Structure of an iminosugar complex of a glycoside hydrolase family 5 lichenase provides insights into the active site

Puneet Garg and Narayanan Manoj*

Department of Biotechnology

Bhupat and Jyoti Mehta School of Biosciences

Indian Institute of Technology Madras

Chennai – 600036, India

* Corresponding author: Narayanan Manoj

Email: nmanoj@iitm.ac.in

Running title: Active site features in the GH5_36 subfamily

Highlights

- Lichenase TmCel5B belongs to the dual-specific GH5_36 subfamily
- Two crystal structures of TmCel5B complexed with 1-deoxynojiromycin and Tris
- First crystal structure of a cognate sugar complex of a GH5_36 homolog
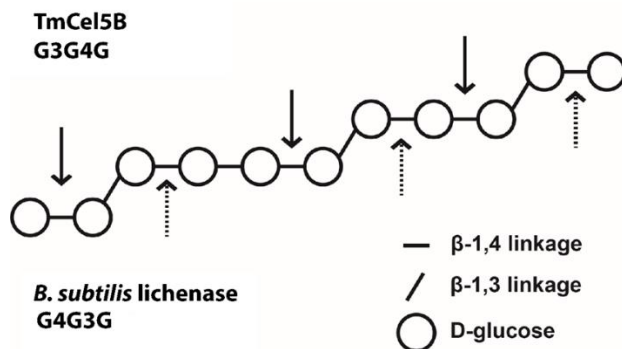- Active site clefts are divergent within the GH5_36 subfamily

Abstract

TmCel5B is a lichenase belonging to glycoside hydrolase family 5 subfamily 36 (GH5_36). To gain insights into the active site of this subfamily which contains multifunctional endoglycanases, we determined the crystal structure of TmCel5B in complex with an iminosugar, 1-deoxynojiromycin (DNJ). DNJ is bound to the -1 subsite, making a network of non-covalent interactions with the acid/base residue Glu139, the nucleophile Glu259, and with other residues that are conserved across the GH5 family. The catalytic site displayed a Glu-Arg-Glu triad of the catalytic glutamates that is unique to the GH5_36 subfamily. Structural comparison of active sites of GH5_36 homologs revealed divergent residues and loop regions that are likely molecular determinants of homolog-specific properties. Furthermore, a comparative analysis of the binding modes of iminocyclitol complexes of GH5 homologs revealed the structural basis of their binding to GH5 glycosidases, in which the subsite binding location, the interactions of the ligand with specific conserved residues, and the electrostatic interactions of the catalytic glutamates with the ring nitrogen, are crucial.

1. Introduction

About 1-3% of genes present in most organisms code for carbohydrate-active enzymes, consistent with the large structural diversity of their natural carbohydrate substrates [1]. Among these enzymes, the Glycoside Hydrolases (GHs) constitute the largest class with 173 different GH families listed in the CAZy database [2]. The GH5 family is among the largest and most diverse GH families with respect to the range of substrates. These enzymes are active on diverse oligo- and poly-saccharide substrates, including cellulose, lichenin, mannan, xylan, arabinoxylan, laminarin, and dextran, among others. GH5 enzymes contain the common $(\beta/\alpha)_8$ barrel protein fold, and at least 29 types of experimentally determined enzymatic activity types are reported within the family [3]. This variability in function across homologs occurs due to the differences in the active site that is primarily built by the loop regions that connect the conserved β-strands and α-helices [4,5].

The GH5 family is classified into 54 distinct subfamilies based on a large-scale phylogenetic analysis in conjunction with biochemical and structural data [6]. While some of these subfamilies are monospecific, several subfamilies display poly-specificity and accommodate a range of substrates. Moreover, the experimental activities of several subfamilies are yet to be determined, and several sequences remain unclassified. The GH5_36 subfamily comprises bacterial and archaeal homologs that are putative endo-glycanases and only three homologs have been biochemically characterized to date. Two enzymes exhibit endo-β-1,4-mannanase activity with minor endo-β-1,4-glucanase activity (EC 3.2.1.78) [7]. Among these two, the structure of the *Caldanaerobius polysaccharolyticus* mannanase (CpMan5B) bound to Tris has been reported at 1.6 Å (PDB ID: 3W0K) [8]. The structure of the complex provided insights into residues that likely contribute to catalysis and specificity. The third enzyme, TmCel5B, from *Thermotoga maritima* has a demonstrated lichenase activity (EC 3.2.1.73) and is devoid of cellulase activity. Lichenase is an endo-β-glucanase specific to catalysing the hydrolysis of (1,4)-β-glucosidic linkages in cereal β-D-glucans containing mixed β-(1,3)- and β-(1,4)-linkages. Moreover, the cleavage occurs on the β-(1,4)-linkage strictly after the β-(1,3)-linkage, with the product having a β-(1,3)-linkage at the reducing end. Interestingly, TmCel5B is reported to be a unique lichenase, producing glucosyl β-(1,3) glucosyl β-(1,4) glucose, instead of the usual glucosyl β-(1,4) glucosyl β-(1,3) glucose (Scheme 1) [9]. Moreover, the gene expression profile of glucomannan-inducible TmCel5B suggests probable mannanase activity as well [10].

Scheme 1. Scheme of cleavage patterns for classical lichenase and TmCel5B on lichenin substrate. Lichenan is a mixed linkage glucose polysaccharide contain β-1,3 and β-1,4 linkages. The classical lichenase represented by *B. subtilis* lichenase, cleaves the lichenin polysaccharide after β-1,3 glycosidic bond and produces a triose and tetraose sugar (glucosyl β-1,4 glucosyl β-1,3 glucose (G4G3G) and glucosyl β-1,4 glucosyl β-1,4 glucosyl β-1,3 glucose (G4G4G3G) respectively). TmCel5B has a unique lichenin hydrolysis profile and produces glucosyl β-1,3 glucosyl β-1,4 glucose (G3G4G), instead of the usual glucosyl β-1,4 glucosyl β-1,3 glucose (G4G3G) [9].

The crystal structure of the apo form of TmCel5B is available in the PDB (PDB ID: 1VJZ, Joint Centre for Structural Genomics, unpublished). Thus, the structural determinants of substrate binding within the subfamily are poorly described due to the lack of a structure of a cognate sugar/analogue complex. Substrates/products and their analogues, transition state analogues, and inhibitors, are commonly used to probe the substrate binding sites and identify the subtle variations in binding interactions employed by diverse GHs. Iminosugars and their analogues are a major group of reversible competitive glycosidase inhibitors that resemble saccharides and contain an endocyclic nitrogen atom. These compounds are known to act by mimicking properties of the glycosidase oxocarbenium-ion-like transition state [11–13]. Ligand trapping of iminosugars and their derivatives in glycosidase crystal structure complexes can thus shed light on the geometry of the active site interactions and help establish the structural basis of transition-state stabilization and inhibition. In this study, we report the crystal structure of TmCel5B bound to a 1-deoxyglucose analogue, 1-deoxynojirimycin (DNJ), at 1.8 Å resolution. DNJ is a naturally occurring iminosugar and together with its derivatives, iminosugars used as glycosidase inhibitors are promising therapeutic agents in pharmaceutical applications [14–17]. The first such sugar-bound complex in the dual-specific GH5_36 subfamily provides insights into the catalytic residues and active site features that may be determinants of substrate binding and specificity.

2. Material and Methods

2.1 Protein expression and purification

TmCel5B (gene *TM1752*, UniProt ID: Q9X274) is 329 amino acids long. The clone procured from the DNASU plasmid repository (Clone ID: TmCD00396393) is in an ampicillin-resistant pMH1 vector with expression under the control of an arabinose-inducible araBAD promoter. The protein contains an N-terminal 12-residue expression and purification tag (MGSDKIHHHHHH). *E. coli* harbouring the plasmid were grown in 400 ml of LB broth in a 1 L flask with shaking at 37 °C. Expression was induced at OD_{600} ~0.6 by the addition of L-(+)-arabinose to a final concentration of 0.2% (w/v) and the cells were grown for 4 h. The subsequent steps were carried out at 4 °C unless otherwise stated. The pelleted cells suspended in lysis buffer (20 mM Tris-HCl, 10 mM imidazole, 300 mM NaCl, pH 7.5) were lysed by sonication. The cell extract was clarified by centrifugation at 10,000 rpm for 10 min. The supernatant was incubated at 65 °C for 20 min in a water bath and centrifuged at 12,000 rpm for 45 min to precipitate heat-labile *E. coli* proteins. The clear lysate was loaded on an Ni-Nitrilotriacetic Acid (Ni^{2+}-NTA) affinity column, pre-equilibrated with lysis buffer. The column was washed with 5 column volumes of wash buffer (20 mM Tris-HCl, 30 mM imidazole, 300 mM NaCl, pH 7.5) and the protein eluted using elution buffer (20 mM Tris-HCl, 300 mM imidazole, 300 mM NaCl, pH 7.5). The purified sample was applied to a HiPrep 26/10 desalting column (GE Healthcare) pre-equilibrated with storage buffer (10 mM Tris-HCl, 25 mM NaCl, pH 7.5) to remove excess imidazole. The molecular weight and purity of the protein were analysed by SDS-PAGE. The concentration of the protein was determined by UV spectroscopy using a theoretical molar extinction coefficient, $\varepsilon = 100380 \ M^{-1} \ cm^{-1}$ at 280 nm. The oligomeric state of TmCel5B was estimated by size-exclusion chromatography (SEC) performed at 4 °C using a SuperdexTM 200 10/300 GL analytical column (GE Healthcare). 50 µl of protein (2 mg ml^{-1}) was loaded onto the column pre-equilibrated with 25 mM NaCl and 10 mM Tris-HCl (pH 7.5) and eluted at a flow rate of 0.4 ml min^{-1}. The chromatogram was monitored by measuring absorbance at 280 nm.

2.2 Protein crystallization, data collection, structure determination and refinement

The enzyme was concentrated to 20 mg ml^{-1}. Initial crystallization screening was performed using commercial screens from Hampton Research (HR2-144, HR2-110 and HR2-112) at 20 °C using the hanging drop vapour diffusion method. 1 µl of protein solution was mixed with 1 µl of reservoir solution and equilibrated with 500 µl reservoir solution. For co-crystallization trials of the TmCel5B-DNJ complex, 25 µl (~0.5 mM) protein was first incubated with 1 µl of 100 mM stock solution of DNJ at 20 °C for 90 min prior to setting up the vapour diffusion drop. Bar shaped crystals were obtained in 100 mM sodium citrate tribasic dihydrate (pH 5.8-

6.2), 8-14% (v/v) isopropanol and 8-16% (w/v) PEG 4000 within 4-6 days. X-ray diffraction datasets were collected at 100 K on a Bruker MICROSTAR rotating anode with MAR345 image plate detector. The data were processed and scaled using MOSFLM and AIMLESS programs from the CCP4 software package [18–20]. The structures were solved by the molecular replacement method using the PHASER program. One subunit of the apo form of TmCel5B (PDB: 1VJZ) was used as the search model. Refinement was performed using the maximum likelihood refinement program implemented in the PHENIX software suite [21]. Model building was carried out using the COOT program [21–23]. The MolProbity program was used to assess the stereochemical quality of the final structure [24]. The summary of the data collection, refinement, and validating statistics is provided in Table 1.

2.3 Sequence and structural analysis

TmCel5B structure was used to carry out a structural homology search using the DALI program. The outcome was compared with the GH5 structures (95 unique proteins) available in the CAZy database and a structure-based sequence alignment was generated [2,25]. Next, 28 sequences from the GH5_36 subfamily were identified from the CAZy database and were added to extend the previous alignment using MAFFT [26]. The final sequence alignment was modified using ESPRIPT [27]. The represented sequence logos of motifs were generated using the WebLogo software [28]. Ligplot and PyMOL programs were used to analyse various structures and generate figures [29,30].

3. Results and Discussion

3.1 Overall structure of TmCel5B

Recombinant TmCel5B was purified to homogeneity and is a monomeric enzyme with a native molecular weight of ~39 kDa based on SEC analysis (Fig. S1). Next, we carried out crystallization trials to obtain the apo form and Michaelis complexes with monosaccharide ligands and analogues. In this study, a new crystal form of TmCel5B belonging to the orthorhombic space group $P2_12_12_1$ was obtained. The asymmetric unit contains two independent subunits. This form is different from that of the previously reported tetragonal form of apo TmCel5B with one subunit in the asymmetric unit (PDB:1VJZ, 2.05 Å resolution, unpublished). The subunit comprises residues 6-329. Residues corresponding to the first 5 residues and the expression tag were not built into the model because of disorder in the electron density.

