PORTLAND
PRESS

Research Article

Structural basis of catalysis and substrate recognition by the NAD(H)-dependent α-d-glucuronidase from the glycoside hydrolase family 4

Samar Ballabha Mohapatra and Ⓓ **Narayanan Manoj**

Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India

Correspondence: Narayanan Manoj (nmanoj@iitm.ac.in)

Members of the glycoside hydrolase family 4 (GH4) employ an unusual glycosidic bond cleavage mechanism utilizing NAD(H) and a divalent metal ion, under reducing conditions. These enzymes act upon a diverse range of glycosides, and unlike most other GH families, homologs here are known to accommodate both α- and β-anomeric specificities within the same active site. Here, we report the catalytic properties and the crystal structures of TmAgu4B, an α-D-glucuronidase from the hyperthermophile *Thermotoga maritima*. The structures in three different states include the apo form, the NADH bound holo form, and the ternary complex with NADH and the reaction product D-glucuronic acid, at 2.15, 1.97 and 1.85 Å resolutions, respectively. These structures reveal the step-wise route of conformational changes required in the active site to achieve the catalytically competent state, and illustrate the direct role of residues that determine the reaction mechanism. Furthermore, a structural transition of a helical region in the active site to a turn geometry resulting in the rearrangement of a unique arginine residue governs the exclusive glucopyranosiduronic acid recognition in TmAgu4B. Mutational studies show that modifications of the glycone binding site geometry lead to catalytic failure and indicate overlapping roles of specific residues in catalysis and substrate recognition. The data highlight hitherto unreported molecular features and associated active site dynamics that determine the structure–function relationships within the unique GH4 family.

Introduction

The glycoside hydrolases (GHs) constitute a large group of enzymes of 167 GH families classified currently in the carbohydrate active enzyme database (CAZy) [1,2]. In general, GHs catalyze hydrolysis through one of two distinct processes that are reliant on an acid/base pair of carboxylic residues. The stereochemical outcome is either an inversion or retention of the anomeric center of the glycone product [3]. In both cases, the steps involve a nucleophile stabilized oxocarbenium ion-like transition state [3–5]. Members of the GH4 and GH109 families employ an unusual NAD(H)-dependent glycosidic bond cleavage mechanism [6–9]. Here, the cofactor oxidizing power is used to generate a transient keto intermediate at the glycone as a means of activating the ring to achieve the intended cleavage [10–13]. Apart from NAD(H), the GH4 enzymes require a divalent metal ion, usually Mn^{2+}, and reducing conditions for catalysis [14–16]. The mechanism has been dissected in detail through a combination of studies by structure determination, spectroscopy, mutational analysis, kinetic isotope effects (KIE) and computational studies [14–17]. In the proposed scheme, oxidation at the C3 position of the glycone is kickstarted with deprotonation of C3-OH of glycone by a base/metal-bound hydroxide, with concurrent hydride transfer at C3 to the NAD^+. Deprotonation of C2-H of the glycone by a basic residue with the elimination of the aglycone generates an α,β-unsaturated ketone intermediate. Next, the addition of water to the C1 followed by reprotonation at the C2 Michael acceptor results in

the reduction of the C3 ketone by NADH. The metal ion acts to stabilize the anions and polarize the carbonyl group, generated in the process [15–17]. KIE studies reveal that both C3 oxidation and C2 deprotonation are rate-limiting [15–20].

GH4 homologs are widespread in Bacteria and Archaea but are notably absent from the eukaryotes. Nevertheless, few GH4 members have been characterized for their structure–function properties. The GH4 possesses a unique fold that includes a Rossman-fold domain, typical of dehydrogenases and oxidoreductases [13]. Moreover, homologs act on a diverse range of glycosides, and this family is exceptional in its ability to accommodate both α- and β-anomeric specificity within the same active site. At least six different specificities are known, namely, α-D-galactosidase [21], α-D-galacturonidase [22], phospho-α-D-glucosidase [15], phospho-β-D-glucosidase [16], α-D-glucosidase [14] and α-D-glucuronidase [9]. Only eight structures of homologs are available in the PDB. Among these, five have been reported in the literature. The most well-studied members are that of *Bacillus subtilis* phospho-α-D-glucosidase (BsGlvA) (PDB: 1U8X) [15] and *Thermotoga maritima* phospho-β-D-glucosidase (TmBglT) (PDB: 1UP4,1UP6,1UP7) [16], where structures of complexes with Mn^{2+}, NAD(H) and the reaction product 6-phospho-glucopyranose (G6P) have been determined. In the other available structure of the complex of *T. maritima* α-D-glucosidase (TmAglA) with NAD^+ and substrate maltose (PDB: 1OBB), the binding geometry of the NAD_N is in a *syn* orientation and represents a non-productive ternary complex [14]. The structure of metal-bound apo form of *Thermotoga neapolitana* TnAgl (PDB: 3U95) was reported with thermostability properties although its activity is unknown [23]. The other structures, namely, KpAglB (PDB: 6DVV), GsLicH (PDB: 1S6Y), BsLpID (PDB: 3FEF), are of homologs from structural genomics efforts, which are apo/metal-bound holo forms and remain unpublished. In particular, finer details of the structural basis of substrate specificity and stereoselectivity, and the conformational changes associated with the mechanism is not well described due to the limited structural data from homologs of this unique family. Because of their unique mechanism and taxonomic occurrence, these enzymes can be useful targets for inhibitor design. It is noteworthy that *T. maritima*, an anaerobic, hyperthermophilic and heterotrophic bacterium, with a relatively small genome, encodes a large number of GHs compared with other bacterial and archaeal genomes. This is related to its capacity to utilize an extensive range of simple and complex carbohydrates and is a likely outcome of evolutionary processes that shaped its ability to grow in diverse global geothermal conditions [24,25].

We have previously reported the structure of citrate (a competitive inhibitor) and Co^{2+}-bound complex of the *T. maritima* GH4 α-D-glucuronidase (TmAgu4B) (PDB: 6KCX, 1.95 Å resolution) [26]. In the present study, we report three crystal structures of TmAgu4B, in the apo form, the holo form with NADH, and the ternary complex with NADH and product D-glucuronic acid (GlcA). These structures, supported by activity studies of a series of substitution variants, reveal the structural determinants of the mechanism and the specificity for glucuronides. Also, the structures illustrate the step-wise route of conformational changes in the active site, beginning from the apo through to a holo form and achieving the final ternary complex that is competent for hydrolysis. These studies expand our understanding of the distinct molecular features and dynamics that define the mechanism, and the substrate binding and specificity, in this unusual family of GHs.

Experimental procedures

Recombinant DNA constructs and site-directed mutagenesis

The *TM0752* gene encoding protein TmAgu4B (471 residues) was obtained as a plasmid from the DNASU plasmid repository. The gene in pMH2T7 vector contains a N-terminal 12 residue His_6 tag. All mutants were generated by PCR using site-directed mutagenesis method. Briefly, Dpn1 enzyme was added to PCR products and the reaction mixture was kept for 4 h at 37°C for digestion of parental template. After Dpn1 digestion, the PCR product was transformed into *Escherichia coli* DH5α cells. Positive clones were selected by ampicillin resistance and grown over night in Luria Bertani (LB) broth with 0.1 mg ml^{-1} ampicillin. Mutant plasmids were isolated by Qiagen Plasmid Mini Kit as per manufacturer's instructions. The mutations were confirmed by DNA sequencing analyses.

Protein expression and purification

The plasmids were transformed into BL21 CodonPlus (DE3)-RIL cells and were grown in LB Broth supplemented with 0.1 mg ml^{-1} ampicillin and 0.034 mg ml^{-1} chloramphenicol, and grown at 37°C with shaking conditions. Gene expression was induced at $A_{600} \sim 0.6$ by adding 0.2% (w/v) L (+) arabinose and the cell culture grown at 30°C for another 4–5 h. The harvested cells were sonicated in buffer (50 mM Tris–HCl pH 7.5,

20 mM imidazole, 300 mM NaCl). The lysate was centrifuged and the supernatant was subjected to heat shock (60°C for 35 min) and pelleted down by centrifugation. The clear supernatant was applied to Ni-Nitrilotriacetic acid (Ni^{2+}-NTA) sepharose column (GE Healthcare) equilibrated with lysis buffer. The column was washed with wash buffer (50 mM Tris–HCl pH 7.5, 50 mM imidazole, 300 mM NaCl,) and protein was eluted with elution buffer (50 mM Tris–HCl, pH 7.5, 250 mM imidazole, 300 mM NaCl). The eluted protein was desalted and stored in buffer (50 mM Tris–HCl pH 7.5, 100 mM NaCl). The oligomeric states of purified samples were estimated by size exclusion chromatography performed at 4°C using a Superdex^{TM} 200 (10/300 GL) analytical column (GE Healthcare). The protein sample (5 mg ml^{-1} in 200 µl) was first incubated with 300 mM DTT, 100 mM NAD$^+$, 100 mM MnCl$_2$ at room temperature. Protein concentration was determined by measuring absorbance at 280 nm, using a calculated value of extinction coefficient of 101 885 M^{-1} cm^{-1}.

