Article

An alternative broad-specificity pathway for glycan breakdown in bacteria

https://doi.org/10.1038/s41586-024-07574-y

Received: 22 June 2023

Accepted: 16 May 2024

Published online: 19 June 2024

Check for updates

Seyed Amirhossein Nasseri[1,2], Aleksander C. Lazarski[3,4], Imke L. Lemmer[1,2], Chloe Y. Zhang[2], Eva Brencher[2], Hong-Ming Chen[1], Lyann Sim[2], Deepesh Panwar[2], Leo Betschart[1], Liam J. Worrall[3,4], Harry Brumer[1,2,3], Natalie C. J. Strynadka[3,4] & Stephen G. Withers[1,2,3] ✉

The vast majority of glycosidases characterized to date follow one of the variations of the 'Koshland' mechanisms[1] to hydrolyse glycosidic bonds through substitution reactions. Here we describe a large-scale screen of a human gut microbiome metagenomic library using an assay that selectively identifies non-Koshland glycosidase activities[2]. Using this, we identify a cluster of enzymes with extremely broad substrate specificities and thoroughly characterize these, mechanistically and structurally. These enzymes not only break glycosidic linkages of both α and β stereochemistry and multiple connectivities, but also cleave substrates that are not hydrolysed by standard glycosidases. These include thioglycosides, such as the glucosinolates from plants, and pseudoglycosidic bonds of pharmaceuticals such as acarbose. This is achieved through a distinct mechanism of hydrolysis that involves oxidation/reduction and elimination/hydration steps, each catalysed by enzyme modules that are in many cases interchangeable between organisms and substrate classes. Homologues of these enzymes occur in both Gram-positive and Gram-negative bacteria associated with the gut microbiome and other body parts, as well as other environments, such as soil and sea. Such alternative step-wise mechanisms appear to constitute largely unrecognized but abundant pathways for glycan degradation as part of the metabolism of carbohydrates in bacteria.

Most glycosidic bonds are highly stable under typical biological conditions, with half-lives of around 5 million years for the degradation of cellulose and starch[3]. To hydrolyse such substrates on biologically useful timescales, glycosidases have evolved to be among the most proficient of catalysts[3]. For the vast majority of cases, hydrolysis of the glycosidic bond is achieved by one of two well-characterized Koshland mechanisms[1] involving acid/base catalysed reactions through oxocarbenium ion-like transition states (Fig. 1a). Minor variations exist in the identities of the catalytic residues[4] or the recruitment of intramolecular catalysis[5]. Achievement of such high proficiency in catalysis by these enzymes requires the evolution of an enzyme active site that exquisitely stabilizes the transition state of the reaction, and this is realised through complementary interactions of all types with essentially all parts of the substrate on both sides of the scissile bond[6]. As a consequence, glycosidases tend to be quite specific for their natural substrates.

The question then arises as to whether alternative, less-demanding and therefore less-specific mechanisms for degradation of glycosides may have evolved that would endow organisms with the ability to degrade pools of diverse substrates that would be otherwise inaccessible to them. This could include complex *O*-glycans of unusual linkage type that are not cleaved by specific Koshland glycosidases or glycosides linked by atoms other than oxygen, such as carbon, nitrogen or sulfur.

Intrigued by this idea, we examined its possibility in the bacterial world using a functional metagenomic approach. The large number of unassigned genes, representing up to 25% of those in known bacterial genomes[7], certainly allows room for this possibility. As screening substrates for the identification of new mechanistic classes of glycosidases, we used the fluorogenic alkylthioglycosides (Fig. 1b) that we had previously[2] shown to not be hydrolysed by Koshland glycosidases, yet cleaved efficiently by non-Koshland glycosidases from CAZy[8] families GH4/109[9] and GH88/105[10]. These substrates will therefore detect not only thioglycosidases but also other possible glycoside-cleaving enzymes that do not use Koshland mechanisms.

Here, we applied this strategy to a large-scale screen of a human gut microbiome metagenomic library using a diverse set of thioglycoside substrates. A total of 12 unique hits was identified, with at least one hit for each of the substrates used. Through kinetic, mechanistic and structural analysis of two of these hits and a homologous system in another bacterium, we identified and characterized an operon that carries out the hydrolysis of a highly broad set of substrates. Analogues of these enzymes are present in a wide variety of organisms and the components of this system are, in many cases, interchangeable between organism and substrate type. These enzyme systems represent an underappreciated mode of glycan degradation.

[1]Department of Chemistry, University of British Columbia, Vancouver, British Columbia, Canada. [2]Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia, Canada. [3]Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia, Canada. [4] Center for Blood Research, University of British Columbia, Vancouver, Canada. ✉e-mail: withers@chem.ubc.ca

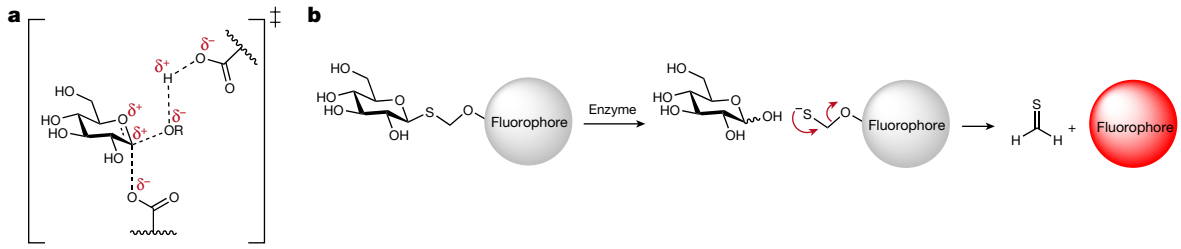


Fig. 1 | Koshland glycosidases and non-Koshland thioglycoside substrates. a, Oxocarbenium-like transition state of a typical retaining Koshland glycosidase. **b**, The structure of the selective substrates used in this study and the mechanism of the self-immolative linker disintegration.

Screen of a human gut metagenome library

We synthesized the seven fluorogenic thioglycoside substrates shown in Fig. 2, based on β-glucose, β-galactose, α-mannose, β-glucuronic acid, 4,5-unsaturated β-glucuronic acid, β-*N*-acetylglucosamine and β-6-phosphoglucose, as described in Supplementary Schemes 1–4. These substrates cover a broad spectrum of commonly occurring glycosides as well as two (**5** and **7**) as controls because enzymes from GH4 (6-phospho-β-glucosidase) and/or GH88/105 (α-4-5-ene-glucuronidase) were likely to be present in our libraries.

Our 20,000-member human gut metagenomic fosmid library[11] was chosen for the screen, as the vast amount of genome sequences available for the human gut microbiome simplifies and enriches the downstream analysis. Anticipating few hits, we pooled the substrates for our first screen, which led to approximately 120 hits (Supplementary Fig. 2); a hit rate comparable to those of metagenomic screens for well-established enzymatic activities[12]. Hits were picked and arrayed into two 96-well 'hit plates', copies of which were then screened separately with the seven synthesized substrates (Fig. 2), as well as the equivalent *O*-glycosides. A summary of all of the results is provided in Supplementary Table 1. At least one hit was found for each thioglycoside substrate.

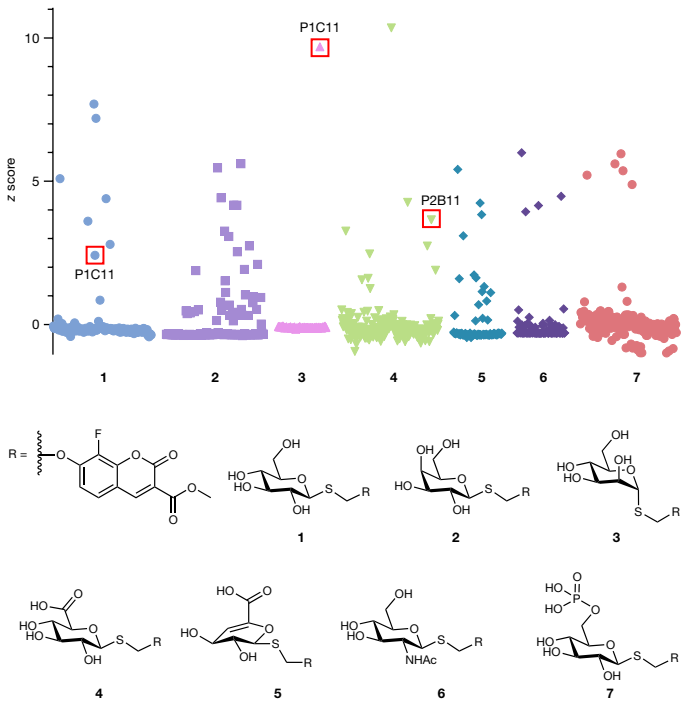


Fig. 2 | Substrate structures and screening results. The structures of substrates **1**–**7** used in this study (the R group represents the fluorophore Jericho Blue), and the deconvoluted *z*-score plots for the second round of screening.

Sequencing of the hit fosmids revealed that many of them were derived from the same regions of bacterial genomes. After removal of redundant hits, about a dozen candidates remained. A summary of these and their most probable source organisms is provided in Supplementary Table 2. The majority of these fosmids contain no known glycosidase genes, yet hydrolytic activity was unequivocal. Here, we focus on two of the hits for which we have fully characterized the sets of proteins responsible for the activity. Detailed investigation of the other hits is ongoing.

Investigations of the P1C11 hit

A particularly interesting hit was P1C11, the single clone that was active against the α-*S*-mannoside substrate **3** and the β-*S*-glucoside substrate **1**, hinting at an unusual mechanism that is able to hydrolyse both α- and β-glycosides (Fig. 2 and Supplementary Table 1). Sequencing of this fosmid revealed its origin from an *Alistipes* species with open reading frames (ORFs) encoding many uncharacterized genes along with several annotated as carbohydrate-related enzymes, but no known glycosidases (Fig. 3a). The closest was a domain of unknown function 1080 (DUF1080)-containing gene, encoding a protein with structural similarity to GH16 endo-1,3-1,4-β glucanases[13] along with nearby genes encoding a putative sugar phosphate isomerase, two Gfo/Idh/MocA-like oxidoreductases and a DUF6377 predicted to be a transcriptional regulator. In silico analysis of these proteins, hereafter called AL1, AL2, AL3, AL4 and AL5 (Fig. 3b), predicts that—in addition to AL5, which is membrane associated[14]—three of the other proteins have signal peptides[15]: AL1 has a general secretory system signal peptide, AL2 has a lipoprotein signal peptide and AL3 has a twin-arginine signal peptide suggesting periplasmic locations, while AL4 is cytosolic.

Similar genes in *Bacteroides* species

Notably, a very similar operon (BT2156-2160; Fig. 3b) in another prevalent gut commensal bacterium, *Bacteroides thetaiotaomicron*, was recently identified as responsible for metabolism of glucosinolates, naturally occurring β-thioglucosides in plants[16]. At least three out of the four enzymes, as well as the SusR-like sensor/regulator[17], were needed for activity in vivo, while preliminary data suggested that only two of the enzymes were required to observe the activity in vitro, although the roles of these proteins were not elucidated. The same cluster was also identified through a functional genetics study on *B. thetaiotaomicron*[18] as being responsible for the degradation of several α-linked disaccharides, including α,α-trehalose, leucrose and maltitol. Homologues of these enzymes also occur in many other human-gut-associated *Bacteroides* species[16]. Further details are provided below.

The similarity of our *Alistipes* sp. enzymes to those in *B. thetaiotaomicron* and the limited mechanistic and structural insights into how this *B. thetaiotaomicron* cluster functions to degrade such diverse glycans led us to perform parallel structural and mechanistic investigations of the two systems. To that end, we cloned each of the enzyme-encoding

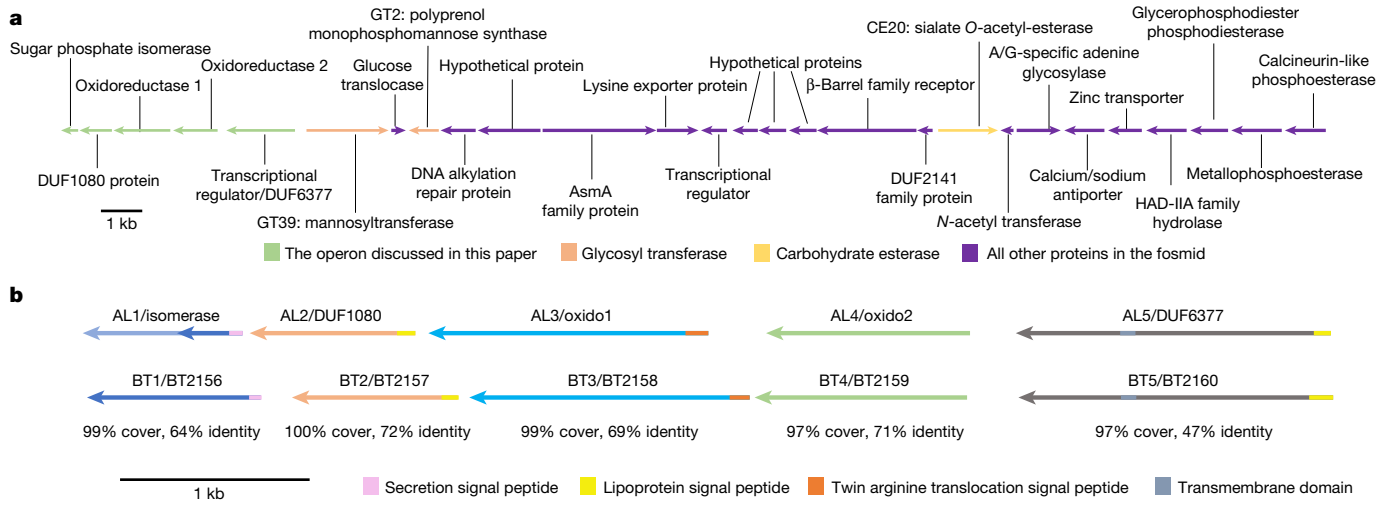


Fig. 3 | Genes in the P1C11 fosmid and *B. thetaiotaomicron*. **a**, ORFs present in the hit, P1C11, including glycosyltransferases from GT2 and GT39 families and a carbohydrate esterase; the five genes discussed in this report are shown in green. **b**, The organization of the genes in the two clusters, showing the percentage identity of the AL enzymes with their BT counterparts. The ORF for AL1, present in the fosmid, was truncated (dark blue); the full length ORF found in multiple *Alistipes* sp. genomes is shown in light blue.

genes and expressed and purified them from *Escherichia coli* as described in the Supplementary Information.