The structure of TmCel5B is a typical $(\beta/\alpha)_8$ TIM-barrel fold, with parallel strands forming an inner barrel, and flanking α-helices. The overall fold is like that observed in all GH5 homologs. However, TmCel5B displayed an incomplete barrel, where the region corresponding to the canonical α5 helix is replaced by a loop region A5 (Fig. 1, Fig. S2). The equivalent region in GH5_36 β-mannanase/endoglucanase, CpMan5B, is a loop region containing a short 3_{10} helix. A similar feature has earlier been identified in some other GH5 enzymes such as CtCel5E from *C. thermocellum* [31], E1CD from *A. cellulolyticus* [32], Exg from *C. albicans* [33] and MeMan5A from *M. edulis* [34]. Besides, TmCel5B contains additional major structural elements that decorate the core fold. These regions, mostly loops connecting the core secondary structure elements, are divergent compared to homologs from other GH5 subfamilies (Fig. S3). In a structural homology search of the PDB, TmCel5B exhibits the highest structural similarity to CpMan5B as expected (3W0K; 42% sequence identity, rmsd of 1.7 Å over 324 Cα atoms) [8,25]. The next closest structural and sequence neighbours among the GH5 are the endoglucanase CtCelC from *C. thermocellum* (1CEN; 28% identity, rmsd of 2.2 Å over 334 Cα atoms) of subfamily GH5_37 [31] and members of the GH5_25 subfamily. Among the GH5_25 homologs include the well-characterized dual-specificity mannanase/endoglucanase TmCel5A from *T. maritima*, (PDB: 3AMD; 23% identity, rmsd of 2.3 Å over 312 Cα atoms) [35]. Together, the GH5_37 and the GH5_25 subfamilies contain one and five structurally characterized members, respectively (Fig. S4). This structural similarity is consistent with a previous sequence-based subfamily level phylogenetic classification where the GH5_37 and GH5_25 subfamilies are closely related to GH5_36 [6].
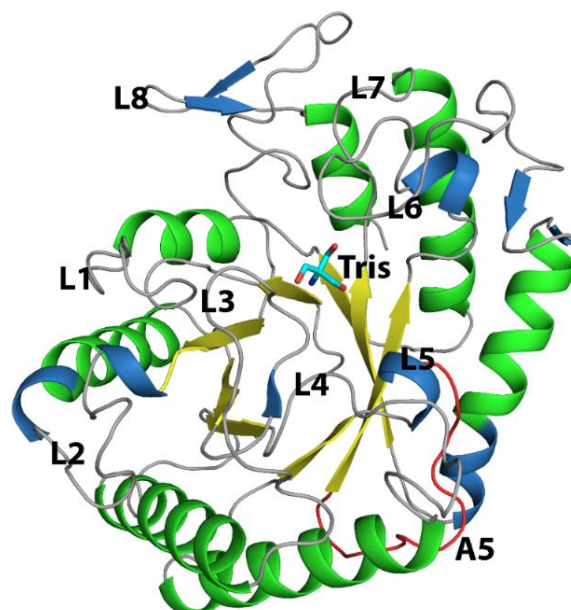
Figure 1. Structure of TmCel5B. Ribbon representation of the TmCel5B-Tris complex (PDB: 7EFZ). The structure displays an incomplete $(\alpha/\beta)_8$ barrel fold, with the canonical $\alpha5$ helix replaced by an irregular loop (A5, red). The secondary structure elements of core fold are shown in green (helix) and yellow (strands). Additional secondary structure elements are displayed in blue, and loops are shown in grey color. Bound Tris molecule at the -1 subsite is shown in stick representation with cyan, red and blue color for carbon, oxygen and nitrogen, respectively. Loop regions connecting the $(\alpha/\beta)_8$ secondary structure elements and constituting the substrate binding site at the C-terminal end of the barrel are labelled L1-L8.

TmCel5B residues Glu139 (on loop connecting $\beta4$ and $\alpha4$) and Glu259 (on the C-terminal end of $\beta7$) were identified as the catalytic acid/base and nucleophile, respectively, based on the structure-based sequence alignment (Fig. S4). The distance between the carboxylate carbon atoms of the catalytic residues is 4.4 Å, demonstrating that TmCel5B adopts a catalytic mechanism typical of the retaining type GH5 enzymes [36–38]. Interestingly, electron density corresponding to a Tris molecule was identified at the active site of the structure that was refined at 1.69 Å (Fig. 2b). Tris, a purification buffer component, is a known competitive inhibitor in several GHs since the hydroxyl groups mimic the sugar hydroxyls. The bound Tris molecule interacts with multiple residues of the expected -1 subsite. The charge on a protonated amino nitrogen which mimics the oxocarbenium ion-like transition state, is at a distance of 2.6 Å from the catalytic nucleophile, Glu259 [36]. Also, a water-mediated H-bond is present between the catalytic acid/base residue Glu139 and Tris. The structure of the TmCel5B-Tris complex (PDB: 7EFZ) is largely identical to that of the apo form of TmCel5B (PDB:1VJZ) (rmsd of 0.32 Å and 0.23 Å over 324 Cα atoms for two subunits). Besides, the side chain conformations of residues constituting the Tris-binding site remains unchanged, indicating a rigid binding site. Minor differences in conformation between the two structures are limited to regions that occur at crystal packing interfaces.

Given the Tris-bound crystals of the apo form, it was imperative to identify new crystallization conditions to obtain a cognate sugar-bound complex. To this end, we tried co-crystallization and soaking experiments with several monosaccharides and glucose analogues. Fortunately, a co-crystallization experiment using DNJ resulted in a TmCel5B-DNJ complex. The complex structure in the orthorhombic crystal form was determined at a resolution of 1.80 Å. Unambiguous electron density maps clearly indicate that DNJ displaced the bound Tris at the catalytic site in both subunits, under the same crystallization conditions (Fig. 2a). DNJ was modelled in an undistorted $^4C_1$ conformation. Henceforth, all description of the structure corresponds to that of the TmCel5B-DNJ complex. Little structural variation was observed between the Tris and DNJ bound forms (rmsd 0.18 Å over all Cα atoms).
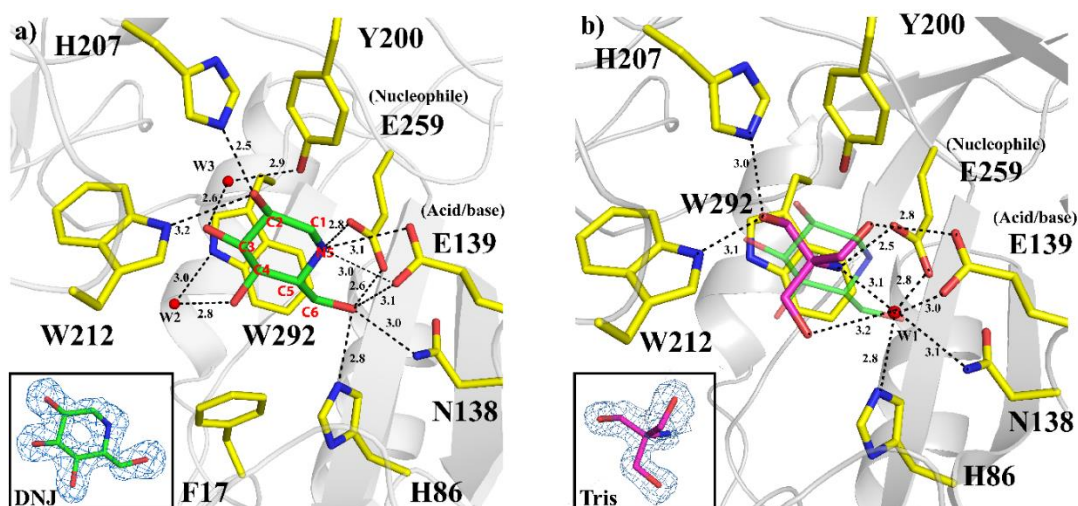
Figure 2. Active site geometry and interactions of bound DNJ and Tris in TmCel5B. a) DNJ, shown in stick representation (green carbon atom), is bound to the expected -1 subsite (PDB:7EC9). The endocyclic N-atom of DNJ makes H-bonds with catalytic residues E259 and E139. DNJ molecule also makes H-bonds with other -1 subsite residues including H86, N138, H207 and W212. Highly conserved W292 and Y200 make water mediated H-bonds with DNJ. Besides this, Y200, W292, W212, H207 and H86 make stacking interactions (cut off ≤ 4 Å) with DNJ. Water molecule (W2 and W3) are present near -2 subsite and are not involved in catalysis. The carbon atoms and endocyclic N-atom of DNJ are marked by red labels. **b**) Tris, shown in stick representation (pink carbon atom), is bound to the -1 subsite (PDB: 7EFZ). DNJ (green stick) from the TmCel5B-DNJ complex is shown superimposed on the Tris complex for comparison. Tris makes H-bonds with both the catalytic residues E139 and E259 as well as with -1 subsite residues namely, H86, N138, H207 and W212. Y200, W292, W212, H207 and H86 also make non-covalent interactions (cut off ≤ 4 Å) with Tris. Side chains of proximal residues (≤ 4 Å) are shown in stick representation (yellow carbon atom). Blue and red color correspond to nitrogen, and oxygen atoms, respectively. The dashed lines correspond to H-bond interactions. All the distances are measured in Å. The inset in each panel shows the corresponding omit map of the bound ligand (Fo-Fc, contoured at 4.0 σ, blue color mesh). The map was generated using POLDER MAPS program in the PHENIX suite.

3.2 Active site residues and interactions of TmCel5B with DNJ

In a scan of the PDB for all complexes of GH homologs with DNJ or its derivatives, a total of 29 PDB entries belonging to 12 different GH families were identified (Table S1). The TmCel5B-DNJ complex is the first instance of a piperidine bound to a GH5 enzyme. The active site of TmCel5B is located at the C-terminal end of the (β/α)$_8$ barrel. DNJ, a 1-deoxyglucose analogue, with a nitrogen replacing the pyranose ring oxygen, is stabilized through extensive hydrogen bonding to residues Glu259, Glu139, Trp212, His207, and Asn138. Moreover, residues Tyr200 and Trp292 form water-mediated hydrogen bonds. Besides, other protein-DNJ interactions include carbohydrate-aromatic interactions, van der Waals interactions and those that contribute to non-polar interactions with residues Tyr200, Trp212, Phe17, His86, and Trp292 (Fig. 2a). Among these, residues His86, Asn138, Glu139, Tyr200, His207, Glu259, and Trp292 are mostly conserved across the GH5 family [31] (Fig. S4, Fig. S5).

A comparison of the TmCel5B-DNJ complex with that of the TmCel5A-cellotetraose complex (PDB: 3AZT) indicates that DNJ is bound at the -1 subsite [35]. Moreover, several residues in contact with the respective ligands are also conserved between TmCel5B and TmCel5A, providing a rationale for the reported endo-glucanase activity of TmCel5B [9]. The presumably protonated endocyclic N5 atom of DNJ is positioned at hydrogen bonding distances of 3.0 Å and 3.1 Å, from the carboxylate oxygen atoms of Glu139 and at 2.8 Å from a carboxylate oxygen atom of the nucleophile Glu259, suggesting strong electrostatic interactions. The C1 carbon atom is positioned at ~3.6 Å and 4.0 Å from the carboxylate oxygen atoms of Glu139. Interestingly, the orientation of DNJ in TmCel5B is flipped relative to the cognate sugar in GH5 homologs so that the 6-OH group mimics the 2-OH group, for example, in TmCel5A (PDB: 3AZT) and CtCelC (PDB: 1CEN). Although this orientation is at odds with the usual binding mode of DNJ mimicking a transition-state analogue, the interactions of the DNJ 6-OH with conserved residues His86 and Asn138 are equivalent to the vital transition-state-stabilizing interactions of substrate 2-OH group of the Michaelis complex. Additionally, the DNJ 2-OH interactions with conserved residues His207 and Trp212 are equivalent to that observed between the substrate 6-OH group with the corresponding residues in TmCel5A and CtCelC. The typical pyranose ring CH-π interaction geometry with the conserved residues Tyr200 and Trp292 (TmCel5B numbering) is also maintained (Fig. 2a). Together, the overall semblance of the interaction geometry to that observed for a competitive inhibitor with transition-state-like features is strongly suggestive of opportunistic binding of DNJ at the -1 subsite, akin to a bound inhibitor or glucose analogue [39].