Enzyme assays

In general, *in vitro* characterization of enzymatic properties of GHs are carried out using *p*-nitrophenyl derivatives of monosaccharides, although the natural substrates of these enzymes may be poly-, oligo-saccharides and other sugar moieties. In our study, six synthetic monosaccharide derivatives, namely, *p*-nitrophenyl-α-D-glucuronic acid (*p*NP-α-GlcA), *p*-nitrophenyl-β-D-glucuronic acid (*p*NP-β-GlcA), *p*-nitrophenyl-α-D-glucose (*p*NP-α-Glc), *p*-nitrophenyl-β-D-glucose (*p*NP-β-Glc) and *p*-nitrophenyl-α-D-galactose (*p*NP-α-Gal), *p*-nitrophenyl-β-D-galactose (*p*NP-β-Gal), purchased from TCI India Pvt. Ltd, were used to examine the substrate specificity of TmAgu4B. The NAD$^+$, MnCl$_2$ and D-glucuronic acid (GlcA) were purchased from Sigma–Aldrich. All chemicals and buffers were of analytical grade. α-D-glucuronidase activity of TmAgu4B was determined by monitoring the continuous release of product *p*-nitrophenol at 405 nm over 2 min in a PerkinElmer Lambda 25 UV–visible spectrophotometer with a Peltier system. The reaction mixture contains MnCl$_2$, NAD$^+$, DTT and enzyme (preincubated at 40°C for 2 min) to final concentration of 0.2 mM, 0.5 mM, 30 mM and 1.8 µM, respectively. The reaction was initiated by adding and mixing of 0.5 mM *p*NP-α-GlcA (preincubated at 40°C for 2 min) to a final volume of 100 µl [24]. The molar absorption coefficient for *p*-nitrophenol under assay conditions used is 7200 mM^{-1} cm^{-1}. One unit of α-D-glucuronidase activity of TmAgu4B is defined as the amount of enzyme that releases 1 µmol of *p*-nitrophenol per min under standard assay conditions.

The effect of pH and temperature on the α-glucuronidase activity of TmAgu4B were determined by measuring the specific activity over the pH in the range of 6.5–10 and temperature in the range of 30–100°C. The effect of different metals on α-glucuronidase activity of TmAgu4B was determined by measuring the specific activity in the presence of MnCl$_2$, MgCl$_2$, CaCl$_2$. Activities of mutants were measured following the protocol described above.

Bisubstrate initial velocity studies were performed by kinetic parameter estimation for one of the two substrates, A and B, at different concentrations of the second substrate. The substrate-dependent rates were fitted to Michaelis–Menten kinetics using GraphPad Prism 5.0, following substrate depletion method. In the first set of experiments, the concentration of *p*NP-α-GlcA was varied at three fixed concentrations of NAD$^+$. In the second set, the substrates A and B were reversed and the above-described protocol was repeated. The two sets of data obtained above used to test bisubstrate mechanisms using equations (1) and (2), representing sequential single-displacement and ping-pong double-displacement mechanisms, respectively [27].

$$v = \frac{V_{max}[A][B]}{K_s^A K_m^B \; + \; K_m^B[A] + K_m^A[B] + [A][B]} \tag{1}$$

$$v = \frac{V_{max}[A][B]}{K_m^B[A] + K_m^A[B] + [A][B]} \tag{2}$$

where V_{max} is the maximum possible velocity when A and B are both saturating, K_m^A is the concentration of A which gives ½ V_{max} when B is saturating, K_m^B is the concentration of B which gives ½ V_{max} when A is saturating, K_s^A is the dissociation constant for the EA complex.

All measurements were performed in triplicates. To differentiate between the sequential and ping-pong mechanisms, the corresponding Lineweaver–Burk double-reciprocal plots were plotted for each experiment using GraphPad Prism 5.0. Furthermore, the global non-linear regression fitting of data to the equations ($n = 3$) were carried out using the ANEMONA.XLT Excel template to generate the kinetic parameters [28].

Determination of K_d values of MnCl$_2$, NAD$^+$, DTT binding to TmAgu4B

TmAgu4B was first extensively dialyzed in 15 l of buffer 50 mM Tris–HCl, 100 mM NaCl, pH 7.5 to remove any bound divalent metal ions and NAD$^+$. For determination of the K_d value of Mn^{2+}, MnCl$_2$ concentration was varied gradually from 0.01 mM to 1.0 mM while keeping the NAD$^+$, DTT and enzyme (preincubated at 40°C for 2 min) concentrations constant at 0.5 mM, 30 mM and 1.8 µM, respectively, in 50 mM HEPES pH 7.5 buffer. The reaction mixture was left at 40°C for another 2 min and the reaction was initiated by adding 0.5 mM pNP-α-GlcA (preincubated at 40°C for 2 min) to a final volume of 100 µl. For determination of the K_d value of DTT, the DTT concentration was varied gradually from 0.5 mM to 30 mM while keeping the MnCl$_2$ and NAD$^+$ and enzyme (preincubated at 40°C for 2 min) concentrations constant at 0.2 mM, 0.5 mM and 1.8 µM, respectively. The K_d values were estimated by fitting the specific activity against the concentration of the respective ligand binding curve using GraphPad Prism 5.0.

Isothermal calorimetry (ITC) experiments were performed using the MicroCal VP-ITC (GE Healthcare) to measure the K_d value of Mn^{2+} binding. The experiments were carried out at 25°C in buffer A (50 mM Tris–HCl pH 8.0, 100 mM NaCl, 0.1 mM DTT). 30 injections of MnCl$_2$ solution were added to the calorimetric cell containing protein (50 µM) in 2 µl increments from a computer-controlled syringe with an initial concentration of 5 mM. Injections were performed at 200 s interval with a stirring speed of 1000 rpm. The binding isotherms from the heat changes were used to calculate the standard free energy of binding ($\Delta G°$) according to the equation $\Delta G° = RT\ln K_a$, where R is gas constant, T is absolute temperature and K_a is association constant. The number of binding sites (n) was obtained from the binding isotherm. Curve fitting and derivation of thermodynamic parameters were carried out with ORIGIN version 7.0 software package (MicroCal). Data were fitted with single site binding equation (3) for $\underline{i}$th injection

$$\Delta Q(i) = Q(i) + \frac{\partial V_i}{V_0}\left[\frac{Q(i) + Q(i-1)}{2}\right] - Q(i-1) \tag{3}$$

where ∂V_i is the displaced volume, V_0 = active cell volume, $\Delta Q(i)$ is the heat change per ith injection and Q is defined by equation (4)

$$Q = \frac{nM_t\Delta H V_0}{2}\left[1 + \frac{X_t}{nM_t} + \frac{1}{nKM_t} - \sqrt{\left(1 + \frac{X_t}{nM_t} + \frac{1}{nKM_t}\right)^2 - \frac{4X_t}{nM_t}}\right] \tag{4}$$

where ΔH is the molar heat capacity of the ligand binding, K = binding constant, n = number of sites, M_t is free concentration of macromolecule in V_0, X_t is the free concentration of ligand.

Monitoring hydride transfer through steady-state accumulation of NADH

The reduction of NAD$^+$ through the course of reaction was monitored by recording the UV absorbance spectra of NADH at 340 nm. The reaction mixture (100 µl) contains 1.8 µM of enzyme, 0.2 mM MnCl$_2$, 30 mM DTT, 1 mM pNP-α-GlcA and 50 mM HEPES (pH 7.5) at 40°C. The reaction was started by adding 0.5 mM NAD$^+$ and was monitored over 30 s. The conversion of NAD$^+$ to NADH was also monitored by measuring the emission fluorescence spectra of NADH between 400 nm and 550 nm after exciting the sample with 340 nm in Fluorolog HORIBA spectrofluorometer. The 650 µl of sample reaction constitutes 1.8 µM of TmAgu4B, 0.5 mM NAD$^+$, 0.2 mM MnCl$_2$, 30 mM DTT, 20 mM Tris–HCl (pH 8.0), 0.5 mM pNP-α-GlcA and blank without pNP-α-GlcA. Both the sample and blank were preincubated for 30 min at 50°C before the emission spectra were recorded.