Kinetic studies of AL and BT enzymes

With all of these proteins in hand, we demonstrated that the four BT enzymes, used together, hydrolyse both our fluorogenic β-*S*-glycoside substrate **1** and the α-*S*-mannoside **3** as well as both the α and β anomers of methylumbelliferyl *O*-glucoside and mannoside, with a decided preference for the β-glycosides. The four AL enzymes together likewise hydrolyse both anomers of both the glucoside and mannoside, but with a preference for the α anomers (Supplementary Information Section 2.5.2.3). As implied previously[16], neither AL4 or BT4 are essential for activity in vitro (Supplementary Information Section 2.5.2.2). Importantly, we showed (Supplementary Information Section 2.5.1.2) that the enzymes AL2 and BT2 (AL/BT2) and AL3 and BT3 (AL/BT3) are largely interchangeable between the two systems, suggesting that they have similar roles and also that they do not work as obligate multienzyme complexes.

While the cloned enzymes were active together in vitro, the overall rates under the conditions used were extremely low, as also reported previously[16], making mechanistic studies challenging and, importantly, suggesting a missing component. However, addition of various metal ions and varying the buffer salts and ionic strength led to only minor rate improvements (Supplementary Information Sections 2.5.1.3 and 2.5.1.4). By contrast, addition of EDTA ablates the reaction for both series of enzymes, indicating important previously bound metals.

Despite the fact that the clusters contain a pair of oxidoreductases, addition of NAD(P)(H) had no pronounced effect on the reaction rates, nor did any other tested cofactors (Supplementary Information Section 2.5.1.5). Assuming that NAD(H) is indeed involved, based on the precedent of other Gfo/Idh/MocA oxidoreductases, this would suggest a tightly bound cofactor carrying out an oxidation of the substrate and getting recycled back into its oxidized form by some means other than cofactor exchange, consistent with the periplasmic location. On the basis of our previous studies of the GH4 enzyme family[19], the most likely mechanism involves an initial oxidation by AL/BT3 to generate a 3-keto-sugar intermediate, as also suggested previously[18]. This acidifies the proton at C2, and facilitates glycoside cleavage through an elimination reaction, distinct from the nucleophilic substitution reactions of Koshland glycosidases[1].

In the absence of NAD(H) exchange in the in vitro system, AL/BT3 will get stuck in the reduced form, unable to perform another catalytic cycle. To bypass this restriction, we added methyl β-3-ketoglucoside (synthesized using a site-selective palladium catalyst[20]) to the reaction mixtures of both series of enzymes as a stoichiometric re-oxidant. This indeed increases the hydrolysis rates of our fluorogenic substrates by more than 100-fold (Supplementary Fig. 9), indicating that this was the rate-limiting step and opening up the opportunity of detailed mechanistic investigations.

Roles of individual enzymes

Roles of individual enzymes were next identified by ^{1}H nuclear magnetic resonance (^{1}H-NMR) analyses of their reactions. We first confirmed the roles of AL/BT3 as the initial oxidants by showing that addition of either of AL/BT3 to equimolar mixtures of a glucoside and a 3-keto-glucose derivative (Glc-α-pNP and 3-keto-1,5-anhydroglucitol in Extended Data Fig. 1a) results in oxidation of the glucoside to its 3-ketoglucoside, yielding an equilibrium mixture (54:46) of the reduced:oxidized species. Performing the same reaction with the reverse reduced/oxidized pair of substrates gives the same equilibrium mix (Supplementary Information Section 2.7.6).

Roles for the other components were assigned by monitoring reactions of chemically synthesized 3-ketoglycoside versions of the minimalist substrates, methyl α- and β-glucoside. As seen in Extended Data Figs. 1b and 2a, AL/BT1 carry out the elimination reaction on the β anomer of 3-keto-GlcOMe, while AL/BT2 do so on the α anomer. Hydration of the resultant 3-keto-2-hydroxy-glucal is then performed efficiently by AL/BT2, while AL/BT1 do so much more slowly (Supplementary Information Sections 2.7.2–2.7.5). Hydration leads to a mixture of isomers of 3-keto-glucose, dominant among which is the 3-keto-gluco-furanose[21] shown in Extended Data Fig. 2a. Thus, the enzymes AL/BT1, previously annotated as sugar phosphate isomerases, are in fact 3-keto-β-glucoside-1,2-lyases while AL/BT2, members of DUF1080 family, are both 3-keto-α-glucoside-1,2-lyases and 3-keto-2-hydroxy-glucal hydratases, and not 3-keto-glucoside hydrolases as previously suggested[18]. The dual activity of AL/BT1 and AL/BT2 is not surprising, considering the very similar transition states for the elimination and hydration reactions (Extended Data Fig. 2b). Control experiments (Supplementary Figs. 93 and 94) confirm that these enzymes alone have no effect on simple glucoside substrates

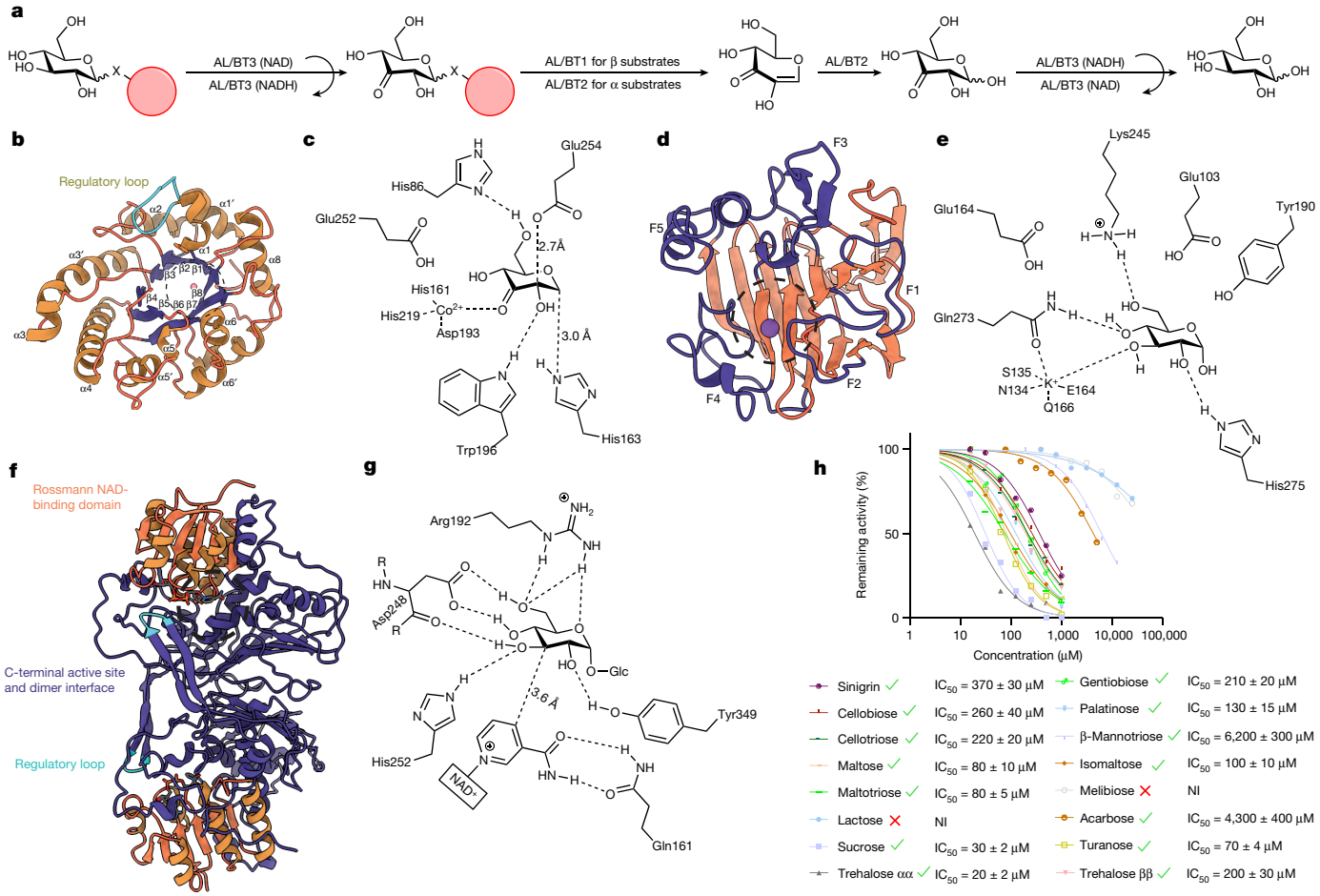


Fig. 4 | Structures of the enzymes and the reactions that they catalyse. **a**, The mechanism of stepwise hydrolysis of α- and β-glycosides by the enzymes AL and BT1–3. **b**, Cartoon representation of AL1. Barrel strands (purple), helices (orange), regulatory loop (cyan) and the active site (dashed circle) are indicated. **c**, Schematic of the active site of BT1 bound to 3-keto-1,5-anhydroglucitol. **d**, Side view of the overall fold of BT2; the β jelly roll (orange), fingers 1–5 (purple; F1–F5) and the active site (dashed circle) are indicated. **e**, Schematic of the active site of BT2 with glucose bound. **f**, The overall fold of AL3. The Rossmann-like NAD fold (orange), C-terminal active site region and dimerization interface (purple), regulatory loop (cyan) and active site (dashed circle) are indicated. **g**, Schematic of the active site of AL3 bound to α,α-trehalose. **h**, Inhibition of AL enzymes by a select number of disaccharides: those that were hydrolysed are shown with green ticks and those that were not hydrolysed are marked with red crosses. NI, no inhibition (compounds that cause no significant inhibition up to the highest concentration tested, 50 mM).

such as the *para*-nitrophenyl glucosides. Moreover, by manipulating enzyme concentrations or by stepwise addition of the enzymes, the intermediates of each step in the overall reaction for hydrolysis of both anomers of *para*-nitrophenyl glucosides are detected by NMR (Supplementary Figs. 91, 92 and 95–98). On the basis of these findings, the overall pathway for hydrolysis of glycosides by the enzymes AL/BT1–3 is depicted in Fig. 4a.

X-ray crystallography

To gain further insights into the mechanisms of these enzymes, high-resolution models of AL1, AL2, AL3, BT1 and BT2 were determined using X-ray crystallography (details are provided in Supplementary Information Section 2.8 and Extended Data Table 2). The resulting models of AL1 and BT1 displayed a classical overall α/β (TIM) barrel fold, typical of the isomerases that they are related to, with an eight-stranded core barrel surrounded by eight helices on the outer layer (Fig. 4b). The active site contains a conserved catalytic Glu (272 and 275 in AL1 and BT1, respectively) for proton extraction (Fig. 4c and Extended Data Fig. 3a,b). In agreement with the results of kinetic assays (Supplementary Information Section 2.5.4.2), a Cobalt metal was observed and confirmed by X-ray fluorescence (XRF) scans of AL1 and BT1 crystals (Supplementary Fig. 103a,b). This is coordinated by His178/182, Asp211/214 and His237/240 as well as the 3-keto group of the substrates in the substrate-bound models (Fig. 4c and Extended Data Fig. 3c). In addition to this ion, the substrate binds to His103/107, Trp214/217 and His180/184, while the predicted catalytic Glu272/275 side-chain OE2 is positioned 2.7 Å above carbon 2 of the bound substrate analogue, 3-keto-1,5-anhydroglucitol (Fig. 4c and Extended Data Fig. 3a).

Apo BT2 and AL2 formed monomers adopting a β jelly roll fold that creates a catalytic channel flanked by five finger domains comparable to proteins from the GH16 family[13] (Fig. 4d and Extended Data Fig. 3d). BT2 and AL2 displayed highly similar folds with an root mean squared deviation (r.m.s.d.) of 0.5 Å across 229 Cα pairings (Supplementary Fig. 110). Notable divergences from GH16 structures were observed within the catalytic channel where no homologous E-[ILV]-D-[IVAF]-[VILMF] (0,1)-E motif[22] was observed, indicating that although the overall fold is similar, BT2 contains different catalytic residues. A potassium ion was observed in the channel (Supplementary Fig. 103i), which probably replaced the functional metal (proposed to be Ca^{2+} based on the kinetic data; Supplementary Information Section 2.5.4.2) during crystallization due to its high concentration (Fig. 4e, Extended Data Fig. 3e and Supplementary Table 5). To identify possible catalytic and metal coordinating

residues, BT2 crystals were soaked with α,α-trehalose before diffraction. Soaked crystals contained a glucose-like molecule in the active site coordinated at the 2, 3, 4 and 6 OH groups by His275, the metal ion, Gln273 and Lys245, respectively (Fig. 4d, Extended Data Fig. 3f and Supplementary Fig. 107). Consistent with this, sequence alignments confirmed conserved amino acids in the channel where Glu103 (>95% conserved as Glu/Asp), Asn134, Glu164, Gln166, Tyr190, Lys245, Gln273 and His275 were all highly conserved across 3,183 homologous sequences and probably comprise the metal-binding and active site of BT2 (Fig. 4d, Extended Data Fig. 3e,f and Supplementary Fig. 104b).

AL3 structures crystallized as dimers with an N-terminal Rossmann-like fold, and a C-terminal domain containing the active site, dimerization interface and a regulatory loop that blocks the opposing monomer active site in the dimer pair (Fig. 4f). Both apo and trehalose-bound enzymes crystallized with NAD co-factors (Supplementary Fig. 108c); however, despite XRF scans indicating the presence of Co^{2+} (Supplementary Fig. 103e), no Co^{2+} was observed. Trehalose-bound AL3 identified the primary glycoside-binding pocket as comprising Gln161 (which mediates the adjacent NAD cofactor), as well as Arg192 (coordinating O-5 and OH-6 of the glucose ring), Trp195, Asp248, His252 (OH-3) and Tyr349 (OH-2) (Fig. 4g and Extended Data Fig. 3g). Numerous additional distinct trehalose binding sites were observed beyond that of the active site just described, with some arranged in a chain-like configuration. It is tempting to suggest this continuous path of sugar binding may show how AL3 and BT3 bind to larger oligosaccharides for catalysis (Extended Data Fig. 3h; see below for further discussion), but given the high concentration of trehalose present in the soaking experiment, non-specific binding may be responsible.

