TIP5P for water molecules, whereas the CHARMM2.2 all-atom carbohydrate force field90 was used for Glc.

**Docking of Glc and disaccharides in HvExo.** Molecular docking calculations of Glc and β-D-glucopyranosyl-(1,2)-β-D-glucosyl (G2O2G) or β-D-glucopyranosyl-(1,3)-β-D-glucosyl (G3O3G) to HvExo for selected protein structures were performed with Gold3,25 applying a search of the 20 Å radius and using the Goldscore scoring function. Side-chain flexibility for Glu491 was allowed, whereas Trp434 and Trp286 were flexible as specified below. Complex 1 was generated by docking Glc to the HvExo-G2O2G complex derived from the HvExo-G2O2G-OMe crystal structure (PDB 6MID6). Complex 2 was generated by docking G2O2G to the HvExo-Glc complex derived from the native HvExo crystal structure (PDB 3WLH73) using complex energy minimised parameters. Conformations and protein-sugar interactions were consistently maintained during classical MD.
with the major conformation of Trp334 (CA-CB.CG-CD1 dihedral angle of around −30°). MD simulations were carried out under periodic boundary conditions of the 1x1x1 LxLxL cubic ensembles, as described for MD simulations of HvExoI in complex with Glc. Docking of ZGO2, G3OG and β-1,6-glucopyranosyl-(1,6)-d-glucose (G6OG) lacking the Glc product in the active site was carried out as described above.

**GaudiMM calculations.** Ternary complexes 1–3 before MD simulation (as described above) and complex 3 after MD simulation with substrates bound in a productive mode, were used as starting points to investigate Glc displacement pathways from the −1 subsite of HvExoI. The protein contained Glc at the −1 subsite, and ZGO2 or G3OG attached at +1 and putative +2 substrates. GaudiMM20 (https://github.com/insilicemg/audi), a recently developed modular multi-objective genetic algorithm platform allows conformational exploration of defined genes (defined below) with multiple evaluation operators of fitness or objectives. The GPathFinder extension was used, defining as genes protein and substrates molecules, rotational bonds of ligands and rotamers of the protein, and as objectives, minimisation of van der Waals contacts and maximisation of separations between Glc geometric centres at the −1 subsite. This afforded the low-cost computational pre-identification of putative exit channels. The ProDy (Protein Dynamics)92 Normal Mode Algorithm (NMA) was used to generate alternative starting structures for the GPathFinder. Normal mode calculations were performed with Tangram implementation on the UCSF Chimera suite93, using residue grouped structures. The PDB accessions of putative Glc in a productive mode, generating 10 modes with a cut-off of 15.0 Å and a Gamma Lj of 1.0. In the case of the crystal structure (PDB 3WHL), the second frame of the mode with frequency 31.2 cm−1 was chosen, while the second frame of the mode with frequency 28.7 cm−1 was selected for complex 3. In both cases, the displacement parameter in the Tangram interface was set to 150.

**PELE calculations.** The PELE software (https://pele.bsc.es)94 was used to simulate Glc displacement and to identify the Glc migration route from the −1 subsite. PELE is a Monte Carlo based algorithm that generates configurations along a path through a sequential ligand and protein geometric perturbation scheme, side-chain conformational prediction and minimisation steps. The ligand (Glc) is perturbed by random translations and rotations, and by internal rotations of protein rotatable bonds, while C(Glc) is perturbed by random translations and rotations, and by internal rotations around