Protein crystallization

Initial crystallization trials were performed by hanging drop vapor diffusion method using commercially available crystallization screens. One microliter of protein at 10 mg ml^{-1} was mixed with 1 µl of reservoir solution and equilibrated with 0.5 ml of reservoir solution. The best hits were optimized by varying additives, pH, precipitants, protein concentration. For crystallization of the apo form, the protein was dialyzed in 15 l of storage buffer (50 mM Tris–HCl, 100 mM NaCl, pH 7.5) and was incubated in ice with 7.27 mM of DTT. Thin plates of crystals of TmAgu4B apo form stacked together were obtained at 20°C by mixing 2 µl of protein (40 mg ml^{-1}) with 2 µl

of reservoir solution composed of 14% PEG 3350, 0.2 M trilithium citrate, 0.1 M imidazole with pH 5.8 or 6.2 and 2-propanol. The NADH bound holo form of TmAgu4B crystals were grown by co-crystallizing protein (40 mg ml^{-1}) solution with 7.27 mM of DTT, 7.27 mM of NAD$^+$ in the same condition as in the apo form. The ternary complex form crystals were prepared by soaking apo form crystals in mother liquor containing 7.27 mM of MnCl$_2$, 50 mM of NAD$^+$, 50–100 mM D-glucuronic acid (GlcA) for 10–15 min prior to data collection.

Data collection, structure determination and refinement

X-ray diffraction data were collected on a MAR345 image plate detector mounted on a BRUKER MICROSTAR rotating anode X-ray generator (CuKα = 1.5418 Å). Crystals soaked in a cryoprotectant solution (40% PEG 3350, 0.2 M trilithium citrate, 0.1 M imidazole) for ~5–10 s were mounted and flash-frozen in a nitrogen gas stream at 100 K. All data collected were processed and scaled using MOSFLM and Aimless programs, as implemented in the CCP4 software package [29-31]. Structure solution and refinement was carried out using programs from the PHENIX program suite [32]. The structure was solved by molecular replacement using the PHASER program using one subunit of TmAgu4B (PDB: 6KCX) as the search model [26]. Iterative rounds of restrained maximum-likelihood refinement using phenix_refine were carried out. Model bias was removed by implementing simulated annealing protocols. The structure and restraints of the ligands were generated using electronic ligand builder and optimization workbench (eLBOW) program, as implemented in the PHENIX suite [33]. Model building was carried out using the Coot program [34]. The stereochemical quality of the final models was evaluated using the MolProbity program. Structure representations were generated using the PyMOL program [35].

Results and discussion
Substrate specificity and kinetic properties

Recombinant TmAgu4B was expressed in *E. coli* and purified to homogeneity. The protein, composed of 471 amino acids, exists as a dimer in solution (Supplementary Figure S1A,B). Thermal unfolding studies revealed a canonical two-state unfolding process and a melting temperature T_m of 90°C, consistent with a hyperthermophilic enzyme [36]. Our first objective was to establish the substrate specificity of TmAgu4B using six synthetic substrates. Activity assays revealed that TmAgu4B is specific to *p*-nitrophenyl-α-D-glucuronic acid (*p*NP-α-GlcA), with a specific activity of 0.31 U/mg under assay conditions. However, *p*-nitrophenyl-β-D-glucuronic acid (*p*NP-β-GlcA), *p*-nitrophenyl-α-D-glucose (*p*NP-α-Glc), *p*-nitrophenyl-β-D-glucose (*p*NP-β-Glc) and *p*-nitrophenyl-α-D-galactose (*p*NP-α-Gal), *p*-nitrophenyl-β-D-galactose (*p*NP-β-Gal) did not show detectable activity, confirming that the enzyme is an α-D-glucuronidase. The enzyme exhibits the highest activity at the optimum pH and temperature of 8.0 and 90°C, respectively (Supplementary Figure S1C,D). We performed bisubstrate kinetic characterization by varying NAD$^+$ and *p*NP-α-GlcA concentrations in parallel independent experiments. Hyperbolic kinetics obtained in all cases indicated that the enzyme follows Michaelis–Menten kinetics (Figure 1A,C, Supplementary Figure S2). Furthermore, the intersecting Lineweaver–Burk plots clearly indicate that the TmAgu4B reaction follows a sequential mechanism and not a ping-pong mechanism (Figure 1B,D). The values of the kinetic constants for the sequential mechanism models are listed in Supplementary Table S1.

Characterization of Mn^{2+}, DTT and NAD$^+$ binding

Binding affinities of the metal and reducing agent were estimated using specific activity measurements in the presence of different concentrations of Mn^{2+} and DTT. Using data fitted to a simple hyperbolic binding equation, the calculated K_d values of binding are 0.037 ± 0.005 μM and 1.2 ± 0.06 mM, for Mn^{2+} and DTT, respectively (Supplementary Figure S3A,B). The ITC thermogram of Mn^{2+} shows an affinity value in the sub-micromolar range, consistent with a single binding site (Supplementary Figure S3C). The enzyme has no detectable activity in the absence of Mn^{2+} or in the presence of non-transition metals like Mg^{2+} or Ca^{2+}. Since the recombinant enzyme copurifies with NAD, samples were extensively dialyzed to remove any trace of the cofactor before assaying for activity in the presence of increasing concentrations of the cofactor. The steady-state accumulation of NADH, monitored by absorbance at 340 nm, is consistent with the course of the reaction where NAD$^+$ is converted to NADH (Supplementary Figure S4A) [37]. Addition of substrate *p*NP-α-GlcA to the reaction mixture containing the enzyme, DTT, NAD$^+$ and Mn^{2+}, increased the fluorescence intensity in the region between 400−550 nm with a peak ~445 nm, corresponding to the characteristic emission spectra of NADH. The same was absent from the reaction mixture without substrate (Supplementary
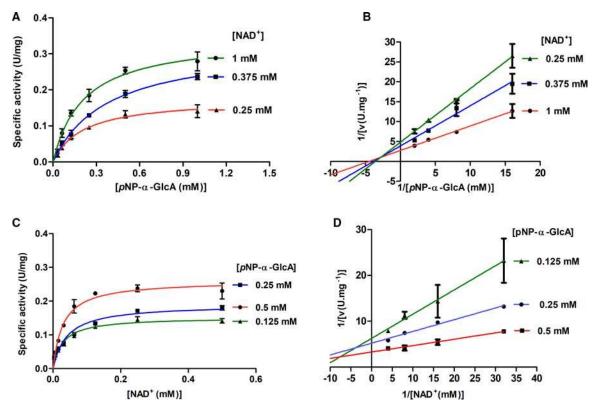
Figure 1. Bisubstrate kinetic initial rate studies of TmAgu4B.
(**A**) Specific activity profiles at variable concentrations of NAD$^+$ at three different fixed concentrations of pNP-α-GlcA. (**B**) Corresponding Lineweaver–Burk plots. (**C**) Specific activity profiles at variable concentrations of pNP-α-GlcA at three different fixed concentrations of NAD$^+$, and (**D**) corresponding Lineweaver–Burk plots. The data are fit to the Michaelis–Menten equation.

Figure S4B). Together, the data confirm the obligate NAD$^+$-dependent activity and the requirement for the divalent transition metal ion and reducing conditions and is in agreement with data from GH4 homologs [15–20,37].

Crystal structures of TmAgu4B

Crystal structures in three forms were determined to understand the structural basis of the mode of action of TmAgu4B. The structures include the apo form (apo-TmAgu4B) and two complexes, with NAD (holo-TmAgu4B), and with NAD and the glycone reaction product, D-glucuronic acid (ter-TmAgu4B). The crystal forms in space group C2 contain a single subunit in the asymmetric unit, with the second subunit of the functional homodimer related by a crystallographic two-fold symmetry. The crystallographic statistics are summarized in Table 1.