Broad substrate specificities

The finding that enzyme systems of this class are able to hydrolyse simple non-activated substrates such as methyl glucosides of both α and β configurations is unprecedented. We were therefore interested in exploring whether this broad specificity extends to commonly occurring oligosaccharides, as this might illuminate the reasons for the parallel evolution of these enzymes in the presence of highly efficient Koshland glycosidases.

As an initial screen, using a mixture of all of the enzymes, we assayed inhibition of the hydrolysis of our fluorogenic substrates by a series of oligosaccharides, as this should detect both competitive substrates and competitive inhibitors (Supplementary Information Section 2.5.5). As shown in Fig. 4h, with the exception of galactosides, a wide range of α- and β-glycosides of varying linkage were inhibitory. Further analysis using thin-layer chromatography and ^{1}H-NMR (Supplementary Information Sections 2.6.2, 2.7.9 and 2.7.10) reveals that all of the inhibitory oligosaccharides are indeed completely hydrolysed. This includes both β- and α-glucosides with essentially any glycosidic linkage to the next sugar unit. The next sugar unit can be any monosaccharide (examples include sucrose (Glcα(1→2)βFruc) and maltose (Glcα(1→4)Glc) or indeed larger saccharide units as cello- and malto-oligosaccharides are all efficiently hydrolysed. This broad activity also extends to glucosides with aglycones such as phenols (salicin, fluorogenic MU-glycosides) as well as thio-glycosides such as sinigrin. The enzymes appear to be strict exo-glycosidases degrading from the non-reducing end and therefore have little effect on polysaccharides in vitro. However, in vivo, they would work in concert with suitable endo-glycosidases to degrade such substrates.

As the mechanism of action of these enzymes does not require an endocyclic oxygen, we hypothesized that they could also cleave pseudo-glycosides. Notably, we found that the enzyme systems are able to break the C-N bond of the pseudo-oligosaccharide acarbose (Supplementary Figs. 70, 72 and 130), a type II diabetes therapeutic, hinting at a possible degradation pathway for this drug molecule and presumably for other compounds of this class in human gut environments that contain these enzymes. This provides another example of how the human gut microbiome can metabolize drugs, this time discovered using an indirect mechanistic approach. Indeed, essentially any cyclohexyl ring with the right configuration of hydroxy groups for oxidation of the 3-hydroxy should undergo a subsequent elimination reaction to break the (pseudo)glycosidic bonds, provided that the substrate fits into the enzyme active sites.

Extension of the inhibition assay on AL/BT3 to monosaccharides (Supplementary Information Section 2.5.5) reveals that the only essential hydroxy groups in the case of AL/BT3 are those at the 3 and 4 positions. Modifications at other positions are allowed provided that their size does not preclude active site binding. Thus, for example, quinovose and 6-phospho glucose are inhibitory but not melibiose (Galα(1→6)Glc).

A consequence of this mechanism is that glycosides modified at the 3-position are not hydrolysed, because oxidation does not occur. Notably, many of the aminoglycoside antibiotics have C-3 modifications and none of those that we tested are substrates of these enzymes (Supplementary Figs. 71 and 72 and Supplementary Information Section 2.6.1.3.3). 3-Amino-modified monosaccharides can also be found among many other natural products[23,24], while 3-*O*-methylation, deoxygenation and branching occurs in some other cases[25,26]. It therefore seems quite likely that modifications at the 3-position have evolved as a means of combatting the degradation of glycosylated natural products such as antibiotics through a 3-keto-mediated pathway.

In vivo hydrolysis of α- and β-glycosides

Previous studies have implicated BT1–5 in the hydrolysis of α-glycosides such as α,α-trehalose[18] and β-thio-glycosides such as glucosinolates[16] by *B. thetaiotaomicron*. To examine this more fully in light of the observed broad specificities, we first investigated whether several *Bacteroides* species could hydrolyse the strictly non-Koshland β- and α-thioglycoside substrates **1** and **3**. Grown under the same conditions in rich medium, *B. thetaiotaomicron* and the other strains containing homologues of this operon (*Bacteroides ovatus*, *Bacteroides uniformis*, *Bacteroides xylanisolvens*) can indeed hydrolyse both substrates, while species lacking similar genes (*Bacteroides fluxus* and *Bacteroides finegoldii*) as well as the mutant *B. thetaiotaomicron* strain lacking the oxidoreductase (ΔBT3) show no activity (Extended Data Fig. 4a,b).

However, the observed activities for *B. thetaiotaomicron* cells grown in rich medium are low, prompting us to investigate what conditions induce the expression of these proteins. Analogous to a previous report[18], we observed strong upregulation of BT1–4 transcripts in cells grown in minimal medium with α,α-trehalose as the sole carbon source (Extended Data Fig. 4c). As expected, these cells also hydrolyse **1** and **3** orders-of-magnitude faster than the cells grown in minimal medium with glucose or rich medium (Extended Data Fig. 4d and Supplementary Fig. 112), establishing that α,α-trehalose can act as an inducer. Importantly, although *B. thetaiotaomicron* cells are reported to not grow on substrates such as cellobiose (Glcβ(1→4)Glc) under regular laboratory conditions[27], we found that the addition of small amounts of α,α-trehalose as an inducer results in the consumption of cellobiose and robust growth (Extended Data Fig. 4e). Similar results were obtained for the wild-type strain with a suite of other glycosides, including the natural products salicin and sinigrin as well as methyl manno- and glucosides (Supplementary Figs. 116–120). On the other hand, mutant strains lacking the oxidoreductase BT3 or the SusR-like regulator BT5[17] do not grow on trehalose or other similar substrates, directly implicating these loci in glycoside utilization (Supplementary Figs. 121–128). Finally, we observed that, in addition to trehalose, other α-glucosides such as palatinose and turanose as well as β-glucosides such as cellobiose and salicin can also act as inducers and eventually lead to growth, but with lag phases of variable length (Extended Data Fig. 4f and Supplementary Information Section 2.10.3).

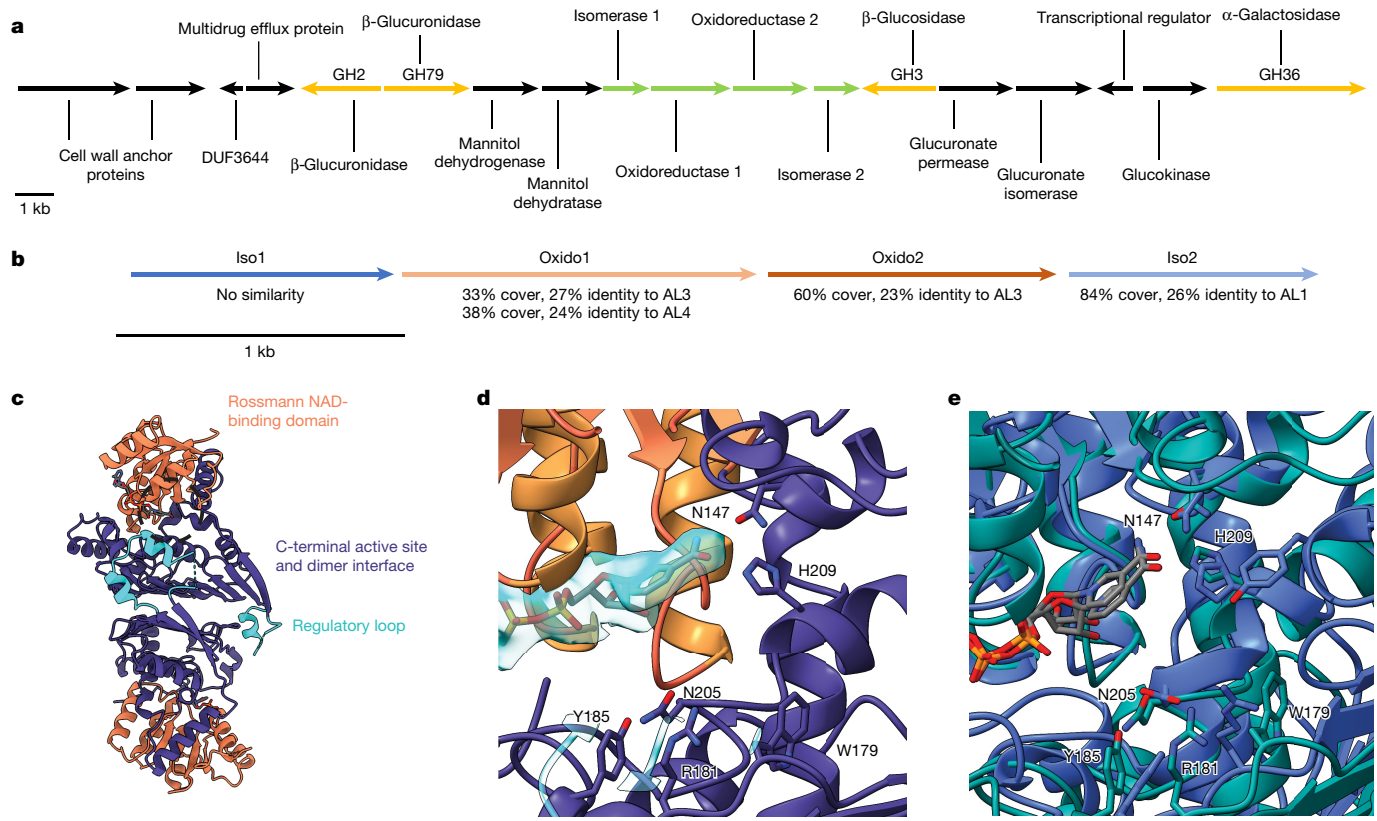


Fig. 5 | Genes in the P2B11 fosmid and structural information. a, ORFs present in the hit P2B11. **b**, The organization of the four genes from P2B11, with the percentage identity of the proteins to their counterparts from P1C11. **c**, The overall fold of P2B11–Oxi1 dimer pair. The Rossmann-like fold (orange), C-terminal active site and dimerization interface (purple) and regulatory loop (cyan), and the active site (dashed circle) are indicated. **d**, P2B11–Oxi1 apo active site with core residues labelled and the 2mFo-dFc omit map around NAD shown as a surface representation at an r.m.s.d of 1.5. **e**, Alignment of the apo P2B11–Oxi1 (teal) active site with apo AL3 (blue); r.m.s.d. = 1.2 Å over 91 Cα pairs. Core residues are labelled.

Although further experiments are required to fully elucidate the regulation of this system and the roles of BT4 and BT5—the two other proteins essential for activity in vivo[16,18]—these observations are in full accordance with the designation of BT1–5 as a broad-spectrum glycoside-utilization operon.

Investigations of the P2B11 hit

Another of our fosmid hits—P2B11, from an unknown Gram-positive bacterium—is active on thio-glucuronide substrate **4**. This fosmid encodes a cluster of four enzymes in which, in contrast to the above cases, the two oxidoreductase genes are flanked by two genes that are annotated as sugar-phosphate isomerases with no apparent nearby regulator (Fig. 5a). While several other genes annotated as glycosidases are found in this fosmid, the similarity of this set of four genes to those previously studied (Fig. 5b) suggested that they were responsible for the activity. Biochemical characterization confirmed this, revealing substantial similarity between these proteins and those discussed above, despite modest sequence similarity. Here also, Oxido1 proved to be the most important for activity, while Iso1 and Iso2 both enhance the rate decidedly and Oxido2 has minor effects (Supplementary Information Section 2.5.6.5). Catalytic activity is enhanced in the presence of Mn^{2+} and Co^{2+} ions, while the presence of NAD(P)(H) cofactors again has no effect on reaction rate. These four proteins together also have a relatively broad substrate specificity, hydrolysing both β- and α-glucuronide substrates at similar rates (Supplementary Information Section 2.5.6.6). While glucosides are not hydrolysed, glycosides bearing negative charges at the 6-position, like 6-PO_4, 6-SO_4 and sulfoquinovoside substrates, are all cleaved, albeit slowly compared with glucuronides. Some of the components of this cluster can be swapped with those found in P1C11, despite different substrate specificities. Thus, AL3 can efficiently oxidize glucuronide substrates, and a mixture of AL3 and Iso1 and Iso2 from P2B11 hydrolyses the glucuronide substrate (Supplementary Information Section 2.5.6.9), confirming the mechanistic parallels.

P2B11–Oxido1 crystallized as a dimer (six dimers in the asymmetric unit of the P21 cell) with the core N-terminal Rossmann domain, C-terminal active site region, dimerization interface and inhibitory loop for opposing monomers displaying an overall fold comparable to that of AL3 (Fig. 5c). XRF scans of the crystals confirmed the presence of Co^{2+} (Supplementary Fig. 103f), suggesting metal binding and specificity; however, as with AL3, no ordered anomalous electron density was evident in the active-site maps. Similar predominantly hydrophobic interfaces were observed between each inhibitory loop and its regulatory partner in both enzymes; however, the relative orientation of the fold and disposition of the active sites on opposite faces in the P2B11–Oxido1 dimer differed compared with the same faces in AL3 (Supplementary Fig. 111b). Glycoside-binding pockets appear to contain the same core in Asn147 for NAD coordination, while Trp179, Arg181, Tyr185, Asn205 and His209 comprise the active-site pocket (Fig. 5d). Most residues appeared in similar positions, except for Tyr185 and Trp179, which lie on opposite sides of Arg181 relative to the analogous residues in AL3 (Fig. 5e).