The deep and extended substrate binding cleft in TmCel5B can likely accommodate at least two glycosyl units on either end of the labile glycosidic bond, a common feature for similar endo-acting enzymes. Notably, the cleft that constitutes the proposed non-reducing end subsites is formed by loop regions L1 (residues 13-26 after strand β1) and L3 (residues 89-106, after strand β3) lining one side, while loop L6 (residues 200-234, after strand β6) lines the cleft containing the reducing end subsites on the opposing side (Fig. 3). Superposition of the DNJ and Tris complexes with the corresponding apo form (PDB: 1VJZ) show that there are minor conformational changes in the residues that make up these subsites (rmsd of 0.23 A among 324 residues), suggesting a rigid and preformed subsite. The only conformational change between the apo/Tris complex and the DNJ complex is that of a rearrangement of the side chain of Lys95 in loop L3 to enclose the active site cleft forming a clamp over DNJ. In this closed-form, Lys95 forms a water-mediated hydrogen bond with the main chain oxygen of Trp212 on loop L6. This movement may facilitate substrate access or product egress from the active site cleft.
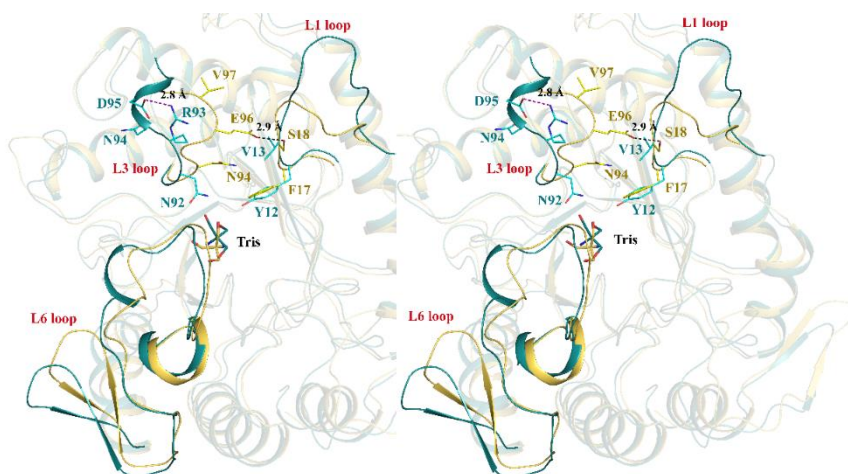
Figure 3. Conformational variability of active site cleft loops within the GH5_36 subfamily. Stereo image for superposition of CpMan5B-Tris complex (teal color) (PDB: 3W0K) and TmCel5B-Tris complex (orange-yellow color) (PDB: 7EFZ). Residues in the L1 and L3 loops are proposed to line the non-reducing end subsites of the cleft, whereas the residues in the L6 loop are expected to line the reducing end subsites. In TmCel5B, a hydrogen bond is present between residue Glu96 in L3 loop and Ser18 of L1 loop, while in CpMan5B, a salt bridge exists between Asp95 and Arg93. These interactions are unique to each homolog and result in significant loop variability between these two close homologs.

Loops L1, L3 and L6 regions are structurally divergent across GH5 subfamilies and play important roles in determining the substrate specificity across paralogs [40]. A *cis*-peptide bond between residues Ile201-Pro202 marks the C-terminal end of the strand β6 and the beginning of the L6 region and likely directly affects the conformation of the substrate-binding cleft at the bottom. Interestingly, Pro202 is conserved only within the GH5_25, GH5_36 and GH5_37 subfamilies (Fig. S4).

In the numerous endoglucanases of the GH5 family, a strictly conserved His residue at position 198 (TmCel5B numbering) has been proposed to be critical for catalysis. This residue positioned within 3.5 Å of the two catalytic glutamates was proposed to constitute a Glu-His-Glu electron-relay network that affects catalysis [20,35]. In the GH5_36, this residue is replaced by a stringently conserved arginine (Arg198) (Fig. S5) [8]. Comparison of the GH5_36 TmCel5B, CpMan5B, and the GH5_37 TmCel5A structures indicate that the Arg and His side chains are structurally equivalent. The side chain of Arg198 maintains a network of interactions with neighbouring residues that is conserved within the GH5_36. This includes strong ionic interactions with the two catalytic glutamates and Asp176 (Fig. S6). Additional hydrogen bonds occur with the side chain of conserved Asn138 and carbonyl oxygen of Cys197. Mutation of Arg to Ala in CpMan5B abolished activity completely, whereas substitution with His lowered activity by about 85 and 93% , for mannanase and endoglucanase activities, respectively [8]. Considering that the pKa values of His and Arg are widely different,

it is likely that this network of ionic interactions may modulate catalysis by influencing the protonation/deprotonation states of the catalytic residues [42]. We believe that this novel variant motif may be an evolutionary adaptation in the GH5_36 subfamily in comparison to its phylogenetically close subfamilies.

3.3 Comparison of active site features within the GH5_36 subfamily

CpMan5B is the sole GH5_36 homolog where mutational analyses of the predicted active site residues combined with activity assays have been carried out [8]. It is to be noted that TmCel5B is an endo-1,3-1,4-β-D-glucanase, whereas CpMan5B is an endo-1,4-β-D-glucanase [8,9]. Hence, we carried out a comparative analysis of TmCel5B-DNJ with CpMan5B to compare and contrast the active site features. Superposition shows that despite high similarity, there are notable differences in the active site residues, which in part may influence the relative mannanase:endoglucanase activities and the specificity profiles between these two close homologs. The two enzymes share several residues that form the proposed -1 and -2 subsites, including His86, Asn94, Asn138, Tyr200, Trp212, and Trp292 (Fig. S4). Substitution of the equivalent His86 and Asn138 to alanine in CpMan5B showed total loss of both activities. These two residues are completely conserved across the GH5_25, GH5_36, and GH5_37 subfamilies as well and have been previously implicated in catalysis by other GH5 endo-glucanases. Tyr200, which forms a stacking interaction with the -1 subsite sugar, is strictly conserved across multiple GH5 subfamilies that include groups of β-mannan-specific, endo-glucan-specific and, dual-specificity enzymes [42]. However, Trp212 in loop L6 is conserved only between GH5_36, GH5_37, and GH5_25 subfamilies (Fig. S5). A comparison of TmCel5B with the cellobiose complexes of TmCel5A (PDB: 3AMG) and CtCelC (PDB: 1CEN) shows that Trp212 forms stacking interactions with the -2 subsite sugar and therefore is expected to play a similar role in the GH5_36 homologs as well.

Phe17 in loop L1 contributes to a strong van der Waals interaction (3.8 Å) with the O4 atom of DNJ. In GH5_36, this residue is either Phe or Tyr, whereas the entire loop region containing this residue is divergent across the GH5 family (Fig. S5). In CpMan5B, the equivalent mutants Y12A, Y12E, and Y12F retained only 28, 12, and 63 % activity, respectively, with mannohexaose, while with cellohexaose, they retained 67, 20, and 29% activity, respectively [8]. Positioning a Tyr instead of Phe in the TmCel5B-DNJ complex indicates a strong hydrogen bond interaction is possible (2.8 Å) between the Tyr OH and the O4 hydroxyl of DNJ (Fig. 2). Considering the protein-DNJ interactions and significant effects of the mutations on the activity profiles, the residue at this position may play an important role

in determining the relative activity profiles and substrate specificities/preferences between homologs within the GH5_36 subfamily. We also speculate that the interaction of DNJ with Phe17 may be responsible for the unusual flipped-over state of the iminosugar at the -1 subsite since the overall geometry of the protein-DNJ interactions is essentially identical when compared to cognate sugar-bound complexes across GH5 homologs.

Notably, Ala substitution of Asn at position 94 in CpMan5B is strongly deleterious for both activities, suggesting a significant role for this residue. While this residue is conserved within the GH5_36 subfamily, this region of loop L3 is divergent compared to subfamilies GH5_25 and GH5_37. Moreover, the residue composition of L3 is distinct between the two GH5_36 subfamily homologs (Fig. S4). Superposition of the CpMan5B and TmCel5B show that L3 residues 93-98 adopt significantly different conformations (rmsd 2.4 Å over the Cα atoms). For instance, the equivalent Cα atoms of residues 96 and 97 are separated by ~8.2 and 6.2 Å, respectively (Fig. 3). Thus, in CpMan5B, the side chain of Asn94 makes a hydrogen bond with the bound Tris, whereas the equivalent Asn in TmCel5B is oriented away from the bound DNJ with no interaction between the two. To evaluate the possible role of Asn94 in TmCel5B, the TmCel5A-cellobiose complex (PDB: 3AMG) was superimposed to examine possible sugar-protein interactions. In this model, Asn94 side chain in TmCel5B is positioned within hydrogen bonding distance to the expected -2 subsite sugar, providing an explanation for the deleterious effect of the Asn94 mutant. A close examination of the L3 regions in CpMan5B and TmCel5B shows that the large conformational difference of L3 between the two proteins is tightly coupled to the conformation and interactions with residues of the adjacent L1 loop. The L1 loop in CpMan5B is three residues longer, with a divergent residue composition. These differences in the conformations and interactions between the L1 and L3 loops in the two proteins, result in dissimilar geometries of their non-reducing active site clefts. Furthermore, notable differences exist in the conformations of the corresponding L6 loops connecting β6 and α6. The rmsd for the Cα atoms of equivalent regions (residue 200-233 in TmCel5B) is ~3.0 Å (Fig. 3). Together, it appears that sequence and structural modifications occur between the GH5_36 subfamily homologs at both the non-reducing and reducing subsites and that these are restricted to loops L1, L3 and L6 that line the extended active site cleft. We propose that these distinct features play a crucial role in determining the structural basis of the contrasting homolog-specific substrate profiles within this subfamily.

3.4 Structural basis of iminosugar binding and inhibition in the GH5 family

Competitive inhibition of glycosidases by iminosugars with a protonated ring nitrogen is attributed to their resemblance to the partial positive charged states of the anomeric carbon or endocyclic oxygen of the glycosidase transition state [11-13]. An exhaustive search of the CAZy database revealed 29 crystal structures of GH complexes of DNJ and their derivatives (Table S1). The first DNJ complex for the GH5 family is reported here in our study. In addition, we identified 7 iminocyclitols complexes of four GH5 homologs (Table S2). Of these, six compounds have been shown to act as inhibitors for the respective GH5 enzymes over a large range of potencies (Ki = 0.03 – 400 µM). These inhibitors constitute a diverse set of iminocyclitols that vary in their ring sizes, stereochemistry and the type and size of the substituents (Fig. S7). For instance, DNJ, CTS, IFL and MNI are monosaccharides while 9MR, CDN and IDC are disaccharides. Moreover, the enzymes in this dataset belong to 5 different GH5 subfamilies and are divergent in their biochemical activities, with distinct substrate binding clefts built around a conserved catalytic site geometry (Table S2). A comparative analysis of the binding modes and interactions of these enzyme-ligand complexes, including the TmCel5B-DNJ, was carried out to identify and assess shared interactions that are most relevant to iminosugar binding and inhibition of the GH5 enzymes.

The iminocyclitol moieties in all cases occupy the -1 subsites, making multiple non-bonded interactions with the neighbouring residues. Among these, CH-π stacking interactions with aromatic residues Tyr200 and Trp292 (TmCel5B numbering) appear invariant (Fig. 4, Fig. S8). Interestingly, in the endoglycoceramidase complex (RhEGC, PDB: 2OYM) with a 5-membered 2,5-imino-D-mannitol ring (MNI), the sugar is present in the same flipped ring orientation as in TmCel5B-DNJ. Moreover, the iminosugar moieties in the RhEGC and TmCel5B complexes superpose well and the interactions of the iminosugar with the conserved -1 subsite residues (Glu139, Glu259, His86, Asn94, Asn138, Tyr200, and Trp292) are almost identical between the two complexes (Fig. 4e). MNI is an inhibitor of RhEGC with a Ki value of 10 µM. However, it is a weaker inhibitor compared to cellobiose like isofagomine (9MR, Ki = 5 µM) and cellobiose like imidazole, (IDC, Ki =0.5 µM), for RhEGC (Table S2). It is probable that an iminosugar bound in this flipped orientation could also influence enzymatic activity of TmCel5B.

Figure 4. Comparative analysis of iminocyclitol interactions with GH5 homologs. Panels **a-g** show pairwise superpositions of the TmCel5B-DNJ complex (PDB: 7EC9, light brown) superposed on structures of seven separate GH5 complexes (dark teal) with bound iminocyclitols. (**a**) CmMan5 (PDB: 1UZ4); (**b, c**) BaCel5A (PDB: 1OCQ, 2V38); (**d, e, f**) RhEGC (PDB: 2OYK, 2OYM, 2OYL); (**g**) CaExg (PDB: 1EQC). All iminosugars occupy the -1 subsite in their respective structures. TmCel5B residues, namely, His86, Asn138, Glu139, Tyr200, Glu259 and Trp292 are conserved across the GH5 family and make interactions with the iminosugar. The endocyclic/anomeric N-atoms make electrostatic interactions with the catalytic glutamates. Ligands are shown as sticks (DNJ in yellow, iminocyclitols in homologs in cyan). IFL- Isofagomine; MNI- Five membered ring iminocyclitol; CTS-Castanospermine; DNJ- 1-deoxynojirimycin; IDC- Imidazole derived cellobiose; 9MR- cellobiose derived IFM, CDN- cellobiose derived neuromycin (Table S2).