Crystals of the apo form were obtained only after extensive dialysis of protein samples in a cofactor-free buffer. NAD(H) is a dissociable prosthetic group, and notably, none of GH4 structures reported so far represent separate snapshots of the apo form and the holo form within the same enzyme. The metal and cofactor-free apo form was refined to 2.15 Å resolution. Extensive crystallization trials were carried out with added NAD$^+$ and Mn^{2+} in co-crystallization and soaking experiments utilizing the apo form to capture the metal and cofactor bound forms. The experiments always had reducing agent DTT. Crystals of the NAD-bound holo form were successfully obtained and the structure determined at 1.97 Å resolution. However, these crystals always lacked the bound metal ion. To determine the cognate ternary complex with a ligand, we carried out multiple co-crystallization trials with substrates and mono- and di-saccharides which were not successful. Therefore, we soaked the crystals of holo-TmAgu4B in mother liquor containing the product and MnCl$_2$ for different time periods. Optimized crystals obtained here allowed the successful determination of the complex

PORTLAND
PRESS

Table 1. Summary of crystallographic statistics

Data set	apo-TmAgu4B	holo-TmAgu4B	ter-TmAgu4B
PDB ID	7CTD	7CTL	7CTM
Wavelength (Å)	1.54	1.54	1.54
Space group	C2	C2	C2
Unit cell parameters			
a, b, c (Å)	72.69, 80.95, 88.38	72.72, 81.28, 89.00	74.15, 80.59, 88.63
Resolution range (Å)[1]	30.30–2.15 (2.27–2.15)	29.65–1.97 (2.02–1.97)	36.80–1.85 (1.89–1.85)
Unique reflections[1]	27 091(3974)	35 755(2570)	40 918(2614)
Multiplicity[1]	6.7 (6.6)	6.5 (6.7)	7.2 (7.2)
Completeness (%)[1]	99.7 (100.0)	100.0 (100.0)	94.2 (89.6)
Mean ($I/\sigma(I)$)[1]	8.7 (1.9)	9.6 (1.5)	12.5 (2.3)
Wilson B-factor	28.4	25.6	17.4
R-merge[1]	0.169 (0.89)	0.129 (1.16)	0.115 (0.797)
$CC_{1/2}$[1],[2]	0.994 (0.581)	0.996 (0.596)	0.997 (0.686)
Model and refinement statistics			
R-work[1],[3]	0.179 (0.283)	0.176 (0.274)	0.1523 (0.256)
R-free[1],[3]	0.225 (0.292)	0.223 (0.323)	0.1963 (0.313)
Non-hydrogen protein atoms	3769	3785	3812
Ligands atoms		44	61
Solvent atoms	187	303	429
Protein residues	469	469	469
Mean B-value (Å^2)			
All atoms (Å^2)	39.8	34.9	24.1
Macromolecules (Å^2)	40.1	34.7	23.4
Ligand name (Å^2)	—	NAI, 43.7	NAI, 23.4, GlcA, 23.8
Solvent (Å^2)	33.1	36.3	31.2
RMS deviation from ideal geometry			
Bond angles (°)	0.53	0.78	0.69
Bond length (Å)	0.002	0.007	0.005
Ramachandran plot analysis			
Favored (%)	97.64	97.43	98.06
Allowed (%)	2.36	2.57	1.94
Outliers (%)	0	0	0

[1]Value in parentheses refer to the highest resolution shell.
[2]$CC_{1/2}$ is the correlation coefficient of the mean intensities between two random half-sets.
[3]$R_{work} = \sum ||F_{obs}| - |F_{calc}||/\sum |F_{obs}|$, R_{free} is calculated for 5% of reflections randomly chosen and not used for refinement; R_{work} is calculated for the remaining reflections.

with NADH and GlcA to a resolution of 1.85 Å. In this context, the three structures represent successive structural snapshots of the enzyme in the apo form, the holo form, and a ternary complex with a cognate sugar acid. Together, these snapshots allow us to capture the structural transitions associated with the entire catalytic cycle and establish the structural basis of the mechanism and substrate specificity and selectivity for a GH4 glucuronidase.

The overall fold of TmAgu4B, as previously described, belongs to the mixed α/β class and contains three separate regions; N-terminal, central catalytic, and C-terminal (Figure 2 and Supplementary Figure S5) [26]. The N-terminal region (residues 1–180), a typical NAD(H)-binding Rossmann fold, comprises of a parallel twisted
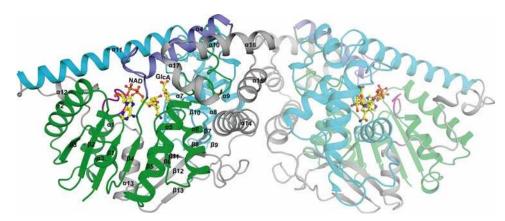
Figure 2. Quaternary structure of TmAgu4B.
Ribbon representation of the ternary complex is shown. The secondary structure elements correspond to the N-terminal Rossmann fold domain shown in lime (residues 1–180), middle catalytic domain shown in gray (residues 181–321) and C-terminal domain shown in cyan (residues 181–321). A part of the N-terminal Rossman fold containing residues 86–132, including α4 helix, is colored blue. Turn I (residues 11–13) is colored purple. All the secondary structures are labeled. The ligands NADH and GlcA are represented as sticks. The second subunit of the dimer is displayed in a lighter shade.

β-sheet sandwiched by α-helices (α1–3, α5–6). The region connecting β4 and α5 (residues 86–132 comprising a loop and helix α4), is an insertion relative to the Rossmann domain and is unique to the GH4 fold. The central catalytic region (residues 181–321) and C-terminal region (residues 322–471), each consisting of a mix of a twisted antiparallel β sheet and α helices, positioned on top of the Rossmann domain, are structural features unique to the GH4 family, and distinguish this family from the evolutionarily related oxyacid dehydrogenases [14–16]. The C-terminal region includes the oligomerization domain that forms the conserved homodimeric interface. The pairwise structural alignments of the eight homologs using the DALI program displayed root mean square deviations (rmsd) in the range 1.2–3.2 Å, Z-scores in the range 31.2–61.8, while the sequence identities range from 18% to 89% [38]. The N-terminal and C-terminal regions are the most structurally conserved across the GH4 family, whereas the central catalytic region is the most divergent and is proposed to be responsible for the diverse substrate specificity and selectivity in this family [36,39].

Interactions of the cofactor and comparison of the apo and holo forms

A previous structure of TmAgu4B at 2.5 Å resolution (PDB: 1VJT, Joint Centre for Structural Genomics, unpublished) lacks the critical nicotinamide (NAD_N) moiety of the bound cofactor. However, the better resolution of data in our study allowed for the modeling of the entire cofactor in the expected binding pocket spanning a conserved cleft between the N-terminal and the central domains (Supplementary Figure S6A). The cofactor modeled as NADH in the holo form makes multiple interactions in the active site (Figure 3A). The adenosyl ring (NAD_A moiety) is well packed in a hydrophobic pocket made of largely aliphatic residues. The O3 and O2 atoms of NAD_N ribose make hydrogen bonds with the main chain NH and the sidechain ND2 atoms of conserved Asn160. Asn160Ala substitution mutant was found to be inactive (Supplementary Figure S7). Water molecule (W1) is structurally conserved in apo-TmAgu4B and in the TmAgu4B complex with Co^{2+} and citrate (PDB: 6KCX) and in the metal bound form of BsLplD (PDB: 3FEF, unpublished), and most likely constitutes a second shell water around the conserved metal coordination network described earlier [26].

Comparison of the apo and holo forms show that the overall structures are largely identical with a rmsd of 0.4 Å, over 469 Cα atoms. However, significant conformational changes occur in the cofactor binding pocket upon NAD binding. These include a rearrangement of the His40 side chain to open up the NAD binding pocket to make a hydrogen bond with Asp93, thereby affecting the neighboring Tyr87 sidechain. As a result, residues in the region 86–96 undergo conformational changes with respect to the apo form (rmsd value of 1.7 Å, over 11 Cα atoms). Several side-chain rearrangements result in key interactions that facilitate cofactor binding. Notably, Arg43, which is disordered in the apo form, is ordered in the holo form. This results in a bidentate
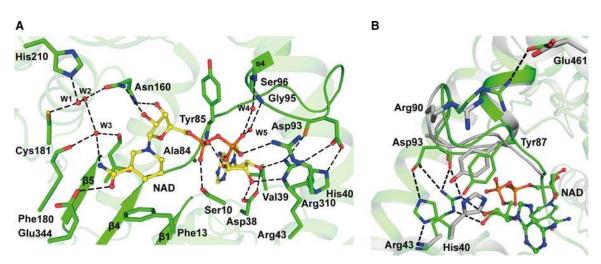
Figure 3. Interactions of holo-TmAgu4B with NAD and structural comparison of apo and holo forms.
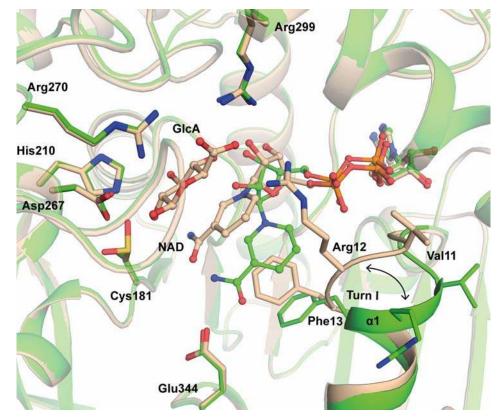(**A**) Ribbon and stick representation of NAD interacting with neighboring residues in holo-TmAgu4B. The adenine (NAD_A) moiety is well packed in a hydrophobic pocket (Val 39, Ala84). W1 is a conserved water molecule. (**B**) Ribbon representation of conformational changes around the NAD_A ribose binding moiety in the holo-TmAgu4B (green) superposed on apo-TmAgu4B (gray). The region 86–96 undergoes major conformational changes in holo-TmAgu4B form along with rearrangements in the side chains of His40, Tyr87. Arg43 and Asp93 are disordered in apo form, whereas these residues are ordered and constitute an ionic lock that anchors the NAD molecule in the holo form. Hydrogen bond/ionic interactions (≤3.2 Å) are shown as dashed lines.