Discussion and conclusions

The operons described in this study endow the bacteria that harbour them with the ability to cleave not only otherwise recalcitrant

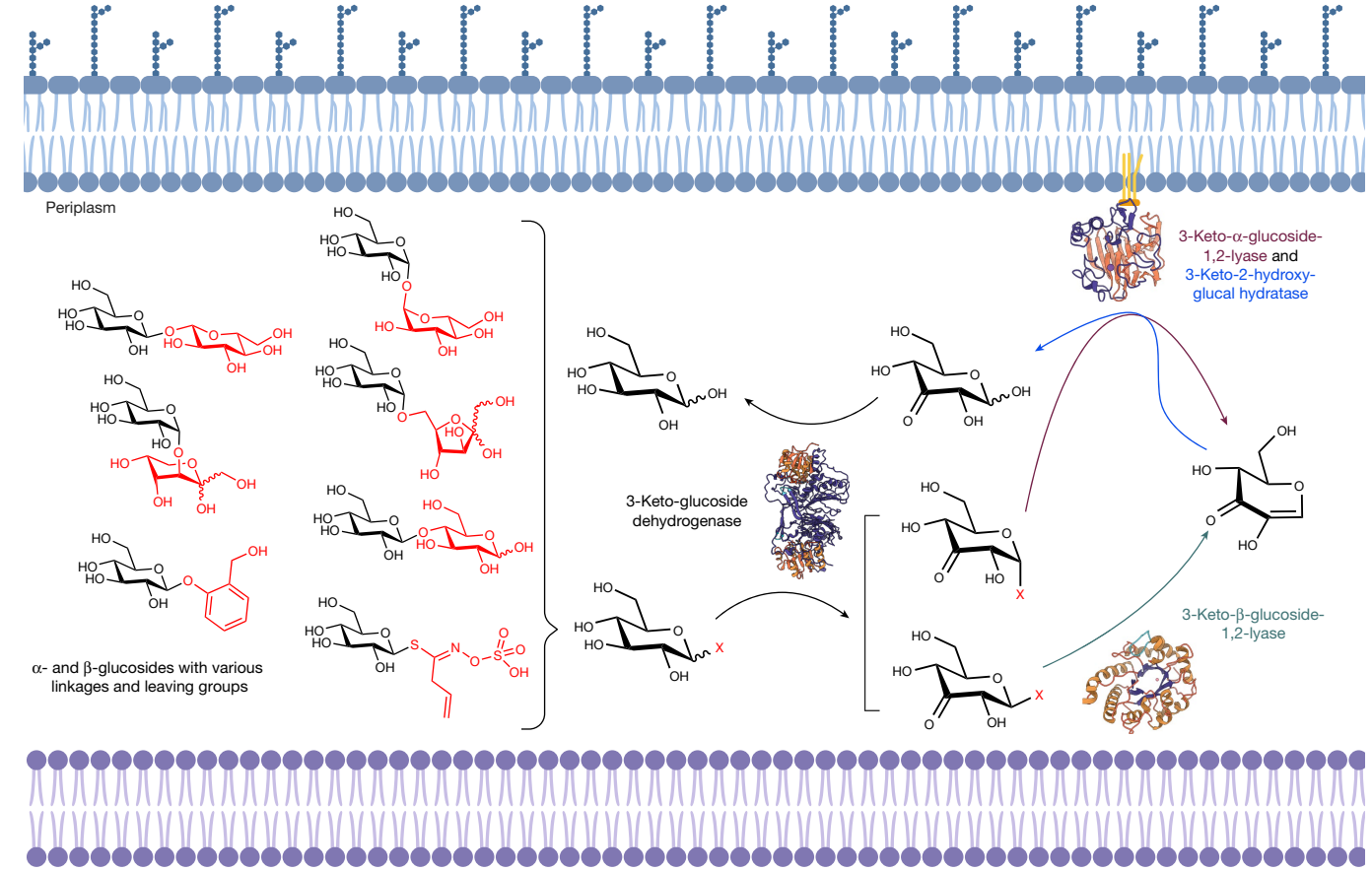


Fig. 6 | The overall proposed mechanism of broad-specificity stepwise hydrolysis. Schematic of the overall proposed mechanism through which the enzymes AL/BT1 (3-keto-β-glucoside-1,2-lyase), AL/BT2 (3-keto-α-glucoside-1,2-lyase and 3-keto-2-hydroxy-glucal hydratase) and AL/BT3 (3-keto-glucoside dehydrogenase) work in concert to break the glycosidic bonds of a diverse set of glycosides. This diagram was created using BioRender.

thioglycosides, but also both the α and β glycosidic bonds of a wide range of commonly occurring oligosaccharides. This versatility derives from their different chemical mechanism, involving anionic transition states rather than the cationic oxocarbenium ion-like transition states of the more common[8] Koshland glycosidases, as well as the separation of the reaction into several steps. We established the functions of the individual enzyme components by NMR monitoring of their reactions, detected each intermediate in the overall reaction, and identified the key residues involved in binding and catalysis by solving the structures of the enzymes.

Figure 6 summarizes the proposed mechanism through which the enzymes AL/BT1–3 can hydrolyse a diverse range of substrates within the periplasm to generate glucose while returning the oxidase AL/BT3 to its starting oxidation state. Thus, the tightly bound NAD cofactor of AL/BT3 is most likely recycled in vivo through reduction of the 3-ketoglucose generated by AL/BT2. Addition of the 3-ketoglucose derivative was necessary for initial rate kinetics in vitro as isomerization of 3-ketoglucose to its furanose form reduces the concentration of the active form. Reoxidation of the enzyme in vivo might also occur through specialized proteins, such as cytochromes[28,29], or through other small-molecule hydride acceptors. To assess this possibility, we screened a range of common carbonyl-containing chemicals and found that D-erythrose and dehydroascorbic acid can also function as co-substrates (Supplementary Information Section 2.5.7). However, the reaction rates are significantly lower than with 3-ketosugars, and the relevance of these specific molecules in vivo when 3-keto-glucose itself will be generated by AL/BT2 remains to be seen.

The substantial differences between this mechanism and that of the Koshland glycosidases have several consequences. As the hydroxyl at C3, rather than the anomeric centre, is the linchpin of this mechanism, both α- and β-anomers can be accommodated. Indeed, comparable pseudo-Michaelis–Menten kinetic parameters are found (Extended Data Table 1) for the reactions catalysed by individual or collective enzymes with both anomers of simple aryl glycosides. Furthermore, by breaking down the reaction coordinate into a series of lower-energy steps, less evolutionary optimization is required to arrive at useful reaction rates for diverse and challenging substrates. Moreover, the interchangeability of these subunits should make the evolution of 'new' activities more facile. However, this versatility does come at the cost of expressing multiple proteins to perform a reaction that Koshland glycosidases carry out efficiently in one step. Our experiments show that the 3-keto pathway is a last-resort pathway in *B. thetaiotaomicron* and, as suggested previously[16], these enzymes are expressed conditionally, only when the bacteria encounter carbohydrates for which no specialized glycoside hydrolases are encoded.

As shown previously[16], close homologues of the BT enzymes are present in about half of the assessed human gut microbiome samples collected worldwide. Moreover, our identification of a divergent glucuronide-hydrolysing homologue, P2B11, shows that bacterial breakdown of glycosides through a 3-keto intermediate is not limited to glucosides and probably extends well beyond. Using our hit sequences as templates, we conducted a search of online genomic databases for other bacteria encoding similar operons. As shown in Extended Data Figs. 5 and 6 and Supplementary Tables 6 and 7, a wide range of both

Gram-positive and Gram-negative bacteria from diverse environments (human and animal bodies, soil and marine environments, hot springs and opportunistic pathogens) carry similar operons with a variety of architectures and often only distant sequence similarities. Specific substrates for each operon await identification.

3-Keto glycosides were first discovered in the 1960s, when bacterial strains that oxidize disaccharides into 3-keto sugars were reported[30]. The best studied example was the plant pathogen *Agrobacterium tumefaciens*, which converts sucrose, among other disaccharides, to 3-keto-sucrose[31,32]. However, the identities of the enzymes responsible were not known and no chemical rationale was provided for the oxidation: the enzymes responsible for breaking the glycosidic bonds were assumed to be regular glycosidases. More recently, ketosugars have been reported as intermediates in the breakdown of anhydrosugars such as levoglucosan[33,34], as well as C-glycoside-cleaving systems that operate through similar oxidation and elimination reactions[35–39]. As mentioned, other glycosidases that transiently oxidize sugar hydroxy groups as part of their mechanism have been reported by us and others[9,40–42]. However, no similarly broad specificity was observed in any of these cases. While levoglucosan is not a substrate of our enzymes, AL/BT3 have the ability to oxidize C-glycosides, but the lyase enzymes AL/BT1 do not further degrade the 3-keto-C-glycoside (Supplementary Information Section 2.6.1.3.2), presumably because they lack the required residues for aromatic enolization[43].

Our findings consolidate and considerably extend these disparate reports, laying out substantial structural and mechanistic support for a distinct, widespread and highly flexible alternative pathway that is used by bacteria for degrading a wide range of glycans. Glycosidases are arguably one of the most extensively studied classes of enzymes. Elucidation of a new biochemical pathway for metabolism of one of the best-studied class of biomolecules in the well-trodden ground of the human gut microbiome serves as a reminder that there remains an immense amount to be learned about the chemistry of biological systems.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-07574-y.

1. Koshland, D. E. Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev.* **28**, 416–436 (1953).
2. Nasseri, S. A., Betschart, L., Opaleva, D., Rahfeld, P. & Withers, S. G. A mechanism-based approach to screening metagenomic libraries for discovery of unconventional glycosidases. *Angew. Chem.* **130**, 11529–11534 (2018).
3. Wolfenden, R., Lu, X. & Young, G. Spontaneous hydrolysis of glycosides. *J. Am. Chem. Soc.* **120**, 6814–6815 (1998).
4. Watts, A. G. et al. Trypanosoma cruzi trans-sialidase operates through a covalent sialyl-enzyme intermediate: tyrosine is the catalytic nucleophile. *J. Am. Chem. Soc.* **125**, 7532–7533 (2003).
5. Vocadlo, D. J. & Withers, S. G. Detailed comparative analysis of the catalytic mechanisms of β-*N*-acetylglucosaminidases from families 3 and 20 of glycoside hydrolases. *Biochemistry* **44**, 12809–12818 (2005).
6. Namchuk, M. N. & Withers, S. G. Mechanism of *Agrobacterium* β-glucosidase: kinetic analysis of the role of noncovalent enzyme/substrate interactions. *Biochemistry* **34**, 16194–16202 (1995).
7. Mistry, J. et al. Pfam: the protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).
8. Drula, E. et al. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.* **50**, D571–D577 (2022).
9. Yip, V. L. Y. & Withers, S. G. Family 4 glycosidases carry out efficient hydrolysis of thioglycosides by an α,β-elimination mechanism. *Angew. Chem. Int. Ed.* **45**, 6179–6182 (2006).
10. Jongkees, S. A. K. & Withers, S. G. Glycoside cleavage by a new mechanism in unsaturated glucuronyl hydrolases. *J. Am. Chem. Soc.* **133**, 19334–19337 (2011).
11. Rahfeld, P. et al. An enzymatic pathway in the human gut microbiome that converts A to universal O type blood. *Nat. Microbiol.* **4**, 1475–1485 (2019).
12. Lorenz, P. & Eck, J. Metagenomics and industrial applications. *Nat. Rev. Microbiol.* **3**, 510–516 (2005).
13. Viborg, A. H. et al. A subfamily roadmap of the evolutionarily diverse glycoside hydrolase family 16 (GH16). *J. Biol. Chem.* **294**, 15973–15986 (2019).
14. Hallgren, J. et al. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. Preprint at *bioRxiv* https://doi.org/10.1101/2022.04.08.487609 (2022).
15. Teufel, F. et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* **40**, 1023–1025 (2022).
16. Liou, C. S. et al. A metabolic pathway for activation of dietary glucosinolates by a human gut symbiont. *Cell* **180**, 717–728 (2020).
17. Ravcheev, D. A., Godzik, A., Osterman, A. L. & Rodionov, D. A. Polysaccharides utilization in human gut bacterium *Bacteroides thetaiotaomicron*: comparative genomics reconstruction of metabolic and regulatory networks. *BMC Genomics* **14**, 873 (2013).
18. Liu, H. et al. Functional genetics of human gut commensal *Bacteroides thetaiotaomicron* reveals metabolic requirements for growth across environments. *Cell Rep.* **34**, 108789 (2021).
19. Yip, V. L. Y. et al. An unusual mechanism of glycoside hydrolysis involving redox and elimination steps by a family 4 β-glycosidase from *Thermotoga maritima*. *J. Am. Chem. Soc.* **126**, 8354–8355 (2004).
20. Jäger, M., Hartmann, M., de Vries, J. G. & Minnaard, A. J. Catalytic regioselective oxidation of glycosides. *Angew. Chemie Int. Ed.* **52**, 7809–7812 (2013).
21. Morris, P. E., Hope, K. D. & Kiely, D. E. The isomeric composition of D-ribo-hexos-3-ulose (3-keto-D-glucose) in aqueous solution. *J. Carbohydr. Chem.* **8**, 515–530 (1989).
22. Allouch, J. et al. The three-dimensional structures of two β-agarases. *J. Biol. Chem.* **278**, 47171–47180 (2003).
23. Lemke, A., Kiderlen, A. F. & Kayser, O. Amphotericin B. *Appl. Microbiol. Biotechnol.* **68**, 151–162 (2005).
24. Arsic, B. et al. 16-membered macrolide antibiotics: a review. *Int. J. Antimicrob. Agents* **51**, 283–298 (2018).
25. Zhao, Q. et al. Neriifolin from seeds of *Cerbera manghas* L. induces cell cycle arrest and apoptosis in human hepatocellular carcinoma HepG2 cells. *Fitoterapia* **82**, 735–741 (2011).
26. Sugawara, K. et al. Elsamicins A and B, new antitumor antibiotics related to chartreusin. 2. Structures of elsamicins A and B. *J. Org. Chem.* **52**, 996–1001 (1987).
27. Martens, E. C. et al. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* **9**, e1001221 (2011).
28. van Beeumen, J. & de Ley, J. Hexopyranoside: cytochrome c oxidoreductase from *Agrobacterium tumefaciens*. *Eur. J. Biochem.* **6**, 331–343 (1968).
29. Takeuchi, M., Asano, N., Kameda, Y. & Matsui, K. Physiological role of glucoside 3-dehydrogenase and cytochrome c_{551} in the sugar oxidizing system of *Flavobacterium saccharophilum*. *J. Biochem.* **103**, 938–943 (1988).
30. Bernaerts, M. J. & De Ley, J. Microbiological formation and preparation of 3-ketoglycosides from disaccharides. *J. Gen. Microbiol.* **22**, 129–136 (1960).
31. Grebner, E. E. & Feingold, D. S. D-Aldohexopyranoside dehydrogenase of *Agrobacterium tumefaciens*. *Biochem. Biophys. Res. Commun.* **19**, 37–42 (1965).
32. Hayano, K. & Fukui, S. Purification and properties of 3-ketosucrose-forming enzyme from the cells of *Agrobacterium tumefaciens*. *J. Biol. Chem.* **242**, 3655–3672 (1967).
33. Kuritani, Y. et al. Conversion of levoglucosan into glucose by the coordination of four enzymes through oxidation, elimination, hydration, and reduction. *Sci. Rep.* **10**, 20066 (2020).
34. Kaur, A. et al. Identification of levoglucosan degradation pathways in bacteria and sequence similarity network analysis. *Arch. Microbiol.* **205**, 155 (2023).
35. Braune, A., Engst, W. & Blaut, M. Identification and functional expression of genes encoding flavonoid *O*- and *C*-glycosidases in intestinal bacteria. *Environ. Microbiol.* **18**, 2117–2129 (2016).
36. He, P. et al. Structural mechanism of a dual-functional enzyme DgpA/B/C as both a *C*-glycoside cleaving enzyme and an *O*- to *C*-glycoside isomerase. *Acta Pharm. Sin. B* **13**, 246–255 (2023).
37. Mori, T. et al. C-Glycoside metabolism in the gut and in nature: Identification, characterization, structural analyses and distribution of C-C bond-cleaving enzymes. *Nat. Commun.* **12**, 6294 (2021).
38. Zheng, S. et al. A newly isolated human intestinal bacterium strain capable of deglycosylating flavone C-glycosides and its functional properties. *Microb. Cell Fact.* **18**, 94 (2019).
39. Taborda, A. et al. Mechanistic insights into glycoside 3-oxidases involved in C-glycoside metabolism in soil microorganisms. *Nat. Commun.* **14**, 7289 (2023).
40. Liu, Q. P. et al. Bacterial glycosidases for the production of universal red blood cells. *Nat. Biotechnol.* **25**, 454–464 (2007).
41. Bell, A. et al. Uncovering a novel molecular mechanism for scavenging sialic acids in bacteria. *J. Biol. Chem.* **295**, 13724–13736 (2020).
42. Kaur, A. et al. Widespread family of NAD-dependent sulfoquinovosidases at the gateway to sulfoquinovose catabolism. *J. Am. Chem. Soc.* **145**, 28216–28223 (2023).
43. Bitter, J. et al. Enzymatic β-elimination in natural product *O*- and *C*-glycoside deglycosylation. *Nat. Commun.* **14**, 7123 (2023).