**References**

1. Warshel, A. et al. Electrostatic basis for enzyme catalysis. *Chem. Rev.* **10**, 3213–3335 (2006).
2. Adamczyk, A. J., Cao, J., Kamerlin, S. C. & Warshel, A. Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proc. Natl Acad. Sci. USA* **108**, 14115–14120 (2011).
3. Fried, S. D. & Boxer, S. G. Thermodynamic framework for identifying free energy inventories of enzyme catalytic cycles. *Proc. Natl Acad. Sci. USA* **110**, 12271–12276 (2013).
4. Marc, A. & Engasser, J. M. Influence of substrate and product diffusion on heterogenous kinetics of enzyme reversible reactions. *J. Theor. Biol.* **94**, 179–189 (1982).
5. Elber, R. Long-timescale simulation methods. *Curr. Opin. Struct. Biol.* **15**, 151–156 (2005).
6. Varghese, J. N., Hrmova, M. & Fincher, G. B. Three-dimensional structure of a barley β-1,3-glucan exohydrolase, a family 3 glycosyl hydrolase. *Structure* **7**, 179–190 (1999).
7. Hrmova, M. et al. Catalytic mechanisms and reaction intermediates along the hydrolytic pathway of plant β-1,3-glucan glucohydrolase. *Structure* **9**, 1005–1016 (2001).
8. Hrmova, M. & Fincher, G. B. barley β-1,3-glucan exohydrolases substrate specificity and kinetic properties. *Carbohydr. Res.* **305**, 209–221 (1998).
9. Hrmova, M. & Fincher, G. B. Dissecting the catalytic mechanism of a plant beta-1,3-glucan glucohydrolase through structural biology using inhibitors and substrate analogues. *Carbohydr. Res.* **342**, 1613–1623 (2007).
10. Yoshida, E. et al. Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3 β-glucosidase from *Kluyveromyces marxianus*. *Biochem. J.* **431**, 39–49 (2010).
11. McAndrew, R. P. et al. From soil to structure, a novel dimeric β-glucosidase belonging to glycoside hydrolase family 3 isolated from compost using metagenomic analysis. *J. Biol. Chem.* **288**, 14985–14992 (2013).
12. Back, J.-P., Whitworth, G. E., Stubbs, K. A., Vodaclo, D. J. & Mark, B. L. Active site plasticity within the glycoside hydrolase NagZ underlies a dynamic mechanism of substrate distortion. *Chem. Biol.* **19**, 1471–1482 (2012).
13. Suzuki, K. et al. Crystal structures of glycosides hydrolase family 3 β-glucosidase 1 from Aspergillus aculeatus. *Biochem. J.* **452**, 211–221 (2013).
14. Pac, P. et al. Crystal structure of native α-1,6-rhamnosidase from *Aspergillus terreus*. *Acta Crystallogr.* **D74**, 1078–1084 (2018).
15. Hrmova, M. et al. Three-dimensional structure of the barley β-1,3-glucan glucohydrolase complex in a transition state mimic. *J. Biol. Chem.* **279**, 4970–4980 (2004).
16. Hrmova, M. et al. Structural rationale for low nanomolar binding of transition state mimics to a family GH β-1,3-glucan glucohydrolase from barley. *Biochemistry* **44**, 16529–16539 (2005).
17. Hrmova, M. et al. Structural basis for a broad specificity in higher plant β-1,3-glucan glucohydrolase complexPlant Cell **14**, 1–22 (2002).
18. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Database* **2013**, 1–9 (2013).
19. Adamczyk, A. J., Cao, J., Kamerlin, S. C. & Warshel, A. Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proc. Natl Acad. Sci. USA* **108**, 14115–14120 (2011).
20. Fried, S. D. & Boxer, S. G. Thermodynamic framework for identifying free energy inventories of enzyme catalytic cycles. *Proc. Natl Acad. Sci. USA* **110**, 12271–12276 (2013).
21. Borrelli, K. W., Vitalis, A., Alcantara, R. & Guallar, V. PELE: protein energy landscape exploration. *Curr. Opin. Struct. Biol.* **22**, 267–276 (2013).
22. Hrmova, M. et al. Structural rationale for low nanomolar binding of transition state mimics to a family GH β-1,3-glucan glucohydrolase from barley. *Biochemistry* **44**, 16529–16539 (2005).
23. Hrmova, M. et al. Structural basis for a broad specificity in higher plant β-1,3-glucan glucohydrolase complexPlant Cell **14**, 1–22 (2002).
24. Luang, S., Hrmova, M. & Ketudat Cairns, J. R. High-level expression of barley β-1,3-glucan exohydrolase HvExoI from a codon-optimized CDNA in *Pichia pastoris*. *Prot. Exp. Purif* **73**, 90–98 (2010).
25. Luang, S., Ketudat Cairns, J. R., Streitlov, V. A. & Hrmova, M. Crystalisation of wild-type and variant forms of a recombinant β-1,3-glucan glucohydrolase from barley (Hordesium vulgare L.) by macroseeding with wild-type native microcrystals and preliminary X-ray analysis. *Int. J. Mol. Sci.* **11**, 2759–2769 (2010).
26. Cramer, D. & Pople, J. A. A general definition of ring puckering coordinates. *J. Am. Chem. Soc.* **97**, 1355–1358 (1975).
27. Quicocho, F. A. Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions. *Ann. Rev. Biochem.* **55**, 287–315 (1986).
insights into the catalytic machinery of family I glycosidases. J. Mol. Biol. 371, 1204–1218 (2007).
56. Horton, D. 1-Thio-β-o-glucose. Methods Carbohydr. Chem. 2, 433–437 (1963).
57. Classon, B., Liu, Z. & Samuelsson, B. New halogenation reagent systems useful for the mild one-step conversion of alcohols into iodoles or bromides. J. Org. Chem. 53, 6126–6130 (1988).