The crucial role of the sugar 2-OH group-enzyme interactions in contributing to transition-state stabilization in retaining glycosidases is well known [35]. In most complexes here, favourable interactions are present between an equivalently positioned donor/acceptor group in the ligand and the amide side chain of the stringently conserved Asn that precedes the catalytic acid/base residue and/or the side chain of a mostly conserved His residue (His86 in TmCel5B). Next, the structures show that irrespective of the location of the nitrogen atom in place of the endocyclic oxygen or at the anomeric carbon position, it is always positioned within H-bonding distance from either or both catalytic glutamates. For instance, DNJ has an N atom at the O5 position, while in isofagomine-lactam (IFL), the N atom is at the C1 position. This tight electrostatic interaction is observed for both *exo-* and *endo*-acting enzymes and is independent of the individual conformations (distorted or undistorted) and flipped orientations that the furanosidic or pyranosidic groups adopt within the active sites (Table S2). Thus, the iminosugar positioning, in general, is consistent with the favourable electrostatic interactions between the catalytic glutamates and the developing positive charge at the transition state of retaining glycosidases. Indeed, the protonated form of bound isofagomine is observed in the active site of BsCel5A [43]. The binding mode of imidazole derived cellobiose (IDC) in RhEGC is a minor exception since it is a case of an imidazole-enforced transition-state mimic, and additional interactions with the 2-OH group determine the binding mode of this high affinity inhibitor [44] (Fig. 4, Fig. S8).

Together, the comparative analysis demonstrates the structural basis for binding of an iminosugar in the GH5 family. Three shared structural interactions with the first-shell conserved protein residues around the -1 subsite profoundly influence iminosugar binding and inhibition and appear to extend to all members of this large and diverse family. These include, i) the stacking with a conserved aromatic residue at the -1 subsite, ii) the H-bonding interactions of the sugar OH or an equivalent group with either or both of a stringently conserved Asn and the mostly conserved neighbouring His, and iii) the interaction of the ring nitrogen with the catalytic carboxylates. These complexes also indicate that the plasticity seen in the -2 subsites of GH5 enzymes should also be considered for design of specific and relevant glycosidase inhibitors. For instance, in AaCel9A (PDB: 3RX7), isofagomine by itself is a weak inhibitor, but the addition of a single β-glucosyl moiety at the 4-OH position significantly improved the binding to the enzyme [45].

4. Conclusions

TmCel5B is the first structure of a GH5_36 subfamily homolog bound with a cognate sugar ligand. The identified divergent and conserved structural features in the extended active site cleft that result in the differences in the active site shapes provide a rationale for the diversity in substrate preference profiles within this dual-specific subfamily and across closely related GH5 subfamilies. Future work on obtaining structures of cognate sugar ligands is required to explore the determinants of the substrate specificity in TmCel5B. GH5 enzymes are ubiquitously present in pathogens like *Mycobacterium tuberculosis* or *Yersinia pestis*, and in non-pathogenic microbes. Moreover, numerous studies have shown that iminocylitols can act as inhibitors for various glycosidase enzymes. The structure basis of iminocyclitol binding to a GH5 homologs described in our studies can provide pointers for the design of potent and specific inhibitors for glycosidase enzymes of medical interests.

Data availability: The coordinates and structure factors presented in this manuscript have been deposited in the Protein Data Bank (PDB) with the following codes: 7EFZ, 7EC9. All remaining data are contained within the article.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

REFERENCES

[1]   G.J. Davies, T.M. Gloster, B. Henrissat, Recent structural insights into the expanding world of carbohydrate-active enzymes., Curr. Opin. Struct. Biol. 15 (2005) 637–645.

[2]   V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013, Nucleic Acids Res. 42 (2014) D490–D495.

[3]   B. Henrissat, G. Davies, Structural and sequence-based classification of glycoside

hydrolases., Curr. Opin. Struct. Biol. 7 (1997) 637–644.

[4]   Y. Kumagai, K. Yamashita, T. Tagami, M. Uraji, K. Wan, M. Okuyama, M. Yao, A. Kimura, T. Hatanaka, The loop structure of Actinomycete glycoside hydrolase family 5 mannanases governs substrate recognition., FEBS J. 282 (2015) 4001–4014.

[5]   P.-H.H. Liang, W.-L.L. Lin, H.-Y.Y. Hsieh, T.-Y.Y. Lin, C.-H.H. Chen, S.K. Tewary, H.-L. Lee, S.-F. Yuan, B. Yang, J.-Y.Y. Yao, M.-C.C. Ho, A flexible loop for mannan recognition and activity enhancement in a bifunctional glycoside hydrolase family 5., Biochim. Biophys. Acta. Gen. Subj. 1862 (2018) 513–521.

[6]   H. Aspeborg, P.M. Coutinho, Y. Wang, H. Brumer, B. Henrissat, Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5), BMC Evol. Biol. 12 (2012) 1–16.

[7]   F.A. Fusco, R. Ronca, G. Fiorentino, E. Pedone, P. Contursi, S. Bartolucci, D. Limauro, Biochemical characterization of a thermostable endomannanase /endoglucanase from Dictyoglomus turgidum, Extremophiles. 22 (2018) 131–140.

[8]   T. Oyama, G.E. Schmitz, D. Dodd, Y. Han, A. Burnett, N. Nagasawa, R.I. Mackie, H. Nakamura, K. Morikawa, I. Cann, Mutational and structural analyses of Caldanaerobius polysaccharolyticus Man5B reveal novel active site residues for family 5 glycoside hydrolases, PLoS One. 8 (2013) 1–11.

[9]   M.A.S. Khan, M. Akbar, M. Kitaoka, K. Hayashi, A unique thermostable lichenase from Thermotoga maritima MSB8 with divergent substrate specificity, Indian J. Biotechnol. 6 (2007) 315–320.

[10]   S.R. Chhabra, K.R. Shockley, D.E. Ward, R.M. Kelly, Regulation of endo-acting glycosyl hydrolases in the hyperthermophilic bacterium Thermotoga maritima grown on glucan- and mannan-based polysaccharides, Appl. Environ. Microbiol. 68 (2002) 545–554.

[11]   M. Bols, 1-Aza sugars, apparent transition state analogues of equatorial glycoside formation/cleavage, Acc. Chem. Res. 31 (1998) 1–8.

[12]   T.M. Gloster, G.J. Davies, Glycosidase inhibition: assessing mimicry of the transition state., Org. Biomol. Chem. 8 (2010) 305–320.

[13]   M.I. Simone, L. Mares, C. Eveleens, A. McCluskey, B.B. Pappin, M.J. Kiefel, T.A. Houston, Back to (non-)Basics: An update on neutral and charge-balanced glycosidase inhibitors, Mini-Reviews Med. Chem. 18 (2017) 812–827.

[14]   Y. Ma, W. Lv, Y. Gu, S. Yu, 1-Deoxynojirimycin in mulberry (Morus indica L.) leaves ameliorates stable Angina pectoris in patients with coronary heart disease by improving

antioxidant and anti-inflammatory capacities, Front. Pharmacol. 10 (2019) 1–9.

[15] G. Horne, F.X. Wilson, J. Tinsley, D.H. Williams, R. Storer, Iminosugars past, present and future: Medicines for tomorrow, Drug Discov. Today. 16 (2011) 107–118.

[16] B.E. Tyrrell, A.C. Sayce, K.L. Warfield, J.L. Miller, N. Zitzmann, Iminosugars: Promising therapeutics for influenza infection, Crit. Rev. Microbiol. 43 (2017) 521–545.

[17] R.J. Nash, A. Kato, C.Y. Yu, G.W. Fleet, Iminosugars as therapeutic agents: Recent advances and promising trends, Future Med. Chem. 3 (2011) 1513–1521.

[18] M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R.M. Keegan, E.B. Krissinel, A.G.W. Leslie, A. McCoy, S.J. McNicholas, G.N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin, K.S. Wilson, Overview of the CCP4 suite and current developments., Acta Crystallogr. D. Biol. Crystallogr. 67 (2011) 235–242.

[19] T.G.G. Battye, L. Kontogiannis, O. Johnson, H.R. Powell, A.G.W. Leslie, iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM., Acta Crystallogr. D. Biol. Crystallogr. 67 (2011) 271–281.

[20] W. Kabsch, Automatic indexing of rotation diffraction patterns, J. Appl. Crystallogr. 21 (1988) 67–72.

[21] P.D. Adams, P. V Afonine, G. Bunkóczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.-W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution., Acta Crystallogr. D. Biol. Crystallogr. 66 (2010) 213–221.

[22] P. V Afonine, R.W. Grosse-Kunstleve, N. Echols, J.J. Headd, N.W. Moriarty, M. Mustyakimov, T.C. Terwilliger, A. Urzhumtsev, P.H. Zwart, P.D. Adams, Towards automated crystallographic structure refinement with phenix.refine., Acta Crystallogr. D. Biol. Crystallogr. 68 (2012) 352–367.

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

[24] V.B. Chen, W.B. Arendall III, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, Acta Crystallogr. Sect. D. 66 (2010) 12–21.

[25] L. Holm, DALI and the persistence of protein shape, Protein Sci. 29 (2020) 128–140.

[26] K. Katoh, J. Rozewicki, K.D. Yamada, MAFFT online service: multiple sequence

alignment, interactive sequence choice and visualization., Brief. Bioinform. 20 (2019) 1160–1166.

[27] X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server, Nucleic Acids Res. 42 (2014) W320–W324.

[28] G.E. Crooks, G. Hon, J.-M. Chandonia, S.E. Brenner, WebLogo: a sequence logo generator., Genome Res. 14 (2004) 1188–1190.

[29] D.W. Schrödinger L, PyMOL, Available at: http://www.Pymol.Org/Pymol. (2020).

[30] R.A. Laskowski, M.B. Swindells, LigPlot+: multiple ligand-protein interaction diagrams for drug discovery., J. Chem. Inf. Model. 51 (2011) 2778–2786.

[31] R. Domínguez, H. Souchon, M.-B. Lascombe, P.M. Alzari, The crystal structure of a Family 5 endoglucanase mutant in complexed and uncomplexed forms reveals an induced fit activation mechanism, J. Mol. Biol. 257 (1996) 1042–1051.

[32] J. Sakon, W.S. Adney, M.E. Himmel, S.R. Thomas, P.A. Karplus, Crystal structure of thermostable Family 5 endocellulase E1 from Acidothermus cellulolyticus in Complex with Cellotetraose, Biochemistry. 35 (1996) 10648–10660.

[33] S.M. Cutfield, G.J. Davies, G. Murshudov, B.F. Anderson, P.C.E. Moody, P.A. Sullivan, J.F. Cutfield, The structure of the exo-β-(1,3)-glucanase from Candida albicans in native and bound forms: Relationship between a pocket and groove in family 5 glycosyl hydrolases, J. Mol. Biol. 294 (1999) 771–783.

[34] A.M. Larsson, L. Anderson, B. Xu, I.G. Muñoz, I. Usón, J.-C. Janson, H. Stålbrand, J. Ståhlberg, Three-dimensional crystal structure and enzymic characterization of β-mannanase Man5A from Blue Mussel Mytilus edulis, J. Mol. Biol. 357 (2006) 1500–1510.

[35] T.-H. Wu, C.-H. Huang, T.-P. Ko, H.-L. Lai, Y. Ma, C.-C. Chen, Y.-S. Cheng, J.-R. Liu, R.-T. Guo, Diverse substrate recognition mechanism revealed by Thermotoga maritima Cel5A structures in complex with cellotetraose, cellobiose and mannotriose., Biochim. Biophys. Acta. 1814 (2011) 1832–1840.

[36] A.M. Brzozowski, D.M. Lawson, J.P. Turkenburg, H. Bisgaard-Frantzen, A. Svendsen, T. V. Borchert, Z. Dauter, K.S. Wilson, G.J. Davies, Structural analysis of a chimeric bacterial α-Amylase. High-resolution analysis of native and ligand complexes, Biochemistry. 39 (2000) 9099–9107.

[37] N.N. Mhlongo, A.A. Skelton, G. Kruger, M.E.S. Soliman, I.H. Williams, A critical survey of average distances between catalytic carboxyl groups in glycoside hydrolases., Proteins. 82 (2014) 1747–1755.