Methods

Chemical synthesis

A detailed description of all of the chemicals synthesized for this study is provided in Supplementary Information Section 1.

Metagenomic screen

The screening procedure is based on that reported previously[44] with minor changes. The preparation of the human gut microbiome metagenomic library used in this study has been described previously[11]. In brief, this library has been made from bacterial DNA isolated from a faecal sample of a healthy, male donor and is expressed in *E. coli* strain EPI300*ΔmelA*[44]. Appropriate numbers of 384-well plates (Corning) were filled with LB medium (50 µl per well) containing 100 µg ml^{-1} L-arabinose and 12.5 µg ml^{-1} chloramphenicol. These plates were next inoculated with bacteria from the metagenomic library storage plates using a Qpix2 xt robot (Genetix) and then incubated overnight at 37 °C in closed containers to minimize evaporation. Next, 45 µl of screening buffer (100 mM phosphate buffer pH 7 with 2% Triton X-100 containing 60 µM of each of the substrates **1**, **3**, **5**, **6** and **7**) was added to each well and the plates were again incubated at 37 °C in closed containers. This process was repeated for the mixture of compounds **2** and **4** separately.

The fluorescence of each well was measured after initial addition of the screening buffer and after 2 and 5 h and overnight incubation using a Synergy H1 plate reader (BioTek) with an excitation wavelength of 405 nm and an emission wavelength of 454 nm (ref. 45). Hits were selected based on *z*-score values, as described in Supplementary Information Section 2.1. These were then picked and arranged into two 96-well 'hit plates', which were subsequently grown on a larger scale and screened individually with the selective substrates **1**–**7** (Fig. 2) as well as multiple common *O*-glycoside substrates. The results of these screens are summarized in Supplementary Table 1.

Sequencing results and analysis

Aliquots from the selected wells, containing *E. coli* clones carrying the fosmids, were plated onto LB/agar plates with chloramphenicol (35 mg ml^{-1}) and incubated at 37 °C overnight. Single colonies were picked for growth in 10 ml LB medium containing chloramphenicol (12.5 µg ml^{-1}) and L-arabinose (100 µg ml^{-1}) at 37 °C with shaking overnight. The fosmid was then isolated using GeneJET plasmid Miniprep Kit (Thermo Fisher Scientific) according to the manufacturer's protocol with the only adjustment being that the samples were incubated at 65 °C for 3 min with the elution buffer before elution.

Miniprepped fosmids were then further purified for sequencing. Thus, 2 µl 25 mM ATP solution, 6 µl 10× Plasmid-Safe Reaction Buffer and 2 µl Plasmid-Safe ATP-dependent DNase were added to the fosmid sample and incubated at 37 °C for 1 h. Next, the sample was purified using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) according to the manufacturer's protocol. The concentrations of the samples were measured using the Nanodrop photometer and the DNA was stored at −20 °C before being sent for sequencing (Plasmidsaurus).

The sequencing results were then analysed using the tools available online from the National Center for Biotechnology Information website. The most likely sources of the fosmid hits were identified using standard nucleotide Basic Local Alignment Search Tool (BLAST) searches. The results for selected verified hits after removal of the duplicates are shown in Supplementary Table 2. The ORFs present on each fosmid were detected using the Open Reading Frame Finder tool (https://www.ncbi.nlm.nih.gov/orffinder/) and each ORF was then analysed using protein BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The annotation for each gene was assigned based on the most prevalent similar genes found using BLAST searches. The nucleotide sequences of the hits described in this paper are included in Supplementary Information Section 5 and a summary of the results is reported in Supplementary Table 2.

Cloning

All of the ORFs were cloned into pET28 with a C-terminal His-TAG (or N-terminal His-TAG for the clones specified as pET28(a)). Signal peptides were predicted using SignalP[15] and, when present, the DNA bases for their expression were omitted from the cloned construct. Cloning was performed using the polymerase incomplete primer extension cloning method[46] (the only exception being P1C11-DUF1080, which was cloned using the Golden Gate[47] method). In brief, separate PCR reactions were set up for the vector (v) and the insert (i) using 0.2 µM of each primer (F and R for forward and reverse primers, respectively), 5 ng of template (pET28 for the vector and the purified fosmid or genomic DNA for the insert), 5 µl of GC or HF buffer (Thermo Fisher Scientific), 120 µM dNTP mix (30 µM each, final concentration), 0.5 µl Phusion DNA polymerase (Thermo Fisher Scientific) and double-distilled H_2O to a final volume of 25 µl. The program was set to initial denaturation at 98 °C for 2 min, then 25 cycles of denaturation at 98 °C (30 s each), annealing was done at a temperature 5 °C lower than the calculated T_m of the primers for 30 s and elongation was at 72 °C for 45 s per kb. PCR products were controlled by 1% agarose gel electrophoresis and directly transformed into chemically competent DH5α *E. coli* cells using 2 µl of each PCR product and 50 µl of chemically competent cells. All primers were ordered from Thermo Fisher Scientific and are listed in Supplementary Table 3.

Protein purification

All proteins were expressed in *E. coli* BL21(DE3) cells in either LB medium (induced with 0.1 mM IPTG) or in LBE-5052 Auto-induction medium[48]. The specific growth conditions for each protein are provided in Supplementary Table 4. Cells were collected at 4,000 rpm for 30 min and stored at −20 °C until further purification. Culture pellets were resuspended in lysis buffer (20 ml for 500 ml of culture, lysis buffer: 50 mM Tris pH 7.0, 150 mM NaCl, 1% glycerol, 20 mM imidazole, 10 mM $MgSO_4$, 0.2 µl benzonase, 0.5 mg hen egg white lysozyme, 1 tablet of Pierce EDTA-free protease inhibitor per 50 ml of buffer). Cell disruption was done by sonication at 35% intensity, 5/10 pulse for 3 min. Insoluble fractions were pelleted by centrifugation at 16,000 rpm, 4 °C, 30 min. Supernatant was loaded onto Ni-NTA columns, equilibrated in buffer A (50 mM Tris pH 7.0, 150 mM NaCl, 1% glycerol). The column was washed with an imidazole gradient from 20 mM to 400 mM over 50 ml of eluent. The fractions were analysed using SDS–PAGE and fractions containing the protein of interest were concentrated and buffer exchanged into storage buffer, 50 mM Tris pH 7.0, 1% glycerol, using an ultracentrifugation device (Amicon). The proteins were further purified using size-exclusion chromatography (Sepharose, Superdex 200) using the ÄKTA go liquid chromatography system (Cytiva) eluting with Tris buffer (50 mM Tris pH 7.0). Fractions showing UV-absorbance were analysed using SDS–PAGE, concentrated and directly used for crystallization screens. A picture of an SDS–PAGE gel of the proteins is shown in Supplementary Fig. 1.

Enzymatic kinetics

Kinetic studies on enzymatic reactions were monitored using the Synergy H1 Hybrid plate reader (BioTek). The reactions were all performed at room temperature and were done in either black or clear half-area flat-bottom 96-well plates (Corning). Unless otherwise stated, the rates are reported as the concentration of released aglycone per specified unit of time, calculated on the basis of the calibration curve made for the fluorophore/chromophore in the assay buffer and under the same conditions as in the assay. All of the assays described herein were based on initial rates that are calculated for the first 1–10 min of the reactions, depending on the time necessary to observe the response, as well as the linear response range of the fluorophore/chromophore signal. All of the data analysis was performed using GraphPad (Prism). Wherever error bars are reported, they represent standard error of the reported

value measured from three (n = 3) distinct samples, while the reported values are the mean of these measurements. The specific details for each one of the kinetic assays performed for this study are provided in Supplementary Information Section 2.5.

LC–MS assays

All of the experiments were performed using the Agilent 1260 Infinity HPLC system equipped with a UV detector and an Agilent 6120 Quadrupole LC/MS detector. The buffer for these experiments was the same buffer used in kinetic experiments, 25 mM NaPi buffer pH 7.0 with the appropriate metal ions. In all of the experiments, the reactions were set up using 1 mM of the specified substrates and 0.2 μM of enzymes unless otherwise specified. After addition of the enzyme(s), reaction mixtures were incubated at 37 °C for the specified amount of time. Then, 10 μl of the reaction solutions was diluted into 90 μl of acetonitrile to precipitate the enzyme. The solutions were then centrifuged for 3 min, and the top soluble portion was used to prepare liquid chromatography–mass spectrometry (LC–MS) samples. For each LC–MS run, 5 μl of the sample was injected into the column (Agilent Eclipse XBD C-18), which was washed with water (solvent A) and acetonitrile (Solvent B) using the following steps: 0 to 1 min, 95% A; 1 to 11 min gradual change to 5% A; 11 to 12 min, 5% A; 12 to 15 min, gradual change to 95% A. The identity of the peaks is verified by assessing the MS spectra and, when possible, with the use of standard samples assayed under the same conditions. Detailed description of all of these experiments can be found in Supplementary Information Section 2.6.1.

Thin-layer chromatography assays

All of the reactions were performed in the same buffer that was used in kinetic experiments, 25 mM NaPi buffer pH 7.0, with the appropriate metal ions added for each respective system of enzymes. In all of the experiments, the reactions were set up using 5 mM of substrates (2 mg ml^{-1} for polysaccharides), 4 mM of 3-keto-1,5-anhydroglucitol and 5 μM of each enzyme. After addition of the enzyme, solutions were incubated at 37 °C overnight (16 h). Then, 0.5 μl of the reaction solutions was then spotted onto the thin-layer chromatography plates, which were developed using *t*-butanol:ammonium hydroxide:methanol:water with the volumetric ratios of 5:4:4:1 as eluent. Controls involved mixtures of the substrates in the assay buffer and the standard for the expected product(s) but no enzyme. The plates were then stained using Hanessian's stain (ammonium molybdate 4.8% (w/v), ceric ammonium molybdate 0.2% w/v, sulfuric acid 6% (v/v) in water). The details for all experiments are provided in Supplementary Information Section 2.6.2.

NMR studies

The NMR experiments were performed using a 600 MHz NMR spectrometer equipped with TXI 1.7 mm cryoprobe (Bruker), with solvent suppression using noesygppr1d pulse sequence with a relaxation delay (D1) of 2 s, mixing time (D8) of 0.1 s, acquisition time of 2 s and with the number of scans ranging from 16 to 64 and using the software TopSpin v.4.2.0. In a typical experiment, all of the reagents except for the enzyme(s) were mixed in the buffer with appropriate concentrations (5 mM of each substrate unless otherwise noted) and transferred to the NMR tube and this was used to optimize the NMR acquisition conditions (lock, tuning and matching, shimming). The sample was then removed, mixed with the enzyme(s) (2.5 μM unless otherwise stated) and transferred back into the machine to acquire the spectra. Typically, the first data were collected about 2 min after the enzymes were introduced to the sample and the additional data were acquired every 2 to 10 min for the rest of the experiment time and also measured after overnight incubation. All NMR assays were performed at room temperature. The buffer for these experiments was the same as that used in kinetic experiments, 25 mM NaPi buffer pH 7.0, which was lyophilized and reconstituted in D_2O, with the appropriate metals (except Mn(II)[49]) added for each respective sample before the experiments. The details for each experiment are provided in Supplementary Information Section 2.8.

X-Ray crystallography

AL, BT, and P2B11 enzymes were screened for crystals using PACT and JCSG+ (Qiagen) crystallization screens using sitting-drop vapour diffusion in INTELLI-PLATE 96-well plates (Art Robbins Instruments) with 0.2 μl of protein mixed with 0.2 μl of mother liquor. Hits were optimized by varying the conditions pH, salt or precipitant. Hits and final conditions are described in Supplementary Table 5. For all crystals, either glucose or trehalose at 25% (w/v), glycerol 30% (w/v) or increased PEG concentration of 30% (w/v) was used as cryoprotectant before flash freezing with cryoprotectants for each of the crystals described in Supplementary Table 5.