interaction with the adenosyl ribose oxygen atoms. Furthermore, a hydrogen bond and a salt bridge interaction with the carbonyl oxygen and the OD2 of Asp93, respectively, creates an ionic lock that clamps the cofactor (Figure 3B and Supplementary Figure S8A). While the residue at position 43 is usually a Lys/Arg in the family, residue 93 is part of a hairpin loop lining the edge of the top edge of the NAD binding domain and is an insertion element unique to TmAgu4B and TnAgl and absent from the other homologs (Supplementary Figure S9). These rearrangements mark an intra-domain cleft closure upon cofactor binding and it is likely that the resultant ionic lock is specific to these two homologs and are key structural features that modulate divergent kinetics of cofactor binding and reaction rates across paralogous members of the family.

Conformational changes are required for the formation of the competent active site

The structure of the complex of NADH and the cognate reaction product GlcA allowed us to identify the substrate-binding pocket and describe the conformational changes associated with the formation of the ternary complex, hitherto unreported in the GH4 family. Unambiguous electron density is consistent for β-D-glucuronic acid in the chair conformation ($^4C_1$) bound at the expected glycone binding site (Supplementary Figure S10). The structures of the holo form and the ternary complex are largely identical with an overall rmsd of 0.37 Å (over 469 Cα atoms) between them. The largest conformational changes occur in the N-terminal residues 11–13 of helix α1 (Figure 4 and Supplementary Figure S8B). This remarkable secondary structure transition from an α-helical conformation to a β-turn, results in the spatial positions of Val11, Arg12 and Phe13 moving closer towards the substrate (Cα atoms of Val11 and Arg12 move by 3.5 and 4.3 Å, respectively). The side chains undergo large 'swing' movements accordingly, creating a closed state in the ternary complex. First, Arg12 sidechain makes direct ionic interactions both with the GlcA carboxylic O6A and with the pyrophosphate O1N atom (~2.9 Å). Besides, a water bridge connects Arg12, NAD_N ribose O2D and GlcA O6A (Figure 5). Second, a major reorientation of the NAD_N moiety occurs due to the rearrangement of the sidechain of Phe13 in the ternary complex (χ1 rotation from 154° to −73°). In the holo form, NAD_N maintains an offset parallel π–π stacking interaction with Phe13 (~3.4 Å) (Supplementary Figure S6B). In contrast, in the ternary complex, NAD_N reorients (C4N atom moves by ~4.0 Å) to a perpendicular T–π stacking interaction (Figures 4 and 5, Supplementary Figure S8B). As a result, the pro-R face of NAD_N now stacks parallel with

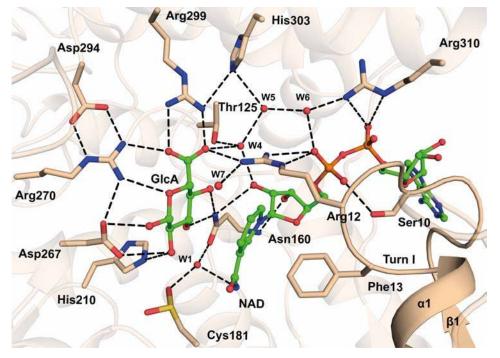**Figure 4. Structural comparison of holo and ternary complex forms.**

Ribbon and stick representations of the superposition of active sites of the holo form (green) and ternary complex (wheat) showing conformational changes upon GlcA binding to the holo form. One set of changes in the ternary complex occurs in N-terminal end of helix α1 that undergoes a transition from an α-helix to a turn (Turn I), enabling a direct interaction of Arg12 with both the carboxylic moiety of GlcA and the pyrophosphate moiety of NAD$^+$. A second set of changes is the conformational changes in the NADH and Phe13 where a $T–\pi$ stacking interaction between Phe13 with NAD$_N$ moiety in the holo form is transformed to a $\pi–\pi$ stacking interaction in the ternary complex.

GlcA bringing the C3 sugar atom to within 3.8 Å of the C4N atom. Moreover, the C3-C4N-N1N and the C4N-C3-C1 angles are 80° and 88°, respectively. The angle between planes defined by C3N-C4N-C5N-C6N of NAD$_N$ and O3-C3-C4-O4 of GlcA is ~11°, mimicking a *syn*-like conformation proposed by Wu et al. [40], for maximum overlap between the C3-H of GlcA and the LUMO-p$_z$ orbital of the C4N of NAD$^+$ (Supplementary Figure S11). This network of CH–π interactions involving the Phe13, the NAD$_N$, and the pyranose ring result in an interaction geometry appropriate for the hydride transfer step in catalysis. Moreover, the NAD$_N$ ribose now makes a new hydrogen bond with Asn160. It appears that the dynamics of residues Arg12 and Phe13 are crucial to positioning the cofactor and the substrate for productive catalysis. Together, we propose that the structural transition of the region Val11-Arg12-Phe13 is substrate-induced and that the cofactor–protein interactions in the ternary complex are coupled to substrate binding (Figure 6). The structural and kinetic data indicate that TmAgu4B adopts the sequential bisubstrate kinetic mechanism and not a ping-pong mechanism. The kinetic analyses cannot distinguish between the random sequential and ordered sequential mechanisms. However, the observed structural transitions from the apo form to holo form and to the ternary complex strongly suggests that TmAgu4B adopts an ordered sequential mechanism where NAD$^+$ is the first substrate while the glucuronide is the second substrate.

Structural insights into the reaction mechanism of TmAgu4B

Inspection of the interactions that GlcA and the cofactor make with the active site residues provides insights into the role of these residues in the mechanism. The GlcA O3 and O4 atoms are hydrogen-bonded to Asn160

Figure 5. Interactions of TmAgu4B with NADH and GlcA in the ternary complex.

Cartoon and stick representation of active site residues interacting with GlcA and NADH. W1 and W4 are conserved water molecules while other water molecules labeled W5, W6, W7 are unique to the complex, and form water bridge interactions connecting the enzyme, NAD and GlcA. The Turn I (residues 11–13) between α1 helix and β1 sheet is labeled. Hydrogen bond interactions (<3.2 Å) are shown as dashed lines.

ND2, while the OD1 atom makes a hydrogen bond with a structurally conserved water (W1) of the metal coordination shell [24]. Asn160, which is present in a *cis* conformation, constitutes a conserved and functionally important Asn160-Pro161 (NP) motif of the GH4 family. The loss of activity in the N160A mutant is consistent with the multiple roles that Asn160 plays in maintaining the precise geometry of interactions of the substrate, the cofactor and the metal coordination via W1 (Figure 5 and Supplementary Figure S7). Catalytic residue Asp267 makes bidentate hydrogen bonds with the GlcA O1 and O2 atoms (2.8–3.0 Å). It is to be noted that the O1 interaction is that for the bound β-conformer of GlcA, whereas this interaction will be absent from
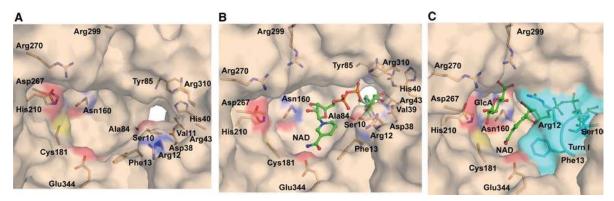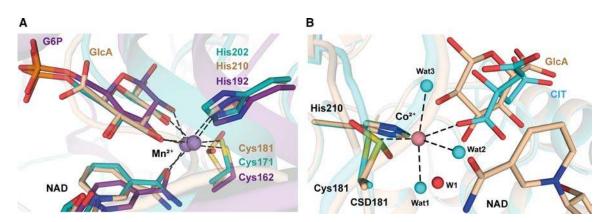

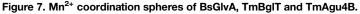**Figure 6. Step-wise sequence of conformational changes in the active site of TmAgu4B.**

Surface and stick representation of active sites in (**A**) apo form, (**B**) holo form, (**C**) ternary complex. The Turn I region in the complex that undergoes a structural transition to occlude the active site is marked in transparent cyan. In all panels, positively charged residues are colored blue and the negatively charged residues are colored red. The bound GlcA and NADH are shown in green stick representation.

the cognate α-linked substrate. Asp267 OD2 atom, 3.4 Å from the C2 atom of GlcA, is expected to act as the base to deprotonate C2. In support, the substitution of Asp267 by Ala abolished the activity (Supplementary Figure S7).