[38] J.D. McCarter, G. Stephen Withers, Mechanisms of enzymatic glycoside hydrolysis, Curr. Opin. Struct. Biol. 4 (1994) 885–892.

[39] C.S. Rye, S.G. Withers, Glycosidase mechanisms, Curr. Opin. Chem. Biol. 4 (2000) 573–580.

[40] E.M. Glasgow, K.A. Vander Meulen, T.E. Takasuka, C.M. Bianchetti, L.F. Bergeman, S. Deutsch, B.G. Fox, Extent and origins of functional diversity in a subfamily of glycoside hydrolases., J. Mol. Biol. 431 (2019) 1217–1233.

[41] B. Zheng, W. Yang, X. Zhao, Y. Wang, Z. Lou, Z. Rao, Y. Feng, Crystal structure of hyperthermophilic endo-β-1,4-glucanase: Implications for catalytic mechanism and thermostability, J. Biol. Chem. 287 (2012) 8336–8346.

[42] Z. Chen, G.D. Friedland, J.H. Pereira, S.A. Reveco, R. Chan, J.I. Park, M.P. Thelen, P.D. Adams, A.P. Arkin, J.D. Keasling, H.W. Blanch, B.A. Simmons, K.L. Sale, D. Chivian, S.R. Chhabra, Tracing determinants of dual substrate specificity in glycoside hydrolase family 5, J. Biol. Chem. 287 (2012) 25335–25343.

[43] A. Varrot, C.A. Tarling, J.M. Macdonald, R. V. Stick, D.L. Zechel, S.G. Withers, G.J. Davies, Direct observation of the protonation state of an imino sugar glycosidase inhibitor upon binding., J. Am. Chem. Soc. 125 (2003) 7496–7497.

[44] M.E.C.C. Caines, S.M. Hancock, C.A. Tarling, T.M. Wrodnigg, R. V. Stick, A.E. Stütz, A. Vasella, S.G. Withers, N.C.J.J. Strynadka, The structural basis of glycosidase inhibition by five-membered iminocyclitols: The clan a glycoside hydrolase endoglycoceramidase as a model system, Angew. Chemie - Int. Ed. 46 (2007) 4474–4476.

[45] S. Moréra, A. Vigouroux, K.A. Stubbs, A potential fortuitous binding of inhibitors of an inverting family GH9 β-glycosidase derived from isofagomine, Org. Biomol. Chem. 9 (2011) 5945–5947.

Supplementary Information

Structure of an iminosugar complex of a glycoside hydrolase family 5 lichenase provides insights into the active site

Puneet Garg and Narayanan Manoj*

Department of Biotechnology

Bhupat and Jyoti Mehta School of Biosciences

Indian Institute of Technology Madras

Chennai – 600036, India

* Corresponding author: Narayanan Manoj

Email: nmanoj@iitm.ac.in

Figure S1. a) SDS-PAGE analysis of TmCel5B. Lane M shows various molecular mass markers and lane 1 shows the band for the purified enzyme corresponding to the molecular weight ~39 kDa. **b) Size-exclusion chromatogram of TmCel5B**. SEC analysis indicates that the protein exists as a monomer in solution. The chromatogram was monitored at an absorbance of 280 nm. The molecular weight of the sample was estimated by plotting the calibration curve using ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (44 kDa) and ribonuclease (13.7 kDa) as standard proteins.

Figure S2. Topology diagram of TmCel5B. α-helices are displayed as cylinders and β-strands are displayed as arrows. Residue numbers marking the extent of the secondary structure elements are indicated. TmCel5B shows an incomplete $(\beta/\alpha)_8$ TIM barrel architecture displayed by yellow arrows (strands β1-β8) and green cylinders (α-helices α1-α8), where α5 helix is replaced by an irregular loop (labelled A5, red rectangular box). The position of catalytic residues (Glu139 and Glu259) and other important residues are indicated by filled purple circles. Besides the core structure, TmCel5B also displays other minor α-helices [α'1- α'5] and 4 β-strands (β'1-β'4) shown in blue. β'1-β'2 and β'3-β'4 form two pairs of antiparallel β-sheets. Loop regions that constitute the substrate binding cleft are labelled L1-L8.

Figure S3. Active site loop variations within closely related GH5 subfamilies. Representation of variation in the conformations of corresponding loop regions (L1-L8) in the GH5_25 (PDB: 3AZT) (green), GH5_37 (PDB: 1CEN) (yellow), GH5_36 (PDB: 3W0K) (blue) and GH5_36 (PDB: 7EFZ) (red) subfamilies.

β1

```
                   1        10          20                       30          β2      α'1
                                                                          40      50
TmCel5B   1 MNNTIPRWRGFNL.L.EAFS.IK.....STG...N.FKEEDFLWMAQWDFNFVRIPMCHLLWSDRGNPFI  58
CpMan5B   1 .....MKKYGFNF.Q.WMYV.WE..EGREPE...P.PDKKALDFLAETGFNFVRIPVDYRFWTRNFDYFN  56
CtCelC    1 ...MVSFKAGINL.GGWISQ.YQVFSKEHFD...TFITEKDIETIAEAGFDHVRLPFDYPIIESDDNVGE  62
TmCel5A   6 FERNKILGRGINI.GNALEA.PN.....EGDWGVV.IKDEFFDIIKEAGFSHVRIPIRWSTHAYAFPPYK  67
FpCel5   51 LEYNKIIGKGVNI.GNALA.PF.....EGAWGVR.IEDEYFEIIKKRGFDSVRIPIRWSAHISEKPPYD  112
FnCel5A  37 FEYNKMIGHGINM.GNALEA.PV.....EGSWGVY.IEDEYFKIIKERGFDSVRIPIRWSAHISEKYPYE  98
Ccsz     37 .....KMKRGINI.GNALESPKD.....FP.WDVK.MSNKFFDDIKDAGFDTVRIPVRFSDYTSD.DNFK  93
CtCel5E 330 FEMVRKMGMGTNLGN.TLEA.PY.....EGSWSKS.AMEYFDDFKAAGYKNVRIPVRWDNHTMRTYPYT  391
                              L1
```

```
          α'2        α2              β3              90          100        α3
           60        70        80                                        110       120
TmCel5B  59 IREDFFEKIDRVIFWGEKYGIHICISLHRAPGYSVNKEVEEKTNLWKDETAQEAFIHHWSFIARRYKGIS 128
CpMan5B  57 PDKKVFEYIDLYLRECSARNIHMCLNLHRAPGYCINRNDIERDNLWLDKRAQDGFVYQWELFAKRYKGVS 126
CtCelC   63 YKEDGLSYIDRCLEWCKKYNLGLVLDMHHAPGYRFQDFK..TSTLFEDPNQQKRFVDIWRFLAKRYINER 130
TmCel5A  68 IMDRFFKRVDEVINGALKRGLAVVINIHHYEEL.......MNDPEE..HKERFLALWKQIADRYKDYP 126
FpCel5  113 IDRNFLERVKHVVDRALENNLTVIINTHHFEEL.......YQEPDK..YGNVLVEIWRQIAKFFKDYP 171
FnCel5A  99 IDKFFLDRVKHVVDVALKNDLVVINCHHFEEL.......YQAPDK..YGPVLVEIWRQVAQAFKDYP 156
Ccsz     94 IDEDFFKKIDKYVDYALDKDLIVVLDLHBFEEI.......MKEPRV..HKEKFLKIWQQIANRYQKYD 152
CtCel5E 392 IDKAFLDRVEQVVDWSLSRGFVTIINSHRDDWI.......KEDYNG..NIERFEKIWEQIAERFKNKS 450
                                         L3
```

β4   **α4**   **β5**   **α'3**   **A5**

```
          130        140        150        160        170        180
TmCel5B 129 STHLSFNLINEPPFPDPQIMSVEDHNSLIKRTITEIRKIDPERLIIIDGLGYG...N...I...PV.D.DLT. 189
CpMan5B 127 SKFLSFDLVNEPPNIGQYGLTRENHASLIIRTVEAIRKIDPDREIVIDGLGGG...N...I...AM.P.ELA. 187
CtCelC  131 .EHIAFELLNEVVEP.....DSTRWNKLMLEYIKAIREIDSTMWLYIGGNNYN..SPDE..LK.N.LAD. 187
TmCel5A 127 E.TLFFSILNEPHG....NLTPSEKWNELLEAKVRSIDKKHTIIIGTAEWG..G...ISALE.K.LSVP 185
FpCel5  172 E.NLFFEIYNEPAQ....NLTAEKWNELYPKVLKVIRESNPTRIVIIDAPNWAHYS..A..VRSL.KLV. 230
FnCel5A 157 D.KLFFEIFNEPAQ....NLTPTKWNELYPKVLGEIRKTNPSRIVIIDVPNWSNYS..Y..VREL.KLV. 216
Ccsz    153 K.KLVFSLLNEPKE....NLYSQLLNEYIEEAIKIIRKTNPKRTIIVGPYNFY..Q..IDYLN.E.LNI. 210
CtCel5E 451 E.NLLFEIMNEPFG....NITDEQIDDMNSRILKIIRKTNPTRIVIIGGGYWN..SYNT..LV.NIKIP. 509
                ★ ●      L4
```