Screening and collection of diffraction data (Extended Data Table 2) was performed at 100 K on beamlines CMCF-BM at the Canadian Light Source using MxDC for data collection, or 23-ID-D at the Advanced Photon Source using JBluIce for data collection. Diffraction data were processed using xia2[50] and XDS[51], with data reduction performed using Aimless[52] as part of the CCP4 package[53]. Anisotropic dataset (trehalose bound AL3) was corrected using the STARANISO server[54]. Phasing was carried out using molecular replacement with Alphafold[55] models for all enzymes except BT2, with AlphaFold models pLDDT scores converted to *b*-factors using process predicted models in the CCP4 suite, and chain A as the search model in Phaser[56], also part of the CCP4 package. BT2 phasing was completed using PDB 3OSD. Sequential rounds of model building and refinement were performed using Coot[57] and Refmac[58]. Validation of the final models was performed using MolProbity[59].

All structure analysis and figure preparation were performed using ChimeraX[60] and Coot, distributed as part of the CCP4 package. Sugar orientations and density fit were checked and analysed using Privateer[61], and dimeric interfaces assessed using PISA[62]. Determination of possible catalytic residues was completed by analysing highly conserved residues in the active site of each enzyme using the ConSurf server[63] (Supplementary Fig. 104). Metal coordination was checked with the check my metal server[64]. Comparisons of homologous AL, BT and CT enzymes were performed using the ChimeraX Matchmaker function. Electron density maps for figures were generated by removing sugars or ligands from the models followed by a subsequent round of TLS refinement resetting *b*-factors to 30 and restrained refinement to make omit maps of substrates and co-factors.

Bioinformatic analysis

The nucleotide sequences of the region encoding the studied proteins in our hits were used to search against the nucleotide collection (nr/nt) available online from the National Center for Biotechnology Information website. The search was done using standard nucleotide BLAST searches, optimized for finding dissimilar sequences (discontiguous megablast). This algorithm was used to find a representative and diverse set of bacteria that encode proteins similar to those we have identified from our metagenomic screen.

Genomes of a select number of the identified bacteria were then inspected to confirm the presence of the corresponding genes for several of the proteins in the vicinity of each other. The percentage similarities of the encoded proteins to each one of the proteins from P1C11 (*Alistipes* sp.) and P2B11 were next calculated using the protein BLAST algorithm. Supplementary Tables 6 and 7 summarize the results for selected verified bacteria that encode proteins similar to those from P1C11 and P2B11, respectively. The results are also shown in Extended Data Figs. 5 and 6. Note that the results represent just a select few strains picked to showcase the diversity of the bacteria that encode similar proteins from various sources. These are by no means comprehensive lists.

Bacteroides cultures, hydrolytic activities and transcription studies

Bacteroides strains were grown in an anaerobic growth chamber at 37 °C using standard microbiology techniques (Supplementary Information Section 2.10) and growth was monitored with a multiwell plate reader at 600 nm. Hydrolytic activities were assessed by incubating aliquots of the bacterial cultures with fluorogenic substrates, monitored with a multiwell fluorescence plate reader (BioTek). Transcription was assessed using standard quantitative PCR with reverse transcription methods. A detailed description of all of the experiments involving *Bacteroides* cultures is provided in Supplementary Information Section 2.10.

Reporting summary

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

Data availability

Data supporting the findings of this study are available within the Article and the Supplementary Information. All sequencing data are available at the end of the Supplementary Information and also from GenBank database with accession numbers PP693324 and PP693325. Structural models are deposited at the Protein Data Bank under accession codes 8TCD, 8TCR, 8TCS, 8TCT, 8TDA, 8TDE, 8TDF, 8TDH, 8TDI and 8V31. Source data are provided with this paper.

44. Armstrong, Z., Rahfeld, P. & Withers, S. G. Discovery of new glycosidases from metagenomic libraries. *Methods Enzymol.* **597**, 3–23 (2017).
45. Chen, H.-M. et al. Synthesis and evaluation of sensitive coumarin-based fluorogenic substrates for discovery of α-*N*-acetyl galactosaminidases through droplet-based screening. *Org. Biomol. Chem.* **19**, 789–793 (2021).
46. Klock, H. E., Koesema, E. J., Knuth, M. W. & Lesley, S. A. Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. *Proteins Struct. Funct. Genet.* **71**, 982–994 (2008).
47. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* **3**, e3647 (2008).
48. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
49. Barnhart, J. L. & Berk, R. N. Influence of paramagnetic ions and pH on proton NMR relaxation of biologic fluids. *Invest. Radiol.* **21**, 132–136 (1986).
50. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
51. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
52. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204–1214 (2013).
53. Agirre, J. et al. The CCP4 suite: integrative software for macromolecular crystallography. *Acta Crystallogr. D* **79**, 449–461 (2023).
54. Tickle, I. J. et al. STARANISO (Global Phasing Ltd, 2018).
55. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
56. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
57. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
58. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
59. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
60. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
61. Agirre, J. et al. Privateer: software for the conformational validation of carbohydrate structures. *Nat. Struct. Mol. Biol.* **22**, 833–834 (2015).
62. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
63. Landau, M. et al. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299–W302 (2005).
64. Gucwa, M. et al. CMM—an enhanced platform for interactive validation of metal binding sites. *Protein Sci.* **32**, e4525 (2023).
65. Sonnenburg, E. D. et al. Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell* **141**, 1241–1252 (2010).

Acknowledgements This work was supported by operating grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institutes of Health Research (CIHR) as well as infrastructure funds from the Canada Foundation for Innovation and BC Knowledge Development Fund. S.A.N. was supported by a Vanier scholarship from NSERC; A.C.L. by a GAP scholarship from the UBC Centre for Blood Research; and L.B. by an Early Postdoc Mobility grant from Swiss National Science Foundation. N.C.J.S. is a Tier I Canada Research Chair in Structure-guided Antibiotic Discovery. Part of the research described in this paper was performed using beamline CMCF-BM at the Canadian Light Source, a national research facility of the University of Saskatchewan, which is supported by the Canada Foundation for Innovation (CFI), the Natural Sciences and Engineering Research Council (NSERC), the National Research Council (NRC), the Canadian Institutes of Health Research (CIHR), the Government of Saskatchewan and the University of Saskatchewan. We thank the GM/CA beamline staff at beamline 23-ID-B at the APS for access and support. GM/CA@APS has been funded in whole or in part by Federal funds from the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006). We thank A. M. Deutschbauer for providing access to the B.T. mutant strains[18]; T. Warren for the initial inspiration for these studies; P. Rahfeld for guidance about the screening process; A. Noonan and S. Hallam for guidance with processing of the sequencing data; M. Ezhova for assistance with the NMR experiments; and F. Liu, S. Macdonald, P. Danby, B. Herring, C. Olagnon, J. Wardman, R. Bains and Y. Tian for suggestions and discussions.

Author contributions S.G.W., S.A.N. and L.B. conceived and developed the screen. Screening was performed by I.L.L. and S.A.N.; S.A.N., C.Y.Z. and E.B. performed the cloning, purification and biochemical characterization of the enzymes. S.A.N., I.L.L., C.Y.Z. and H.-M.C. synthesized all of the chemicals. A.C.L., L.S., L.J.W. and N.C.J.S. performed the crystallography analysis. S.A.N. and D.P. performed microbial growth and RT–qPCR assays under the guidance of H.B.; S.A.N., A.C.L. and S.G.W. wrote the manuscript with input from all of the authors.

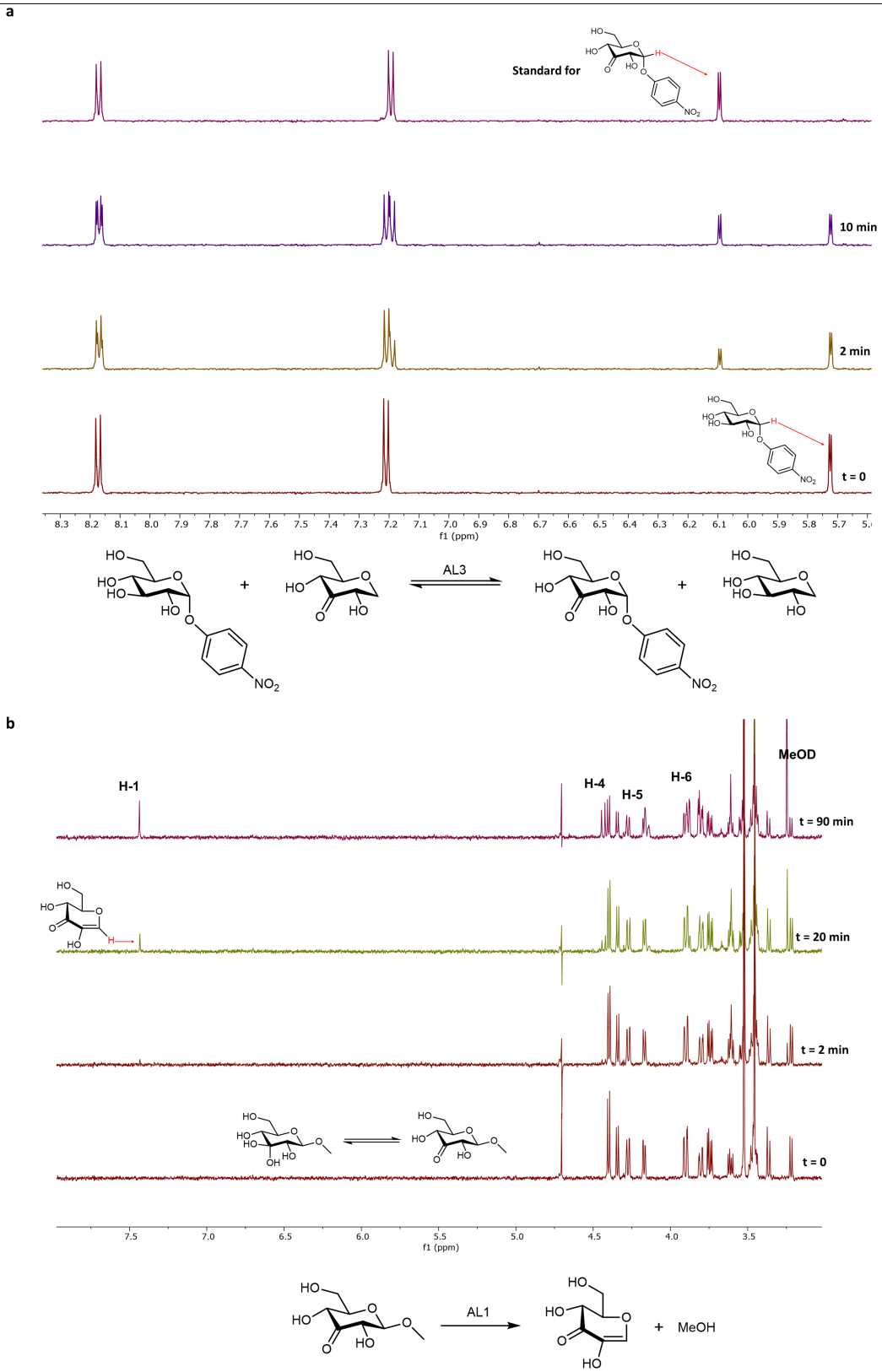
Competing interests The authors declare no competing interests.

Additional information

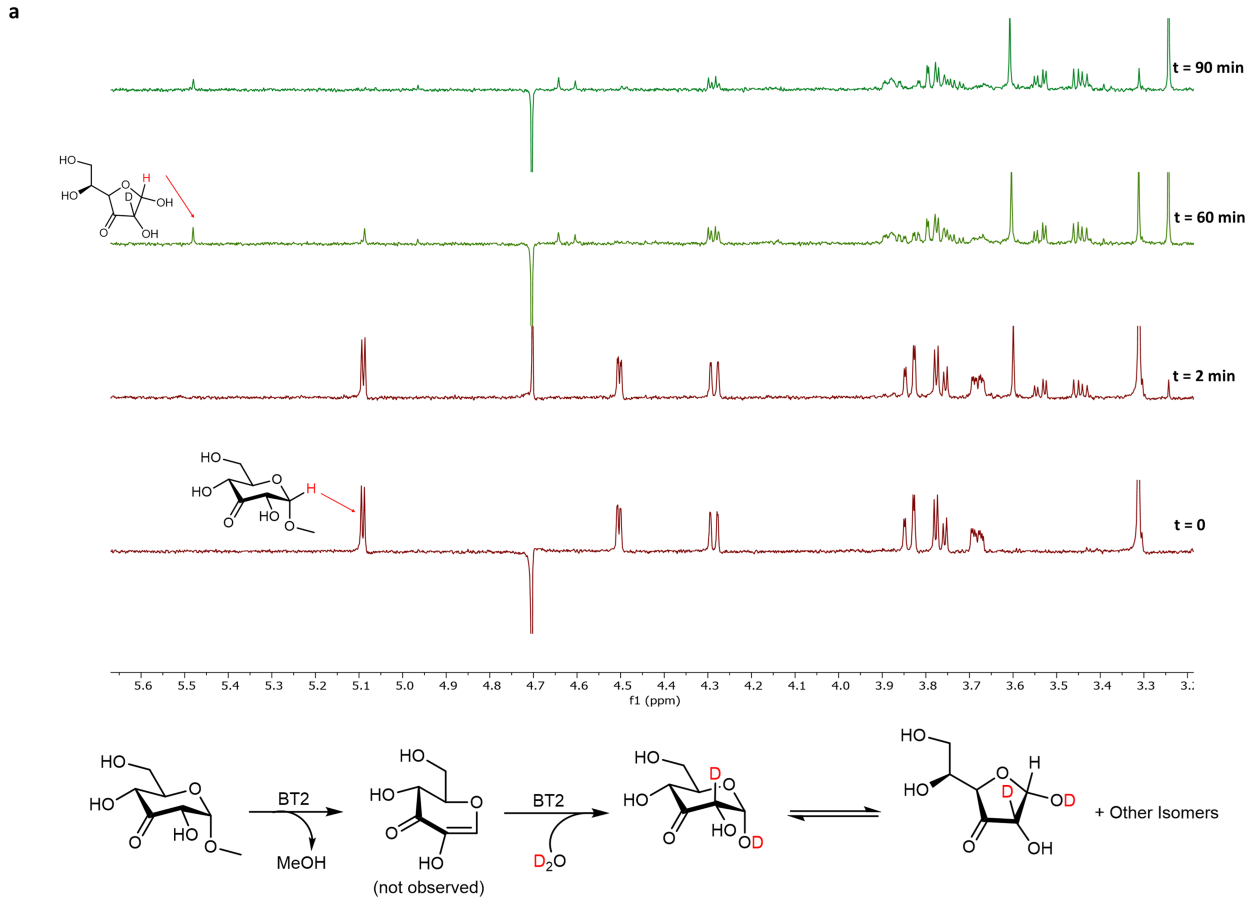
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-07574-y.