We were unable to locate bound Mn^{2+} in the ternary complex although the crystallization conditions contained excess metal ions. It is likely that the metal-coordinating Cys181 residue, which has been oxidized to cysteine sulfinic acid (Cys-SOOH), prevents metal binding. Nevertheless, the relative positioning and the geometry of interactions between GlcA, the cofactor, and the stringently conserved neighboring residues, Asn160, His210 and Asn209, in TmAgu4B, are nearly identical with that observed in the G6P and metal-bound ternary complexes of BsGlvA and TmBglT (PDB: 1U8X, 1UP6) [18,20]. Moreover, previous reports indicate that all steps of the mechanism (oxidation–elimination–addition) occurs at the glycone binding subsite. Therefore, it is reasonable that the active site interactions in TmAgu4B reflect a catalytically productive complex of a cognate glucuronide substrate. For instance, the distances between the GlcA O2 atom and the metal-coordinating atoms SG of Cys181 and NE2 of His210, are 3.4 and 3.0 Å, respectively. The equivalent interactions of Cys171 and His202 in BsGlvA are 3.4 and 3.7 Å (Figure 7A). These metal–protein direct interactions are invariant irrespective of whether the substrate or cofactor binding sites are occupied or not. Hence, we modeled a metal-bound form of the ternary complex using the structure of the Co^{2+}-bound TmAgu4B (PDB: 6KCX) to position Mn^{2+} (Figure 7). The metal binding site and its interactions are structurally equivalent in all metal-bound homologs including, TmBglT, BsGlvA, TnAgl and BsLplD. The Mn^{2+}-bound ter-TmAgu4B model did not require repositioning of any atoms except for a minor reorientation of the NAD_N amide group to bring the O7N atom within the expected octahedral coordination sphere of Mn^{2+} (Supplementary Figure S12). W1, at a distance of ~2.7 Å from GlcA O3 atom in TmAgu4B, is also present in the holo form. As expected, the substitution of Cys181 by Ala abolished the activity of TmAgu4B (Supplementary Figure S7).

Based on this model, we sought to establish the structural basis of the catalytic cycle of TmAgu4B. The proposed mechanism occurs in two half-reactions. We propose that the Mn^{2+} with W1 together constitute the metal-activated hydroxide ion which acts as the general base to deprotonate the C3-hydroxyl group of the glycone and initiate the reaction (Figure 8, I) [19]. The deprotonation acts in concert with hydride transfer to NAD^+ to give a ketone at C3 and has been previously shown to be rate-limiting [17–20,37,39,41]. The carbonyl intermediate is stabilized by Mn^{2+}, leading to a relatively more acidic proton at C2. In the next step, deprotonation at C2 and concurrent β-elimination at C1 of glycone carbonyl intermediate occurs (Figure 8, III). The β-elimination leading to the loss of two substituent atoms has been widely documented across literature in

Figure 7. Mn^{2+} coordination spheres of BsGlvA, TmBglT and TmAgu4B.
(**A**) Superposition of the Mn^{2+} coordinating residues and ligand groups, namely, His, Cys, G6P, GlcA and NAD_N in BsGlvA (PDB: 1U8X, teal) and TmBglT (PDB: 1UP6, purple) and ternary complex of TmAgu4B (wheat). The metal ion is absent from the ternary complex of TmAgu4B. Comparisons show that the bound GlcA in TmAgu4B is as expected for a productive substrate complex. (**B**) Superposition of the ternary complex of TmAgu4B (wheat) and the Co^{2+} bound citrate complex of TmAgu4B (cyan, PDB: 6KCX). Positioning of the metal ion from this citrate complex into the ternary complex completes the expected octahedral sphere composed of interactions with Cys181, His210, W1, O2/GlcA, O3/GlcA and O7N/NAD. Colors of residue labels correspond to that of the structures.
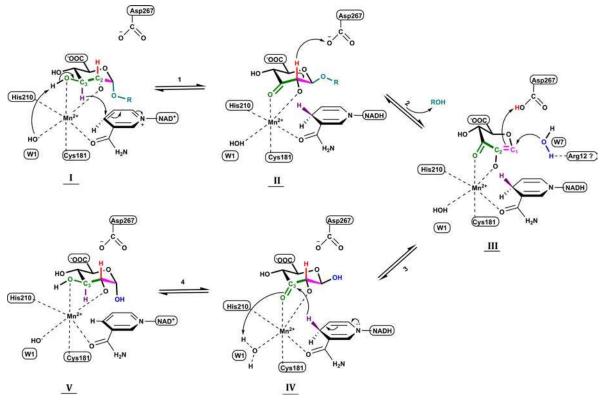
Figure 8. Schematic representation of the proposed TmAgu4B mechanism.

(**I**) Mn^{2+}-OH (W1) deprotonates the C3-OH (green) leading to C3 keto sugar (green) and concurrent C3-H hydride (purple) abstraction by nearby C4N/NAD^+. (**II**) O=C3–C2 enolate species (green) is stabilized by Mn^{2+} leading to slightly acidic C2-H (red). C2 is deprotonated by Asp267 and concerted β-elimination ($E1_{CB}$) of -ROH (teal) leading to a double bond character across the C2=C1 Michael acceptor (pink). (**III**) Water (blue) addition to unsaturated Michael acceptor at C2=C1 (pink) occurs in a Michael addition fashion. (**IV**) Keto group at position C3 (green) is reduced by NADH (purple) and (**V**) Glycone product bound form.

uronic acids since the glycosidic linked residue or the methoxy substituent (at C4 of glycone GlcA in case of TmAgu4B) of the uronic acid residue are good leaving groups [42–43]. Careful KIE studies in combination with structural analysis in TmBglT, suggest that the structurally equivalent Tyr241 deprotonates the C2 atom and elimination of aglycone proceeds through a $E1_{CB}$ or simple β-elimination [16,20,37]. In TmAgu4B, C2-H is deprotonated by the equivalent Asp267 and the β-elimination of the aglycone leaves a α,β-unsaturated Michael acceptor at C1=C2 of GlcA (Figure 8, II, III). This Michael-like acceptor undergoes base-catalysis by water, generating a keto group at C3, while the C2 is reprotonated by Asp267. Finally, the onboard NADH reduces the C3 ketone, completing the overall cycle (Figure 8, IV, V).

Interestingly, a water molecule (W7) is present in the ternary complex at a distance of 4.3 Å from the C1 atom (Figure 5). W7, present only in the ternary complex, is hydrogen-bonded to the sidechain of Arg12, while an equivalent water molecule is also present in the BsGlvA [15]. It is tempting to speculate that W7, activated by Arg12, plays a functional role in catalysis [44] and is an interesting feature that can explored using quantum mechanics/molecular mechanics (QM/MM) studies. The structural basis for the proposed mechanism is in good agreement with earlier structural and computational studies carried out for homologs [17–19,37,39].