β6   **α'4**

```
          190        200        210        220                  β'1              β'2
                                                                                 230
TmCel5B 190 ..I..ENTVQSCRGYIPFSVTHYKAEWV.D.SKDF.PVPEWPNGWHF.................GEYWN 234
CpMan5B 188 ..H..LGVVHSGRGYQPMALTHYQASWW.DGHKGL.PEPYYPDLLWQ.................GKVWN 233
CtCelC  188 ..IDDDYIVYNFHFYNPFFFTHQKAHWS.ESAMAYNRTVKYPGQYEGIEEFVKNNPKYSFMMELNNLKLN 254
TmCel5A 186 KWE..KNSIVTIHYYNPFEFTHQGAEWV.EGSEKW.LGRKWGS.................PDD 227
FpCel5  231 .ND..KRIIVSFHYYEPFNFTHQGAEWV.NPTPPV.GVKWNGE.................EWE 271
FnCel5A 217 .DD..KNIIVSFHYYEPFNFTHQGAEWV.SPTLPI.GVKWEGA.................DWE 257
Ccsz    211 ..PKDSNIVVSFHYYEPNDFAFQGNIYHKG.FEHL.SNITWEG.................TNEQ 253
CtCel5E 510 ..D.DPYLIGTAHYYDPFEFTHQGAEWV.EGSEKW.LGRKWGT.................OED 539
               ★                                                    L6
```
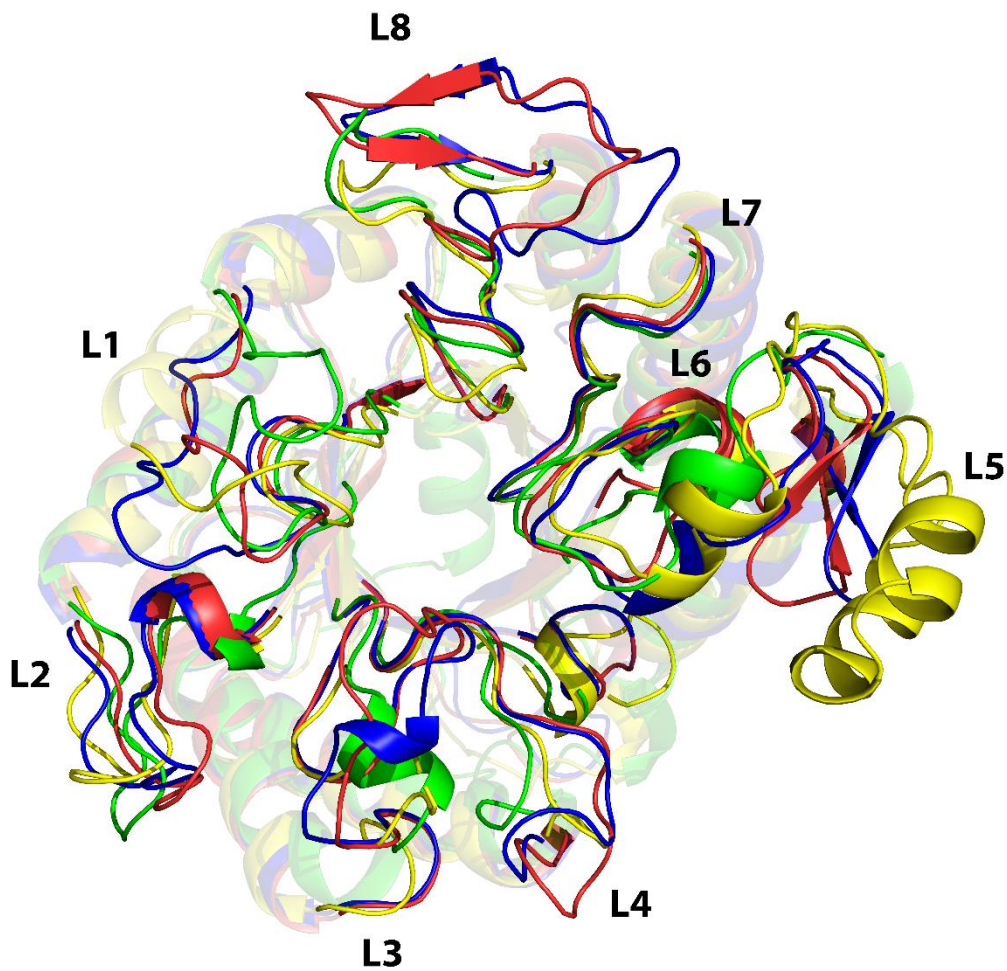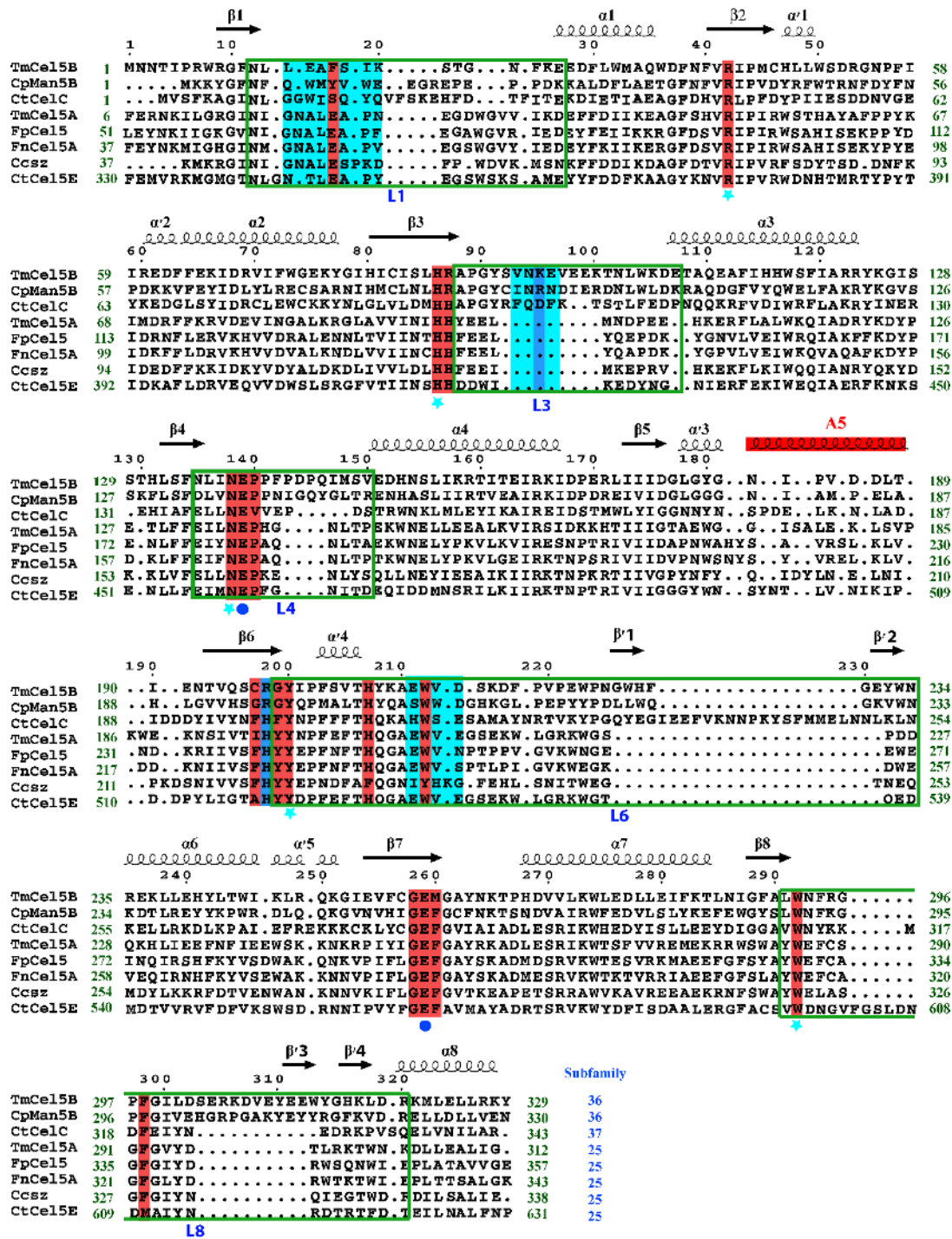
α6   **α'5**   **β7**   **α7**   **β8**

```
          240        250        260        270        280        290
TmCel5B 235 REKLLEHYLTWI.KLR.QKGIEVFCGEMGAYNKTPHDVVLKWLEDLLEIFKTLNIGFALWNFRG...... 296
CpMan5B 234 KDTLREYYKPWR.DLQ.QKGVNVHIGEFGCFNKTSNDVAIRWFEDVLSLYKEFEWGYSLWNFRG...... 295
CtCelC  255 KELLRKDLKPAI.EFREKKKCKLYCGEFGVIAIADLESRIKWHEDYISLLEEYDIGGAVWNYKK.....M 317
TmCel5A 228 QKHLIEEFNFIEEWSK.KNKRPIYIGEFGAYRKADLESRIKWTSFVVREMEKRRWSWAYWEFCS...... 290
FpCel5  272 INQIRSHFKYVSDWAK.QNKVPIFLGEFGAYSKADMDSRVKWTESVRKMAEEFGFSYAYWEFCA...... 334
FnCel5A 258 VEQIRNHFKYVSEWAK.KNNVPIFLGEFGAYSKADMESRVKWTKTVRRIAEEFGFSLAYWEFCA...... 320
Ccsz    254 MDYLKKRFDTVENWAN.KNNVKIFLGEFGVTKEAPETSRRAWVKAVREEAEKRNFSWAYWELAS...... 326
CtCel5E 540 MDTVVRVFDFVKSWSD.RNNIPVYFGEFAVMAYADRTSRVKWYDFISDAALERGFACSVWDNGVFGSLDN 608
                                   ●                                     ★
```

β'3 **β'4**   **α8**

```
          300        310        320              Subfamily
TmCel5B 297 PFGILDSERKDVEYEEWYGHKLD.RKMLELLRKY 329    36
CpMan5B 296 PFGIVEHGRPGAKYEYYRGFKVD.RELLDLLVEN 330    36
CtCelC  318 DFEIYN.........EDRKPVSQELVNILAR.    343    37
TmCel5A 291 GFGVYD.........TLRKTWN.KDLLEALIG.   312    25
FpCel5  335 GFGIYD.........RWSQNWI.EPLATAVVGE   357    25
FnCel5A 321 GFGLYD.........RWTKTWI.EPLTTSALGK   343    25
Ccsz    327 GFGIYN.........QIEGTWD.FDILSALIE.   338    25
CtCel5E 609 DMAIYN.........RDTRTFD.JEILNALFNP   631    25
               L8
```

Figure S4. Structure-based multiple sequence alignment of TmCel5B with representative GH5 homologs. The alignment is based on a multiple structural superposition of available structures from the GH5_37, GH5_36 and GH5_25 subfamilies. The secondary structure elements marked on the top line correspond to that of TmCel5B (PDB: 7EFZ). The catalytic residues are marked with blue circles and conserved motifs and critical residues are highlighted in red and marked with cyan stars. Critical residues, Lys95 and Arg198, unique to the GH5_36 subfamily, are highlighted in blue. The loop regions L3, L4, L6 and L8 are marked using green boxes. The residues within loops L1, L3 and L6 which line the active site cleft are highlighted in cyan. TmCel5B and CpMan5B belongs to GH5_36 subfamily and CtCelC belongs to GH5_37 subfamily. Others belong to GH5_25 subfamily.
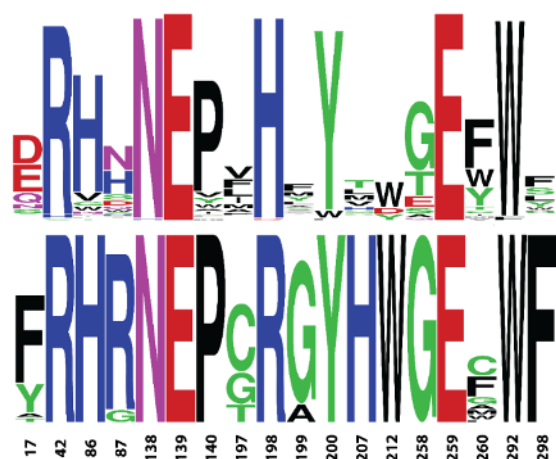
Figure S5. WebLogo representation of conservation pattern of active site cleft residues within the GH5_36 subfamily and comparison with homologs from other GH5 subfamilies. The top panel displays the conservation of specific residues present across the GH5 family (n=122 sequences), while the bottom panel displays the structurally equivalent residues within the GH5_36 subfamily (n=28 sequences). The residue number corresponds to the amino acid number in TmCel5B. Residues R42, N138, E139, Y200, E259 and W292 are completely conserved across the GH5 family. A His residue at position 198 is highly conserved across the GH5 family but is substituted by a completely conserved Arg within the subfamily 36. The subsequent residue (G199) is also highly conserved in subfamily 36, whereas it is highly variable in GH5 family. Other highly conserved residues within the subfamily 36 include F17, P140, H207, W212 and F298.
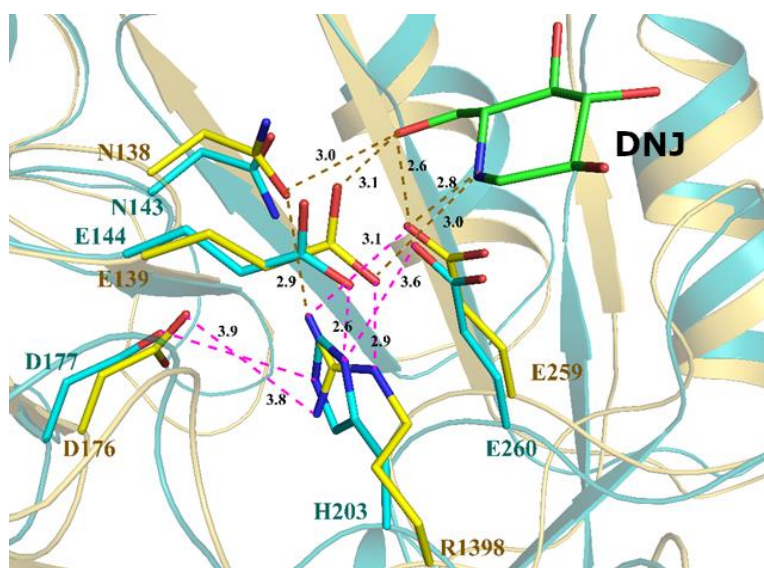