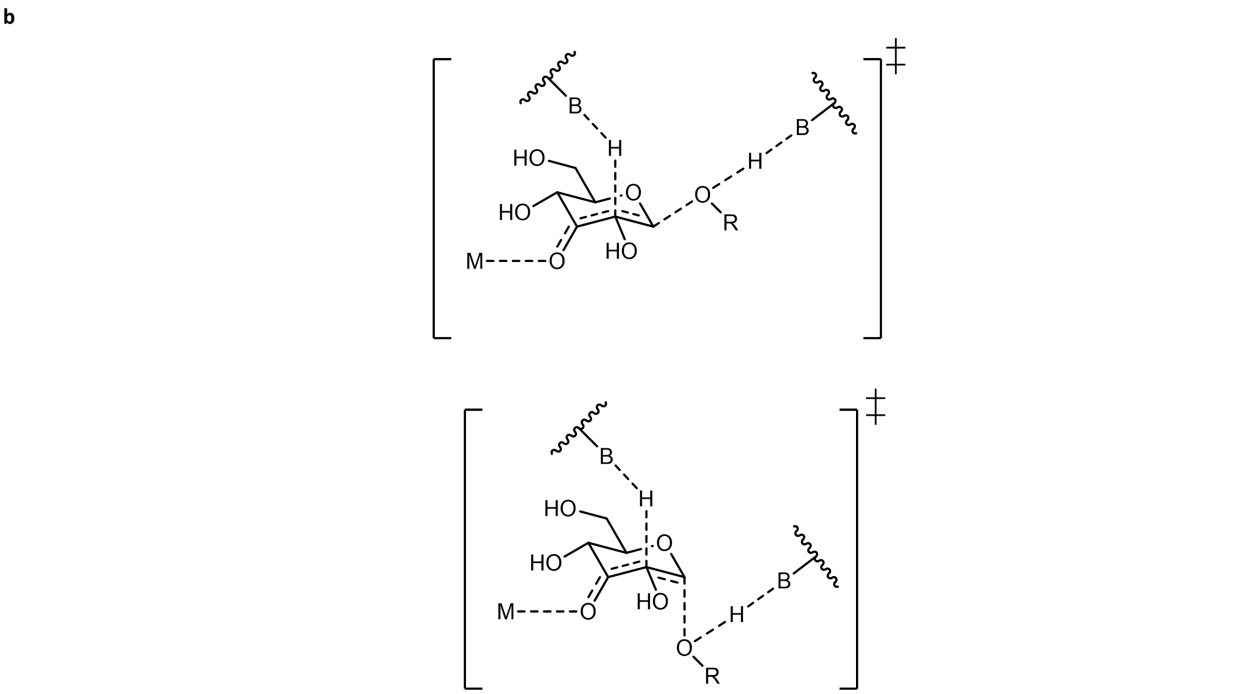
Correspondence and requests for materials should be addressed to Stephen G. Withers.

Peer review information *Nature* thanks Sebastian Hiller and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

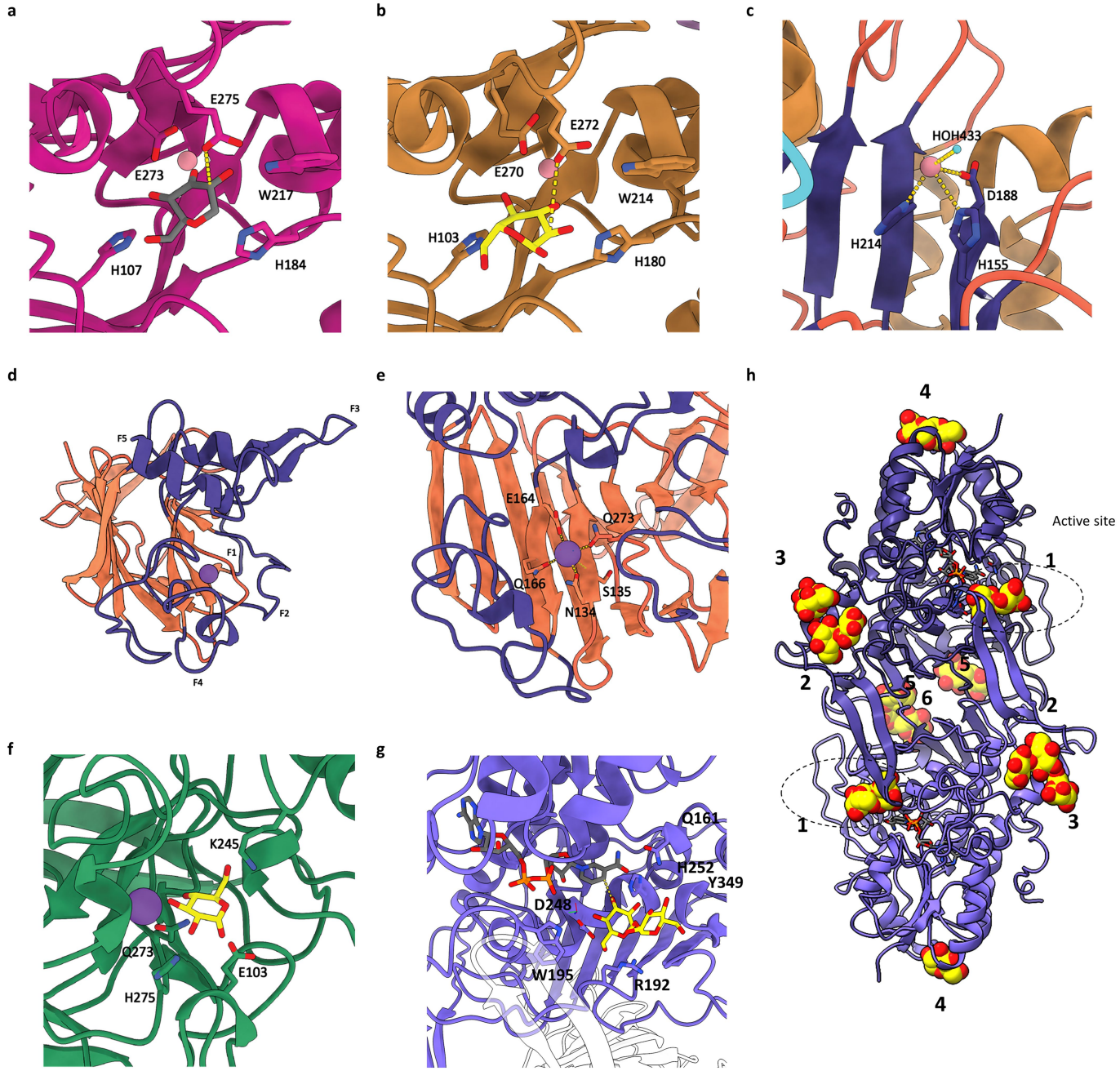


Extended Data Fig. 1 | Reactions catalysed by AL/BT3 and AL/BT1. ^{1}H-NMR characterization of the reactions of a) AL3 and b) AL1.

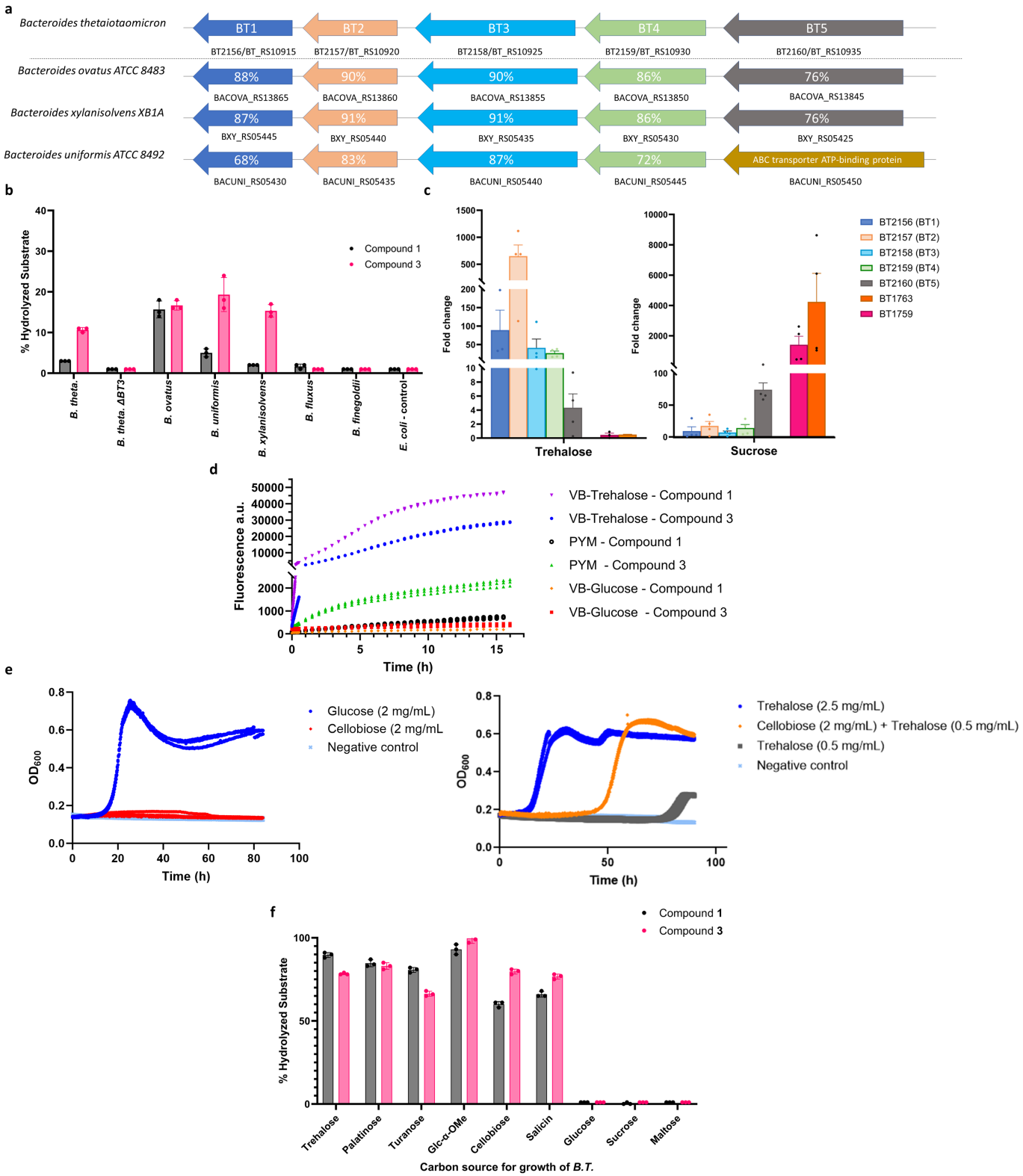




Extended Data Fig. 2 | Reaction catalysed by AL/BT2 and proposed transition states for AL/BT1 and 2. a) ^{1}H-NMR characterization of the reaction of BT2 and b) proposed transition states for AL/BT1 (top) and AL/BT2 (bottom), B refers to protein residues capable of acid/base catalysis and M refers to the metal ions that coordinate the 3-keto groups.

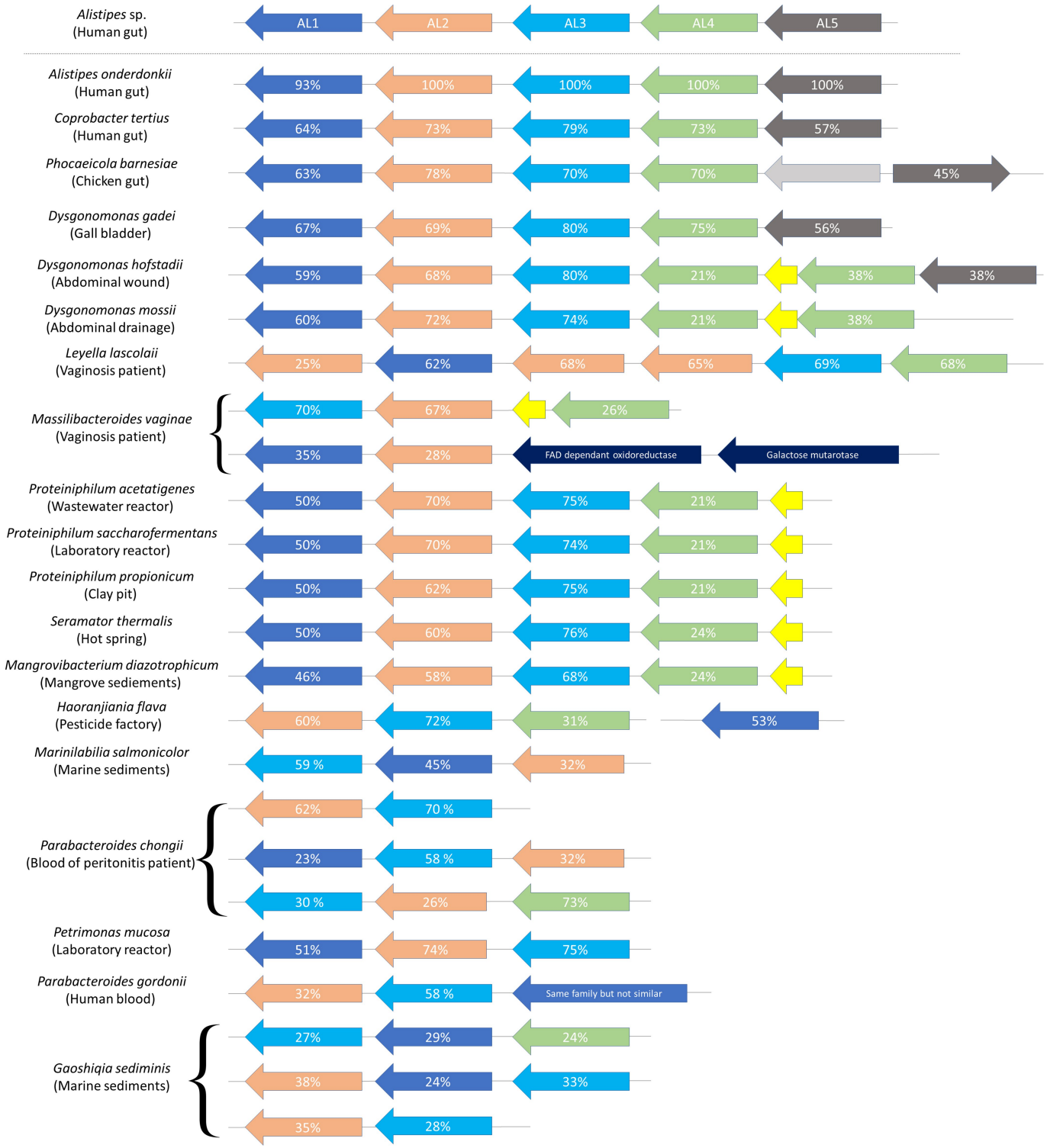


Extended Data Fig. 3 | Additional structural figures for AL/BT 1-3. a) Active site of BT1 bound with substrate analogue 3k-1,5-anhydroglucitol with coordinating residues labelled. b) Active site of AL1 bound with glucose with coordinating residues labelled. c) Metal binding site of apo AL1 with coordinating residues labelled. d) Front view of overall fold of BT2 with jelly roll in orange, fingers (purple) labelled F1-F5, and active site denoted with a dashed circle. e) Metal binding site of apo BT2 with coordinating residues labelled. f) BT2 active site with conserved possible active site residues labelled. g) Active site of AL3 bound with trehalose with key residues labelled. h) Binding sites of trehalose observed in AL3 crystal soaked with trehalose. Active site is labelled as 1 and circled, secondary sites are labelled 2–6.