A close examination of the interactions of the catalytic base Asp267 suggests that its protonation state is likely perturbed by an ionic interaction with the sidechain of the neighboring Arg270 (3.0 Å). The orientation of the Arg270 guanidinium is held in place by a stable bidentate salt bridge with Asp294, located on the opposite side (Figures 5 and 6C). It is likely that the interaction geometry of the Asp267–Arg270–Asp294 triad perturbs the pKa of Asp267. A perturbed pKa is consistent with the role of Asp267 in the deprotonation and

protonation of the C2 atom during each half cycle of the reaction. We next calculated the pKa values of Asp267 using the PROPKA program [45]. As expected, the predicted values of pKa show an increase from 2.4 in the apo form to 6.0 in the ternary complex. This triad (Asp–Arg--Asp) is also present in the α-glucosidase TmAglA, where the structurally equivalent Asp260 is proposed to act as the base [14]. Aspartates with perturbed pKa as high as 8.2 have been reported earlier in human aromatase enzyme active site due to interactions with nearby residues and ligands [46]. Interestingly, both R270K and R270A mutants are inactive (Supplementary Figure S7). It is likely that Lys at this position is unable to maintain the orientation or the appropriate protonation state of the catalytic Asp267. Structural comparisons indicate that this triad region is divergent in conformation and sequence compared with the phosphoglucosidases (Supplementary Figure S9) [18,37]. It thus appears that in the GH4 family, the pKa of the catalytic base is fine-tuned by its interactions with residues in the neighborhood. This hypothesis is consistent with the structural diversity of this region across homologs with different specificities. Lastly, the glycone binding site maintains an exquisite complementarity to the sugar acid. This explains why most mutational modifications to the existing active site geometry leads to catalytic failure (Supplementary Figure S7). Given this limitation, the examination of the precise role of specific residues in TmAgu4B will necessitate extensive studies using KIE and NMR techniques.

Structural basis of glucuronide specificity in TmAgu4B

Next, we sought to understand the strict glucuronidase specificity of TmAgu4B using a comparative analysis of the available three ternary complexes. Among the characterized homologs, TmAgu4B is structurally the most similar to the α-glucosidase TmAglA (rmsd of 1.2 Å over 451 Cα atoms), while the rmsd with the two phosphoglucosidases, TmBglT and BsGlvA, are 2.6 and 2.8 Å (over 461 and 469 Cα atoms), respectively. Notably, the carboxylic moiety of GlcA in TmAgu4B makes ionic interactions with three Arg residues (Arg12, Arg270 and Arg299), constituting a significantly positive, well-packed pocket that confers specificity towards glucuronide substrates (Figure 6C). A perfect bidentate interaction is formed between the carboxylic O6A and O6B atoms with Arg299 (∼2.6 Å). Arg270 makes another bidentate interaction with O5 and O6B atoms (∼3.0 Å), while Arg12 interacts with O6A and O1N of NADH (Figure 5, Supplementary Figure S13).

Structural comparisons show that the four phosphoglucosidases, TmBglT, BsGlvA, KpAglB and GsLicH, contain a conserved Arg equivalent to Arg299 (Arg285 in BsGlvA). In contrast, in the α-glucosidase TmAglA, Arg299 is substituted by Trp293 (Supplementary Figure S13A–D). Presumably, a positively charged residue is imperative at this position to recognize a generic negatively charged moiety in the substrate glycone. Given that the TmAgu4B mutants R299A, R299F and R299W displayed no activity, perhaps the ionic interactions of Arg299 (Trp in glucosidase) is important both for optimal orientation and affinity of the negatively charged sugar (Supplementary Figures S7 and S13A,B). In fact, the specificity for the substrate in the phosphoglucosidases appears to be dictated by its interactions with another conserved Arg (Arg95 in BsGlvA, Arg87 in TmBglT). This Arg is replaced by Trp99/Leu93 in TmAgu4B and TmAglA, consistent with their non-phosphorylated substrates (Supplementary Figure S13B,C).

Thus, specificity in TmAgu4B is likely determined by interactions of the carboxylic atoms with either or both Arg12 and Arg270, positioned on opposite sides (Figures 5 and 6C). Arg270 is located in a divergent loop region that connects helices α9 and α10 in the GH4 family (Supplementary Figure S9). This region in the phosphoglucosidases lacks an Arg at the corresponding position, whereas Arg is conserved in the glucosidase (Arg263, TmAglA) and makes a bidentate interaction with the O6H and O5 atoms of bound maltose, identical with the Arg210–GlcA interactions in TmAgu4B (Supplementary Figure S13A). This suggests that Arg270 cannot serve to discriminate between the glucuronide and glucoside specificities in the TmAgu4B and TmAglA, respectively. Arg12, present in the flexible and dynamic Turn I, is unique to TmAgu4B and is substituted by Val15 in TmAglA (Phe17 in BsGlvA). Moreover, Turn I region is divergent across the family, and presents an open conformation in TmAglA, oriented away from the substrate (Supplementary Figure S13A). These observations strongly suggest that the ionic interactions of Arg12 largely dictates the strict specificity for glucuronides in TmAug4B. It is noteworthy that the role of Arg12 in this context is directly enforced by a substrate-induced coupled swing movement of residues Phe13 and Arg12. Moreover, the substitution mutants R12A and R12K are both catalytically inactive. Considering the network of direct and water-mediated interactions of Arg12 with the substrate and cofactor, this residue has overlapping roles in substrate binding, specificity and catalysis.

Another residue that may dictates the specificity of TmAgu4B is Thr125. The equivalent residue, Asp119 in TmAglA, interacts the glycone (Supplementary Figure S13A). Docking of GlcA into the TmAglA active site

creates an unfavorable electrostatic environment with the glycone carboxylic group, whereas in TmAgu4B, Thr125 makes a favorable water-mediated interaction. Indeed, a T125D mutation in TmAgu4B abolished activity (Supplementary Figure S7). Notwithstanding the close structural and sequence similarity of the glycone binding pockets of TmAgu4B and TmAglA, the evolution of exquisite substrate discrimination between the two homologs is governed by multiple adaptations of the pocket in the immediate neighborhood of the C6 position.

Conclusions

The structural basis of the mechanism and specificity of TmAgu4B studied here reveals unique stepwise sequential conformational rearrangements, previously unreported for this family. Conformational changes associated with binding the cofactor to the apo form are first necessary to create a competent active site to accommodate the substrate. Next, the binding of the substrate induces a subsequent secondary structure transition resulting in a swing movement of key residues ensuring that both the nicotinamide moiety and the substrate are oriented in an alignment required for hydride transfer. The ternary complex also illustrates the direct roles of crucial catalytic residues, including the base Asp267, the protonation state of which is modulated through interactions within a triad of charged residues. Lastly, crucial insights are realized for the structural basis of substrate specificity in the GH4 family by defining clear roles for specific residues that discriminate between glucoside, glucuronide and phosphoglucoside substrates within the glycone recognition pockets of three paralogous members. The GH4 family is another example that demonstrates the exquisite fine-tuning of molecular features and dynamics that have driven the evolution of the enormous number of functionally diverse extant carbohydrate-active enzymes built around a few structural folds and catalytic frameworks.

Data Availability

The co-ordinates and structure factors presented in this paper have all been deposited in the Protein Data Bank (PDB) with the following codes: 7CTD, 7CTL, 7CTM. All remaining data are contained within the article.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

CRediT Contribution

Narayanan Manoj: Conceptualization, Supervision, Funding acquisition, Project administration, Writing — review and editing. **Samar Ballabha Mohapatra**: Resources, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing — original draft.

Abbreviations

GlcA, D-glucuronic acid; BsGlvA, *Bacillus subtilis* phospho-α-glucosidase; BsLplD, *Bacillus subtilis* glucosidase; G6P, Glucose-6-phosphate; GsLicH, *Geobacillus stearothermophilus* phospho-β-glucosidase; HOMO, highest occupied molecular orbital; LUMO, lowest unoccupied molecular orbital; KpAglB, Klebsiella pneumoniae 6-phospho-α-glucosidase; TmBglT, *Thermotoga maritima* phospho-β-glucosidase; TnAgl, *Thermotoga neapolitana*; TmAglA, *Thermotoga maritima* α-glucosidase; TmAgu4B, *Thermotoga maritima* α-glucuronidase.