Figure S6. A unique Glu-Arg-Glu charge-relay network of the catalytic glutamates in the GH5_36 subfamily. Close up view of the superposition of the catalytic sites of GH5_36 TmCel5B (PDB: 7EC9, yellow) and GH5_25 FnCel5A (PDB: 3NCO, cyan). The His and Arg residues are structurally equivalent and make ionic interactions with the catalytic glutamates and Asp176. Hydrogen bonds and ionic interactions are represented in brown and pink dashed lines, respectively. All lengths are indicated in Å. The carbon atoms of DNJ, TmCel5B and FnCel5A are represented by green, yellow, and cyan color respectively. Oxygen and nitrogen are represented by red and blue respectively. The cut-off values for bond distances to define H-bonds and ionic interactions are 3.2 and 4.0 Å, respectively.
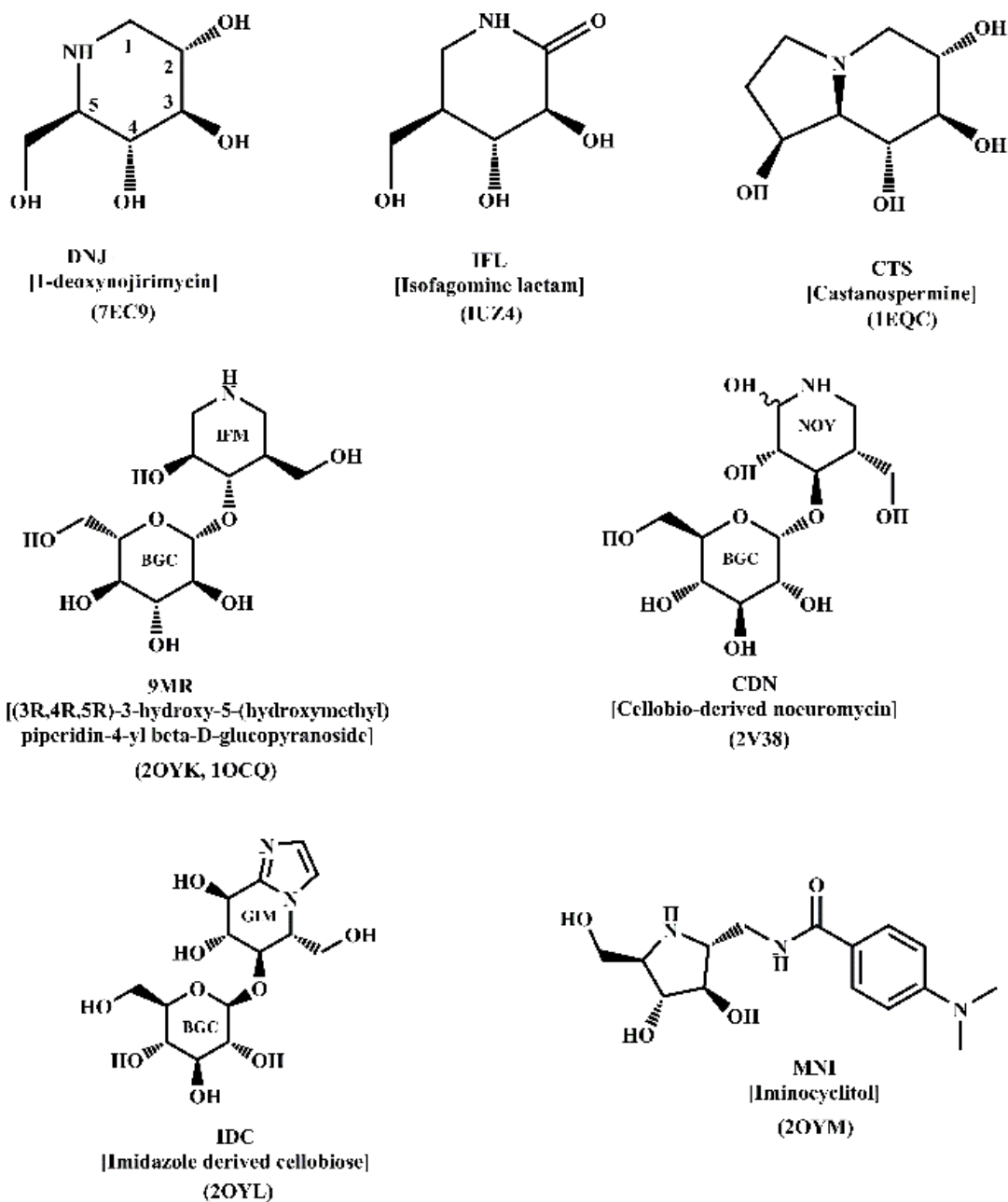
Figure S7. Iminocyclitol ligands bound in structures of complexes of GH5 homologs. DNJ and CTS are iminosugars where the endocyclic O-atom has been replaced by N-atom, whereas in IFM and IFM-derived compounds (9MR and IFL), the anomeric carbon is replaced by N-atom. MNI is a five membered ring iminocyclitol and 9MR and IDC are cellobioses. The atom numbered as position 1 corresponds to the anomeric carbon in glucose. The corresponding PDB ids are also mentioned. IFL-Isofagomine; IFM- Isofagomine; MNI- Five membered ring iminocyclitol; CTS- Castanospermine; DNJ- 1-deoxynojirimycin; IDC- Imidazole derived cellobiose; 9MR- cellobiose derived IFM, CDN-cellobio-derived neuromycin
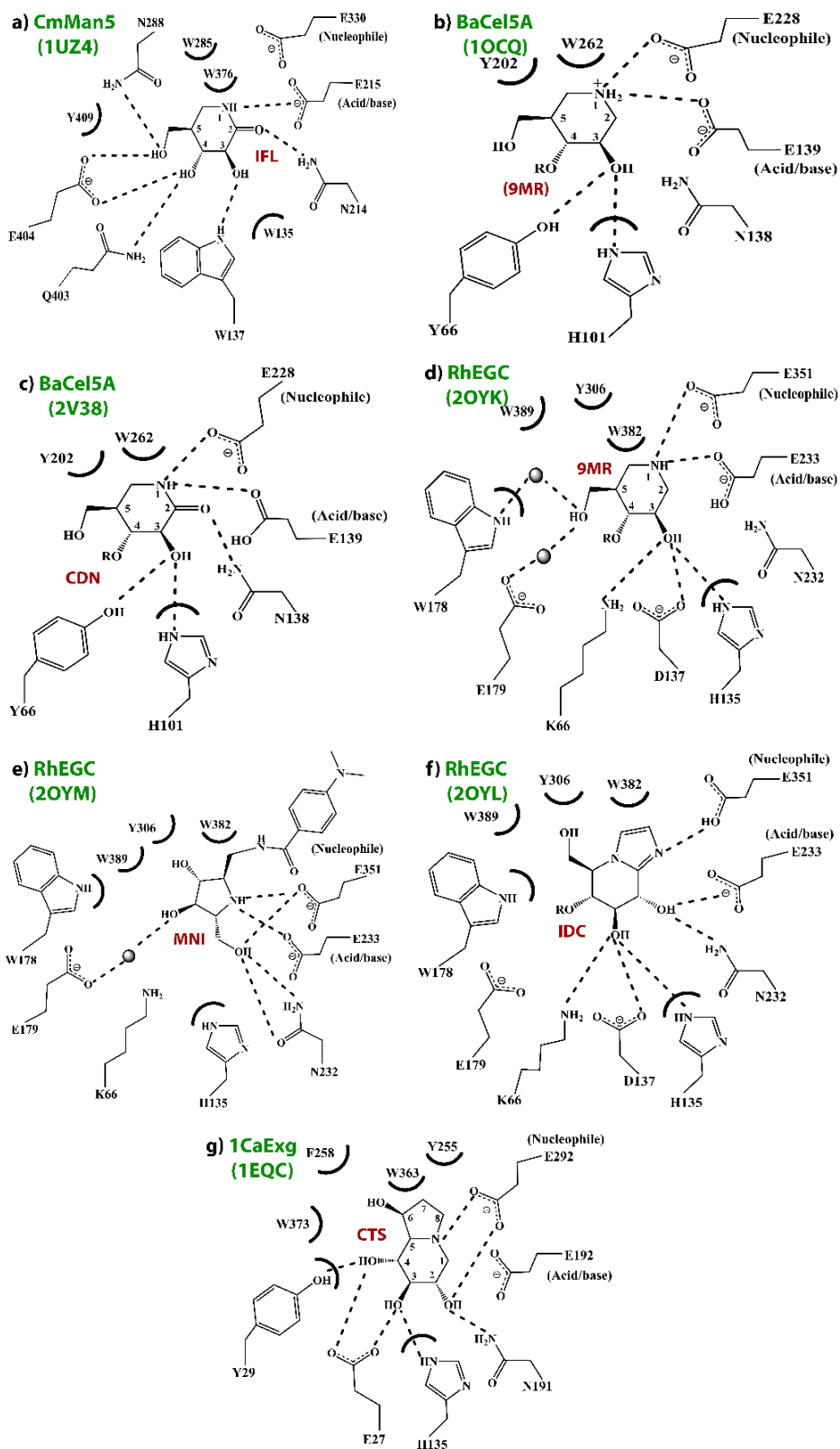
Figure S8. Schematic representation of the -1 subsite interactions of iminocyclitols in complexes of GH5 homologs. The names of GH5 enzymes and PDB IDs are labelled as given in Table S2.

Table S1. List of structures of GH complexes with DNJ and derivatives

S. No	PDB ID	Resolution (Å)	Chain	Iminosugar conformation**	Activity	Reference	Family, [Fold], mechanism
1	1OIM	2.15	A	^4E	β-glucosidase	[1]	GH1 [(β/α)$_8$] Retaining
			B	$^1S_3$			
2	2J77	2.10	A	$^4C_1$		[2]	
			B	$^1S_3$			
3	5NS8	1.55	A, B	$^4H_3$	β-glucosidase	[3]	
			C	^4E			
4	3VIG	0.99	A	$^4H_3$	β-glucosidase	[4]	
5	6R5N	2.00	A, B	$^4H_3$	β-glucosidase	[5]	GH3 Retaining
6	4IID	2.30	A, B	$^4C_1$	β-glucosidase	[6]	
7	6LGC	1.90	A, B	$^4C_1$	sucrose hydrolase	[7]	GH13 [(β/α)$_8$] Retaining
8	1I75	2.00	A, B	$^4C_1$	glucano-transferase	[8]	
9	2PWD	1.80	A, B	$^4C_1$	trehalulose synthase	[9]	
10*	2YA2	2.37	A	$^4C_1$	α-amylase	[10]	
11	3GBE	1.70	A	$^4C_1$	trehalulose synthase	[11]	
12	5WCZ	1.58	A, B	$^4C_1$	α-glucosidase	[12]	
13*	2XG9	1.80	A	$^4C_1$	β-amylase	[13]	GH14 [(β/α)$_8$] Inverting
14*	6F9J	1.67	A	$^4C_1$	β-amylase	N.A.	
15	1DOG	2.30	A	$^4C_1$	α-glucosidase	[14]	GH15 [(α/α)$_6$] Inverting
			B	^4E			
16*	4HP0	1.19	A	$^4C_1$	lysozyme	[15]	GH22 Retaining
17*	4HPI	1.19	A	$^4C_1$	lysozyme		
18	3GXT	2.70	A	$^4C_1$	α-galactosidase	[16]	GH27 [(β/α)$_8$] Retaining
			B	^4E			
19*	5NGL	1.85	A	$^4H_3$	β-glucanase	[17]	GH30 [(β/α)$_8$] Retaining
			B, C	^4E			
20	2X2J	2.35	A, B	$^4C_1$	α-glucan lyase	[18]	GH31 [(β/α)$_8$] Retaining
21	5DKY	1.60	A	$^4C_1$	α-glucosidase	[19]	
22	5IEE	1.92	A	$^4C_1$	α-glucosidase	[20]	
23	5NN5	2.00	A	$^4C_1$	α-glucosidase	[21]	
24	7VKZ	2.00	A, B	$^4C_1$	beta-1,2-glucosyltransferase	[26]	GH35 Retaining
25	3QFZ	2.39	A, B	$^4C_1$	phosphorylase	[22]	GH94 [(α/α)$_6$] Inverting
26	3QG0	2.70	A	$^4C_1$	phosphorylase		
			B	$^1S_3$			
27	2JKE	1.70	A, B	$^4C_1$	α-glucosidase	[23]	GH97 [(β/α)$_8$] Both
28	5BX3	1.96	A	$^4C_1$	β-glucosidase	[24]	GH116 [(α/α)$_6$] Retaining
29	1DIE	2.50	A, B	$^4C_1$	xylose isomerase	[25]	Not assigned

* DNJ derivatives
** The conformation was determined using Privateer software of CCP4i suite [27]

8

Table S2. List of crystal structures of complexes of GH5 homologs with bound iminocyclitols.

S. No.	UniProt ID (activity)	Enzyme name (GH5 Subfamily)	Sequence Identity (%)[#]	PDB ID	RMSD (Å)[*]	Ligand (Ki)[**]	Conformation (-1 site sugar)	Reference
1.	Q6QT42 (exo-mannosidase)	CmMan5 (GH5_7)	19.1	1UZ4	2.0 (234)	IFL (400 μM)	$B_{2,5}$	[28]
2.	O85465 (endo-cellulase)	BaCel5A (GH5_2)	20.2	1OCQ	2.4 (234)	9MR (0.70 μM)	$^4C_1$	[29]
				2V38	2.4 (234)	CDN (0.03 μM)	$^4C_1$	[30]
3.	O33853 (endo-glyco-ceramidase)	RhEGC (GH5_28)	23	2OYK	2.0 (234)	9MR (5 μM)	$^4C_1$	[31]
				2OYM	2.0 (234)	MNI (10 μM)	^4E	
				2OYL	2.0 (234)	IDC (0.5 μM)	$^4C_1$	
4.	P29717 (exo-glucanase)	CaExg (GH5_9)	26	1EQC	1.9 (234)	CTS (ND)	$^{1,4}B$	[29]
5.	Q9X274 (mannanase)	TmCel5B (GH5_36)	-	7EC9	-	DNJ (ND)	$^4C_1$	This study

[#] Sequence identity values correspond to pairwise comparisons with TmCel5B

[*] RMSD values correspond to pairwise comparisons with TmCel5B. The number in bracket indicates the number of Cα atoms used for superpositions.