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | In vivo hydrolysis of substrates by *Bacteroides* species. a) Homologues of the proteins from *Bacteroides thetaiotaomicron* in some other *Bacteroides* strains. The percent similarity of each of the encoded proteins to the corresponding BT homologue is shown in white. b) The percent of hydrolysed selective non-Koshland substrates **1** and **3** by different *Bacteroides* strains. All the strains were grown anaerobically and incubated with the substrates overnight. Bars represent the mean and error bars represent standard deviations (n = 3 technical replicates, derived from the same original culture). c) Transcriptional response of BT genes BT2156-60 to trehalose and sucrose as sole carbon sources. Representative genes from the fructan utilization locus (BT1759 and BT1763, encoding GH32 and SusC homologues, respectively) serve as positive controls for sucrose regulation[65]. Data were acquired at mid-log phase (9 h for trehalose, 5 h for sucrose) and are normalized to transcript levels prior to inoculating cells into the respective carbohydrate medium. Bars represent the mean and error bars represent the standard error of the mean (n = 4 replicates, two technical replicates each derived from two separate original cultures). d) Resulting fluorescence from hydrolysis of substrates **1** and **3** by *B. thetaiotaomicron* cells grown in different media. Growth on α,α-trehalose results in significantly higher activities while growth on glucose represses them, (n = 3 technical replicates, derived from the same original pre-culture, data for all replicates are shown.) e) Growth of *B. thetaiotaomicron* cells on cellobiose in absence and presence of trehalose, performed in triplicate for all the conditions shown here. f) hydrolysis of substrates **1** and **3** by *B. thetaiotaomicron* cells grown on different glycosides as the sole carbon source. Bars represent the mean and error bars represent standard deviations (n = 3 technical replicates, derived from the same original culture).



Extended Data Fig. 5 | Homologues of the proteins from P1C11 in bacteria from different environments. Homologues of the proteins from P1C11 (*Alistipes* sp.) identified in bacteria from various environments (the source from which each bacterium is isolated is noted in parentheses below the name of bacterium). The percent similarity of each of the encoded proteins to the corresponding *Alistipes* homologue is shown in white. Dark blue denotes oxidoreductases that are not from the same family as AL3/AL4 and denoted in yellow are commonly co-occurring genes annotated as a DoxX-domain-containing protein, which is reported to form membrane-associated oxidoreductase complexes. Broken lines mean that a few genes are in between those shown and light grey shows other genes. The length of the arrows and the space between them does not correspond to the actual length of the DNA fragments in this figure. This is a representative list not an exhaustive one.

Extended Data Fig. 6 | Homologues of the proteins from P2B11 in bacteria from different environments. Homologues of the proteins from P2B11 identified in bacteria from various environments (the source from which each bacterium is isolated is noted in parentheses below the name of bacterium). The percent similarity of each of the encoded proteins to the corresponding P2B11 homologue is shown in white. Broken lines mean that a few genes are in between those shown and light grey shows other genes. The length of the arrows and the space between them does not correspond to the actual length of the DNA fragments in this figure. This is a representative list not an exhaustive one.

Extended Data Table 1 | (Pseudo)-Michaelis-Menten kinetic parameters for enzymes

Enzyme(s)	Substrate	Co-substrate (conc.)	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})
AL1	3K-Glc-α-pNP	-	2250 ± 400	50 ± 7	(2.2 ± 0.5) x 10^4
AL2	3K-Glc-β-pNP	-	160 ± 33	4.8 ± 0.1	(3.0 ± 0.6) x 10^4
AL3	Glc-α-pNP	3K-1,5-anhydroglucitol (10 mM)	6700 ± 3500	48 ± 13	(7.2 ± 4.2) x 10^3
AL1,2,3	Glc-β-MU	3K-Glc-β-OMe (50 μM)	231 ± 42	0.39 ± 0.02	(1.7 ± 0.3) x 10^3
AL1,2,3	Glc-α-MU	3K-Glc-β-OMe (50 μM)	20 ± 2	0.82 ± 0.02	(4.1 ± 0.4) x 10^4
AL1,2,3	Compound **3**	3K-Glc-β-OMe (50 μM)	114 ± 33	0.10 ± 0.02	(8.8 ± 3.1) x 10^2
BT1	3K-Glc-β-pNP	-	800 ± 70	17.6 ± 0.6	(2.2 ± 0.2) x 10^4
BT2	3K-Glc-α-pNP	-	26 ± 2	9.9 ± 0.1	(3.8 ± 0.3) x 10^5
BT1,2,3	Glc-β-MU	3K-Glc-β-OMe (50 μM)	66 ± 12	0.19 ± 0.01	(2.9 ± 0.5) x 10^3
BT1,2,3	Compound **1**	3K-Glc-β-OMe (50 μM)	18 ± 3	0.26 ± 0.02	(1.4 ± 0.3) x 10^4

Note that all kinetic parameters were measured at room temperature. Detailed conditions for the experiments can be found in the S.I., sections 2.5.3., 2.5.4.3., and 2.7.12. 3K- denotes a 3-keto group and the chemical structure of all substrates are shown in Scheme S14. Note that all these numbers are approximations, since I) because of the keto-hydrate equilibrium, the actual substrate concentrations are less than those presumed for AL/BT1 II) for AL/BT2 the enzyme will catalyse both elimination and hydration reactions, and we only observe the first step III) whenever AL/BT3 is present, a co-substrate is used which is also an inhibitor/slow substrate of the other enzymes. Moreover, whenever more than one enzyme is used these numbers do not reflect accurate kinetic parameters of any enzyme but rather are 'observed' kinetic parameters. IV) for compounds 1 and 3, observation of the fluorophore relies on spontaneous breakdown of the *S,O*-acetal linker, which we have assumed is faster than enzymatic steps and therefore does not affect the observed rates.

Extended Data Table 2 | Data collection and refinement statistics

	AL1	AL1 Glucose	AL1 Trehalose	BT1 3K-GlcH	AL2	BT2	BT2 Glucose	AL3	AL3 Trehalose‡	P2B11-Oxido-1
Data collection										
Space group	P 21 21 21	P 21 21 21	P 21 21 21	P 32 2 1	P 31 1 2	P 21 21 21	P 21 21 21	I 1 2 1	I 1 2 1	P 1 21 1
Cell dimensions										
a, *b*, *c* (Å)	47.57, 112.61, 209.34	47.517, 113.68, 208.58	47.99, 113.55, 212.44	122.29, 122.29, 68.75	95.09, 95.09, 159.54	75.94, 80.02, 103.15	77.04, 81.17, 103.78	180.64, 56.75, 220.45	186.58, 57.22, 224.96	106.26, 185.88, 145.36
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 120	90, 90, 120	90, 90, 90	90, 90, 90	90, 108.87, 90	90, 106.59, 90	90, 111.20, 90
Resolution (Å)	46.39 – 1.90 (1.97-1.90)	47.40 – 2.08 (2.15 – 2.08)	44.30 – 1.50 (1.55 – 1.50)	45.69 – 1.86 (1.93 – 1.86)	47.59 – 2.65 (2.72 – 2.65)	61.15 – 1.46 (1.51 – 1.46)	43.72 – 1.85 (1.92 – 1.85)	48.00 – 2.10 (2.18 – 2.10)	55.31 – 2.95 (3.06 – 2.95)	48.37 – 2.60 (2.69 – 2.60)
R_{merge}	0.161 (1.562)	0.135 (0.978)	0.071 (1.303)	0.082 (2.707)	0.145 (2.106)	0.152 (1.223)	0.182 (1.051)	0.181 (1.583)	0.141 (0.469)	0.120 (1.721)
$I / \sigma I$	9.69 (1.26)	8.93 (1.66)	11.04 (1.33)	28.97 (1.16)	22.90 (1.90)	13.60 (1.29)	13.18 (1.81)	8.68 (1.58)	3.51 (2.20)	13.73 (1.11)
CC1/2	0.996 (0.557)	0.997 (0.603)	0.999 (0.575)	1 (0.619)	0.999 (0.868)	0.991 (0.506)	0.985 (0.615)	0.992 (0.557)	0.95 (0.497)	0.998 (0.507)
Completeness (%)	99.35 (94.72)	99.68 (98.18)	99.80 (99.90)	99.93 (99.86)	99.12 (97.72)	99.87 (99.94)	99.77 (99.96)	99.66 (99.66)	89.62 (63.96)	99.94 (99.98)
Redundancy	6.4 (6.2)	6.3 (5.3)	6.4 (6.6)	19.9 (19.7)	18.9 (18.5)	6.4 (5.9)	6.5 (6.4)	6.6 (6.8)	2.7 (2.4)	6.9 (6.9)
Refinement										
Resolution (Å)	1.90	2.08	1.50	1.86	2.65	1.46	1.85	2.10	2.95	2.60
No. reflections	89195 (8375)	68822 (6689)	186072 (18379)	49898 (4903)	24911 (3236)	109360 (10835)	56117 (5543)	123787 (12234)	43691 (3079)	161056 (16054)
R_{work} / R_{free}	0.184/0.226	0.177/0.227	0.150/0.194	0.154/0.164	0.197/0.251	0.168/0.209	0.189/0.236	0.199/0.233	0.196/0.246	0.187/0.232
No. atoms										
Protein	8672	8644	8685	2226	7810	4075	4058	14249	14220	36746
Ligand/ion	36	130	100	42	2	2	14	176	705	662
Water	512	318	870	226	40	553	398	414	140	255
B-factors (Å²)										
Protein	25.3	25.5	18.2	44.7	63.1	24.0	33.4	53.7	42.6	74.1
Ligand/ion	35.5	41.9	27.0	86.7	43.6	21.0	52.0	44.8	46.9	74.3
Water	31.7	33.3	33.4	57.4	45.5	31.7	36.4	37.6	23.4	47.7
R.m.s. deviations										
Bond lengths (Å)	0.015	0.016	0.015	0.012	0.012	0.014	0.015	0.016	0.012	0.013
Bond angles (°)	2.02	2.10	1.89	1.83	2.40	1.88	2.08	2.08	1.96	2.09

Statistics for the highest resolution shell are shown in parentheses.
‡Anisotropic cut-off applied to merged intensity data[54].

nature portfolio

Corresponding author(s): S. A. Nasseri, S.G. Withers

Last updated by author(s): Apr 19, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	For the assays with BioTek plate reader: Gen 5 2.09 (BioTek). X-ray data collection programs: For CMCF-BM MxDC, APS 23-ID-D JBluIce, For NMR data: TopSpin 4.2.0 (Bruker)
Data analysis	Microsoft Excel 2022, GraphPad Prism 10.2.2., GraFit v7.0.3, MestReNova 14.2.1., XDS VERSION Jan 10, 2022 BUILT=20220820, Xia2 3.15.0, CCP4i2 2.0.14, Pointless 1.12.15, Aimless 0.7.9, Alpafold2 v1.5.2, CCP4i2 Process Predicted Models (CCP4i2 suite 2.0.14), Phaser 2.8.3, Refmac 5.5, Privateer version MKIV, PISA v2.1.0, Molprobity 4.2, ChimeraX 1.5, CheckMyMetal Server 2.1, The ConSurf Server v1.00 2019, Phenix Generate "Table 1" for Journal 1.20.1-4487, STARANISO v3.350 Server, Coot 0.9.8.93

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The authors declare that the data necessary to support the findings of this study are available within the main text, extended data, and supplementary information. The raw data for the information presented in the main text figures and extended data are provided as an excel file (Source data). All the sequence data are included in the SI and also available from NCBI with accession numbers PP693324 and PP693325. Structural models are deposited with the protein data bank with accession codes 8TCD, 8TCR, 8TCS, 8TCT, 8TDA, 8TDE, 8TDF, 8TDH, 8TDI, 8V31.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Does not apply in this study as the study does not use human research participants.
Reporting on race, ethnicity, or other socially relevant groupings	Does not apply in this study as the study does not use human research participants.
Population characteristics	Does not apply in this study as the study does not use human research participants.
Recruitment	Does not apply in this study as the study does not use human research participants.
Ethics oversight	Does not apply in this study as the study does not use human research participants.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were generally set to n=3 for most experiments, as is the convention in the field for probing the direct interactions between enzymes and substrates.
Data exclusions	For refinement of X-ray crystallography models, 5% of reflections were omitted for calculation of Rfree values to minimize model bias. For the RT-qPCR assay, three of the replicates failed to show any CT value and are thus not included in calculations.
Replication	Unless otherwise stated, all experiments were done with 3 technical replicates. All attempts to replicate were successful.
Randomization	Randomization is not relevant as this study does not involve clinical trials.
Blinding	Blinding is not relevant as this study does not involve clinical trials.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*