58. Dong, H., Pei, Z., Angelin, M., Byström, S. & Ramström, Ö. Efficient synthesis of β-o-mannosides by double parallel or double serial inversion. J. Org. Chem. 72, 3694–3701 (2007).
59. Horton, D., Wolfrom, M. L. & Thiosugars, I. Synthesis of derivatives of 2-amino-2-deoxy-1-thio-o-glucose. J. Org. Chem. 79, 1794–1800 (1962).
60. Johnston, B. D. & Pinto, B. M. Synthesis of thio-linked disaccharides by 1 — 2 intramolecular thioglycolyl migration: oxocarbonium versus episolium ion intermediates. J. Org. Chem. 66, 4607–4617 (2001).
61. Otwinowski, Z. & Minor, W. A processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 267, 307–316 (1996).
62. Marshallov, G. N. P. et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D67, 355–367 (2011).
63. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure determination. Acta Crystallogr. D66, 213–221 (2010).
64. McRee, D. E. XtalView/Xfit - a versatile program for manipulating atomic coordinates and electron density. J. Struct. Biol. 125, 156–166 (1996).
65. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of coot. Acta Crystallogr. D66, 486–501 (2010).
66. Winn, M. D., Isupov, M., Pannu, N. S. & Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular structures. Acta Crystallogr. D57, 122–133 (2001).
67. Painter, J. & Merritt, E. A. A molecular viewer for the analysis of TLS rigid-body motion in macromolecules. Acta Crystallogr. D61, 465–471 (2005).
68. Brünger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nat. Struct. Biol. 35, 472–475 (1992).
69. Brünger, A. T. X-PLOR Version 3.85: a system for crystallography and NMR. (Yale University Press, New Haven, CT, 1992).
70. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharam, V. Stereochemistry of polypeptide chain configurations. J. Mol. Biol. 7, 95–99 (1963).
71. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK—a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26, 283–291 (1993).
72. Lau, E. & Bäcå, A. Capillary gas chromatography of partially methylated alditol acetates on a high polarity, cross-linked, fused silica BPhX70. Column. J. Chrom. 637, 100–103 (1993).
73. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. J. Comput. Chem. 25, 1157–1174 (2004).
74. Kirschner, K. N. et al. GLYCAM06: a generalizable biomolecule force field. Carbohydrates. J. Comput. Chem. 29, 622–655 (2008).
75. 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).
76. Kale, L. et al. NAMD2: greater scalability for parallel molecular dynamics. Comput. Phys. 151, 283–312 (1999).
77. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. Comput. Phys. 6, 1781–1802 (2005).
78. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of molecular dynamics equations. J. Comput. Chem. 4, 1–10 (1983).
79. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. Chem. Phys. 81, 511–519 (1984).
80. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. Phys. Rev. A. 31, 1695–1697 (1985).
81. 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).
82. Car, R. & Parrinello, M. Unified approach for molecular dynamics and density-functional theory. Phys. Rev. Lett. 55, 2471–2474 (1985).
83. Troullier, N. & Martins, J. L. Efficient pseudopotentials for plane-wave calculations. Phys. Rev. B. 43, 1993–2006 (1991).
84. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. Phys. Rev. Lett. 77, 3865–3868 (1996).
85. Barducci, A., Bonomi, M. & Parrinello, M. Metadynamics. WIREs Comput. Mol. Sci. 1, 826–843 (2011).
86. Laio, A. & Parrinello, M. Escaping free-energy minima. Proc. Natl Acad. Sci. USA. 99, 12562–12566 (2002).
87. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. Phys. Rev. Lett. 100, 200603 (2008).
88. MacKerell, A. D. J. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B. 102, 3586–3616 (1998).
Acknowledgements

We thank M. Raab (Slovak Academy of Sciences), B.J. Smith (La Trobe University) and G.B. Fincher (University of Adelaide) for interest in this research. H. Tong (Advanced Photon Source), N. Matsugaki and S. Wakatsuki (Photon Factory), and F. Pettolino (University of Adelaide) for interest in this research. H. Tong (Advanced Photon Source), N. Matsugaki and S. Wakatsuki (Photon Factory), and F. Pettolino (University of Adelaide) are thanked for advice and S. Pradeau for technical assistance.

We thank M. Raab (Slovak Academy of Sciences), B.J. Smith (La Trobe University) and G.B. Fincher (University of Adelaide) for interest in this research. H. Tong (Advanced Photon Source), N. Matsugaki and S. Wakatsuki (Photon Factory), and F. Pettolino (University of Adelaide) are thanked for advice and S. Pradeau for technical assistance.

We thank M. Raab (Slovak Academy of Sciences), B.J. Smith (La Trobe University) and G.B. Fincher (University of Adelaide) for interest in this research. H. Tong (Advanced Photon Source), N. Matsugaki and S. Wakatsuki (Photon Factory), and F. Pettolino (University of Adelaide) are thanked for advice and S. Pradeau for technical assistance.