[**] Ki- inhibition constant, IFL- Isofagomine; IFM- Isofagomine; MNI- Five membered ring iminocyclitol; CTS- Castanospermine; DNJ- 1-deoxynojirimycin; IDC- Imidazole derived cellobiose; 9MR- cellobiose derived IFM, ND- Not determined

Table S3. Interactions of TmCel5B with DNJ and TRS.

Ligand	H-bond interactions			Non-bonded contacts (<4.0 Å)
	Ligand atom	Protein atom	Distance (Å)	
DNJ	O2	His207 NE2	2.5	Phe,17, Asn138, Glu139, Ty200, His207, Trp212, Glu259, Trp292
	O2	Trp212 NE1	3.2	
	O3	HOH890 Tyr200 OH	2.6-w-2.9*	
	O4	HOH72 Trp292 NE1	2.8-w-3.0*	
	N5	Glu259 OE1	2.8	
	N5	Glu 139 OE2	3.1	
	N5	Glu 139 OE1	3.0	
	O6	Glu259 OE2	2.6	
	O6	Asn138 ND2	3.0	
	O6	Glu139 OE1	3.1	
	O6	His86 NE2	2.8	
TRS	O2	His207 NE2	3.0	Tyr200, His207, Glu259, Glu139, His86, Trp292, Trp212
	O2	Trp212 NE1	3.1	
	O1	Glu139 OE2	2.8	
	O3	HOH664 Glu139 OE1	3.2-w-3.0*	
	O3	HOH664 His86 NE2	3.2-w-2.8*	
	O3	HOH664 Asn138 ND2	3.2-w-3.1*	
	O3	HOH664 Glu259 OE2	3.2-w-2.8*	
	N	Glu259 OE1	2.5	
	N	HOH664 Glu259 OE2	2.5-w-2.8*	
	N	HOH664 Glu139 OE1	2.5-w-3.0*	
	N	HOH664 Asn138 ND2	2.5-w-3.1*	
	N	HOH664 His86 NE2	2.5-w-2.8*	

*-w- water mediated interactions

References.

[1]   D.L. Zechel, A.B. Boraston, T. Gloster, C.M. Boraston, J.M. Macdonald, D.M.G. Tilbrook, R. V. Stick, G.J. Davies, Iminosugar Glycosidase Inhibitors: Structural and Thermodynamic Dissection of the Binding of Isofagomine and 1-Deoxynojirimycin to β-Glucosidases, J. Am. Chem. Soc. 125 (2003) 14313–14323.

[2]   T.M. Gloster, P. Meloncelli, R. V. Stick, D. Zechel, A. Vasella, G.J. Davies, Glycosidase inhibition: An assessment of the binding of 18 putative transition-state mimics, J. Am. Chem. Soc. 129 (2007) 2345–2354.

[3]   D. Mhaindarkar, R. Gasper, N. Lupilov, E. Hofmann, L.I. Leichert, Loss of a conserved salt bridge in bacterial glycosyl hydrolase BgIM-G1 improves substrate binding in temperate environments, Commun. Biol. 1 (2018) 1–11.

[4]   W.Y. Jeng, N.C. Wang, C.T. Lin, W.J. Chang, C.I. Liu, A.H.J. Wang, High-resolution structures of Neotermes koshunensis β-glucosidase mutants provide insights into the catalytic mechanism and the synthesis of glucoconjugates, Acta Crystallogr. Sect. D Biol. Crystallogr. 68 (2012) 829–838.

[5]   K. V. Mahasenan, M.T. Batuecas, S. De Benedetti, C. Kim, N. Rana, M. Lee, D. Hesek, J.F. Fisher, J. Sanz-Aparicio, J.A. Hermoso, S. Mobashery, Catalytic Cycle of Glycoside Hydrolase BglX from Pseudomonas aeruginosa and Its Implications for Biofilm Formation, ACS Chem. Biol. 15 (2020) 189–196.

[6]   K. Suzuki, J.I. Sumitani, Y.-W.W. Nam, T. Nishimaki, S. Tani, T. Wakagi, T. Kawaguchi, S. Fushinobu, Crystal structures of glycoside hydrolase family 3 β-glucosidase 1 from Aspergillus aculeatus, Biochem. J. 452 (2013) 211–221.

[7]   T. Miyazaki, E.Y. Park, Structure-function analysis of silkworm sucrose hydrolase uncovers the mechanism of substrate specificity in GH13 subfamily 17 exo-α-glucosidases, J. Biol. Chem. 295 (2020) 8784–8797.

[8]   R. Kanai, K. Haga, K. Yamane, K. Harata, Crystal structure of cyclodextrin glucanotransferase from alkalophilic Bacillus sp. 1011 Complexed with 1-deoxynojirimycin at 2.0 Å resolution, J. Biochem. 129 (2001) 593–598.

[9]   S. Ravaud, X. Robert, H. Watzlawick, R. Haser, R. Mattes, N. Aghajari, Trehalulose synthase native and carbohydrate complexed structures provide insights into sucrose isomerization, J. Biol. Chem. 282 (2007) 28126–28136.

[10]   A. Lammerts Van Bueren, E. Ficko-Blean, B. Pluvinage, J.H. Hehemann, M.A. Higgins, L. Deng, A.D. Ogunniyi, U.H. Stroeher, N. El Warry, R.D. Burke, M. Czjzek, J.C. Paton, D.J. Vocadlo, A.B. Boraston, The conformation and function of a multimodular glycogen-degrading pneumococcal virulence factor, Structure. 19 (2011) 640–651.

[11]   S. Ravaud, X. Robert, H. Watzlawick, R. Haser, R. Mattes, N. Aghajari, Structural determinants of product specificity of sucrose isomerases, FEBS Lett. 583 (2009) 1964–1968.

[12]   M.W. van der Kamp, E.J. Prentice, K.L. Kraakman, M. Connolly, A.J. Mulholland, V.L. Arcus, Dynamical origins of heat capacity changes in enzyme-catalysed reactions., Nat. Commun. 9 (2018) 1177.

[13]   M. Rejzek, C.E. Stevenson, A.M. Southard, D. Stanley, K. Denyer, A.M. Smith, M.J. Naldrett, D.M. Lawson, R.A. Field, Chemical genetics and cereal starch metabolism: Structural basis of the non-covalent and covalent inhibition of barley β-amylase, Mol. Biosyst. 7 (2011) 718–730.

[14]   E.M.S. Harris, R.B. Honzatko, A.E. Aleshin, L.M. Firsov, Refined Structure for the Complex of 1-Deoxynojirimycin with Glucoamylase from Aspergillus awamori var. X100 to 2.4-Å Resolution, Biochemistry. 32 (1993) 1618–1626.

[15]   M. Ogata, N. Umemoto, T. Ohnuma, T. Numata, A. Suzuki, T. Usui, T. Fukamizo, A

novel transition-state analogue for lysozyme, 4-O-β-Tri-N- acetylchitotriosyl moranoline, provided evidence supporting the covalent glycosyl-enzyme intermediate, J. Biol. Chem. 288 (2013) 6072–6082.

[16] J.Q. Fan, S. Ishii, N. Asano, Y. Suzuki, Accelerated transport and maturation of lysosomal α-galactosidase A in fabry lymphoblasts by an enzyme inhibitor, Nat. Med. 5 (1999) 112–115.

[17] M.J. Temple, F. Cuskin, A. Baslé, N. Hickey, G. Speciale, S.J. Williams, H.J. Gilbert, E.C. Lowe, A Bacteroidetes locus dedicated to fungal 1,6-β-glucan degradation: Unique substrate conformation drives specificity of the key endo-1,6-β-glucanase., J. Biol. Chem. 292 (2017) 10639–10650.

[18] H.J. Rozeboom, S. Yu, S. Madrid, K.H. Kalk, R. Zhang, B.W. Dijkstra, Crystal structure ofβ-1,4-glucan lyase, a unique glycoside hydrolase family member with a novel catalytic mechanism, J. Biol. Chem. 288 (2013) 26764–26774.

[19] T. Satoh, T. Toshimori, G. Yan, T. Yamaguchi, K. Kato, Structural basis for two-step glucose trimming by glucosidase II involved in ER glycoprotein quality control, Sci. Rep. 6 (2016) 4–10.

[20] A.T. Caputo, D.S. Alonzi, L. Marti, I.-B.B. Reca, J.L. Kiappes, W.B. Struwe, A. Cross, S. Basu, E.D. Lowe, B. Darlot, A. Santino, P. Roversi, N. Zitzmann, Structures of mammalian ER α-glucosidase II capture the binding modes of broad-spectrum iminosugar antivirals, Proc. Natl. Acad. Sci. U. S. A. 113 (2016) E4630–E4638.

[21] V. Roig-Zamboni, B. Cobucci-Ponzano, R. Iacono, M.C. Ferrara, S. Germany, Y. Bourne, G. Parenti, M. Moracci, G. Sulzenbacher, Structure of human lysosomal acid α-glucosidase-A guide for the treatment of Pompe disease, Nat. Commun. 8 (2017) 1111.

[22] S. Fushinobu, M. Hidaka, A.M. Hayashi, T. Wakagi, H. Shoun, M. Kitaoka, Interactions between Glycoside Hydrolase Family 94 Cellobiose Phosphorylase and Glucosidase Inhibitors, J. Appl. Glycosci. 58 (2011) 91–97.

[23] T.M. Gloster, J.P. Turkenburg, J.R. Potts, B. Henrissat, G.J. Davies, Divergence of catalytic mechanism within a glycosidase family provides insight into evolution of carbohydrate metabolism by human gut flora., Chem. Biol. 15 (2008) 1058–1067.

[24] R. Charoenwattanasatien, S. Pengthaisong, I. Breen, R. Mutoh, S. Sansenya, Y. Hua, A. Tankrathok, L. Wu, C. Songsiriritthigul, H. Tanaka, S.J. Williams, G.J. Davies, G. Kurisu, J.R.K. Cairns, Bacterial β-Glucosidase Reveals the Structural and Functional Basis of Genetic Defects in Human Glucocerebrosidase 2 (GBA2), ACS Chem. Biol. 11 (2016) 1891–1900.

[25] C.A. Collyer, D.M. Blow, Observations of reaction intermediates and the mechanism of aldose-ketose interconversion by D-xylose isomerase, Proc. Natl. Acad. Sci. U. S. A. 87 (1990) 1362–1366.

[26] K. Kobayashi, H. Shimizu, N. Tanaka, K. Kuramochi, H. Nakai, M. Nakajima, H. Taguchi, Characterization and structural analyses of a novel glycosyltransferase acting on the β-1,2-glucosidic linkages, J. Biol. Chem. 298 (2022) 101606.

[27] J. Agirre, J. Iglesias-Fernández, C. Rovira, G.J. Davies, K.S. Wilson, K.D. Cowtan, Privateer: software for the conformational validation of carbohydrate structures, Nat. Struct. Mol. Biol. 22 (2015) 833–834.

[28] F. Vincent, T.M. Gloster, J. Macdonald, C. Morland, R. V Stick, F.M. V Dias, J.A.M. Prates, C.M.G.A. Fontes, H.J. Gilbert, G.J. Davies, Common inhibition of both beta-glucosidases and beta-mannosidases by isofagomine lactam reflects different conformational itineraries for pyranoside hydrolysis., Chembiochem. 5 (2004) 1596–1599.

[29] A. Varrot, C.A. Tarling, J.M. Macdonald, R. V Stick, D.L. Zechel, S.G. Withers, G.J.

Davies, Direct observation of the protonation state of an imino sugar glycosidase inhibitor upon binding., J. Am. Chem. Soc. 125 (2003) 7496–7497.

[30]   P.J. Meloncelli, T.M. Gloster, V.A. Money, C.A. Tarling, G.J. Davies, S.G. Withers, R. V. Stick, D-glucosylated derivatives of isofagomine and noeuromycin and their potential as inhibitors of β-glycoside hydrolases, Aust. J. Chem. 60 (2007) 549–565.

[31]   M.E.C.C. Caines, S.M. Hancock, C.A. Tarling, T.M. Wrodnigg, R. V. Stick, A.E. Stütz, A. Vasella, S.G. Withers, N.C.J.J. Strynadka, The structural basis of glycosidase inhibition by five-membered iminocyclitols: The clan a glycoside hydrolase endoglycoceramidase as a model system, Angew. Chemie - Int. Ed. 46 (2007) 4474–4476.