References

1   Davies, G. and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853–859 https://doi.org/10.1016/S0969-2126(01)00220-9

2   Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 https://doi.org/10.1093/nar/gkt1178

3   Koshland, Jr, D.E. (1953) Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev.* **28**, 416–436 https://doi.org/10.1111/j.1469-185X.1953.tb01386.x

4   Zechel, D.L. and Withers, S.G. (2000) Glycosidase mechanisms : anatomy of a finely tuned catalyst. *Acc. Chem. Res.* **33**, 11–18 https://doi.org/10.1021/ar970172

5   Sinnott, M. (2007) *Carbohydrate Chemistry and Biochemistry*, The Royal Society of Chemistry, Cambridge, UK.

6   Thompson, J., Gentry-Weeks, C.R., Nguyen, N.Y., Folk, J.E. and Robrish, S.A. (1995) Purification from *Fusobacterium mortiferum* ATCC 25557 of a 6-phosphoryl-O-α-D-glucopyranosyl:6-phosphoglucohydrolase that hydrolyzes maltose 6-phosphate and related phospho-α-D-glucosides. *J. Bacteriol.* **177**, 2505–2512 https://doi.org/10.1128/JB.177.9.2505-2512.1995

7   Bibel, M., Brettl, C., Gosslar, U., Kriegshäuser, G. and Liebl, W. (1998) Isolation and analysis of genes for amylolytic enzymes of the hyperthermophilic bacterium *Thermotoga maritima*. *FEMS Microbiol. Lett.* **158**, 9–15 https://doi.org/10.1111/j.1574-6968.1998.tb12793.x

8   Liu, Q.P., Sulzenbacher, G., Yuan, H., Bennett, E.P., Pietz, G., Saunders, K., et al. (2007) Bacterial glycosidases for the production of universal red blood cells. *Nat. Biotechnol.* **25**, 454–464 https://doi.org/10.1038/nbt1298

9   Suresh, C., Kitaoka, M. and Hayashi, K. (2003) A thermostable non-xylanolytic alpha-glucuronidase of *Thermotoga maritima* MSB8. *Biosci. Biotechnol. Biochem.* **67**, 2359–2364 https://doi.org/10.1271/bbb.67.2359

10 Tanner, M.E. (2008) Transient oxidation as a mechanistic strategy in enzymatic catalysis. *Curr. Opin. Chem. Biol.* **12**, 532–538 https://doi.org/10.1016/j.cbpa.2008.06.016

11 Yip, V.L.Y. and Withers, S.G. (2006) Breakdown of oligosaccharides by the process of elimination. *Curr. Opin. Chem. Biol.* **10**, 147–155 https://doi.org/10.1016/j.cbpa.2006.02.005

12 Charnock, S.J., Brown, I.E. Turkenburg, J.P., Black, G.W. and Davies, G.J. (2002) Convergent evolution sheds light on the anti-β-elimination mechanism common to family 1 and 10 polysaccharide lyases. *Proc. Natl Acad. Sci. USA* **99**, 12067–12072 https://doi.org/10.1073/pnas.182431199

13 Major, L.L., Wolucka, B.A. and Naismith, J.H. (2005) Structure and function of GDP-mannose-3′,5′-epimerase: an enzyme which performs three chemical reactions at the same active site. *J. Am. Chem. Soc.* **127**, 18309–18320 https://doi.org/10.1021/ja056490i

14 Lodge, J.A., Maier, T., Liebl, W., Hoffmann, V. and Sträter, N. (2003) Crystal structure of *Thermotoga maritima* α-glucosidase AglA defines a new clan of NAD+-dependent glycosidases. *J. Biol. Chem.* **278**, 19151–19158 https://doi.org/10.1074/jbc.M211626200

15 Rajan, S.S., Yang, X., Collart, F., Yip, V.L.Y., Withers, S.G., Varrot, A. et al. (2004) Novel catalytic mechanism of glycoside hydrolysis based on the structure of an NAD+/Mn2+-dependent phospho-α-glucosidase from *Bacillus subtilis*. *Structure* **12**, 1619–1629 https://doi.org/10.1016/j.str.2004.06.020

16 Varrot, A., Yip, V.L.Y., Li, Y., Rajan, S.S., Yang, X., Anderson, W.F. et al. (2005) NAD+ and metal-ion dependent hydrolysis by family 4 glycosidases: Structural insight into specificity for phospho-β-D-glucosides. *J. Mol. Biol.* **346**, 423–435 https://doi.org/10.1016/j.jmb.2004.11.058

17 Chaklader, S., Cheng, L., Choi, M., Liu, J. and Bennet, A.J. (2011) Mechanistic evaluation of MelA α-galactosidase from *Citrobacter freundii*: a family 4 glycosyl hydrolase in which oxidation is rate-limiting. *Biochemistry* **50**, 4298–4308 https://doi.org/10.1021/bi101808h

18 Yip, V.L.Y., Thompson, J. and Withers, S.G. (2007) Mechanism of GlvA from *Bacillus subtilis*: a detailed kinetic analysis of a 6-phospho-α-glucosidase from glycoside hydrolase family 4. *Biochemistry* **46**, 9840–9852 https://doi.org/10.1021/bi700536p

19 Huang, W., Llano, J. and Gauld, J.W. (2010) Redox mechanism of glycosidic bond hydrolysis catalyzed by 6-phospho-α-glucosidase: A DFT study. *J. Phys. Chem. B* **114**, 11196–11206 https://doi.org/10.1021/jp102399h

20 Yip, V.L.Y. and Withers, S.G. (2006) Mechanistic analysis of the unusual redox-elimination sequence employed by *Thermotoga maritima* BglT : a 6-Phospho-β-glucosidase from glycoside hydrolase family 4. *Biochemistry* **45**, 571–580 https://doi.org/10.1021/bi052054x

21 Nagao, Y., Nakada, T., Imoto, M., Shimamoto, T., Sakai, S., Tsuda, M. et al. (1988) Purification and analysis of the structure of α-galactosidase from *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **151**, 236–241 https://doi.org/10.1016/0006-291X(88)90584-0

22 Thompson, J., Pikis, A., Rich, J., Hall, B.G. and Withers, S.G. (2013) α-Galacturonidase(s): a new class of family 4 glycoside hydrolases with strict specificity and a unique CHEV active site motif. *FEBS Lett.* **587**, 799–803 https://doi.org/10.1016/j.febslet.2013.02.004

23 Yun, B.Y., Jun, S.Y., Kim, N.A., Yoon, B.Y., Piao, S., Park, S.H. et al. (2011) Crystal structure and thermostability of a putative α-glucosidase from *Thermotoga neapolitana*. *Biochem. Biophys. Res. Commun.* **416**, 92–98 https://doi.org/10.1016/j.bbrc.2011.11.002

24 Nelson, K., Clayton, R., Gill, S., Gwinn, M.L., Dodson, R.J., Haft, D.H., et al. (1999) Evidence for lateral gene transfer between archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323–329 https://doi.org/10.1038/20601

25 Conners, S.B., Montero, C.I., Comfort, D.A., Shockley, K.R., Johnson, M.R., Chhabra, S.R. et al. (2005) An expression-driven approach to the prediction of carbohydrate transport and utilization regulons in the hyperthermophilic bacterium *Thermotoga maritima*. *J. Bacteriol.* **187**, 7267–7282 https://doi.org/10.1128/JB.187.21.7267-7282.2005

26 Mohapatra, S.B. and Manoj, N. (2019) Structure of an α-glucuronidase in complex with Co2+ and citrate provides insights into the mechanism and substrate recognition in the family 4 glycosyl hydrolases. *Biochem. Biophys. Res. Commun.* **518**, 197–203 https://doi.org/10.1016/j.bbrc.2019.08.030

27 Palmer, T. and Bonner, P.L.R. (2007) *Enzymes : Biochemistry, Biotechnology and Clinical Chemistry*, Woodhead Publishing Limited, Cambridge, UK.

28 Hernández, A. and Ruiz, M.T. (1998) An EXCEL template for calculation of enzyme kinetic parameters by non-linear regression. *Bioinformatics* **14**, 227–228 https://doi.org/10.1093/bioinformatics/14.2.227

29 Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R. and Leslie, A.G.W. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 271–281 https://doi.org/10.1107/S0907444910048675

30 Kabsch, W. (1988) Automatic indexing of rotation diffraction patterns. *J. Appl. Crystallogr.* **21**, 67–72 https://doi.org/10.1107/S0021889887009737

31 Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242 https://doi.org/10.1107/S0907444910045749

32 Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M. et al. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D. Biol. Crystallogr.* **68**, 352–367 https://doi.org/10.1107/S0907444912001308

33 Moriarty, N.W., Grosse-Kunstleve, R.W. and Adams, P.D. (2009) Electronic ligand builder and optimization workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **65**, 1074–1080 https://doi.org/10.1107/S0907444909029436

34 Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr.* **66**, 486–501 https://doi.org/10.1107/S0907444910007493

35 The PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC

36 Mohapatra, S.B. and Manoj, N. (2021) A conserved π-helix plays a key role in thermoadaptation of catalysis in the glycoside hydrolase family 4. *Biochem. Biophysics. Aca - Proteins Proteom.* **1869**, 140523 https://doi.org/10.1016/j.bbapap.2020.140523

37   Yip, V.L.Y. and Withers, S.G. (2012) Identification of Tyr241 as a key catalytic base in the family 4 glycoside hydrolase BglT from *thermotoga maritima*. *Biochemistry* **51**, 8464–8474 https://doi.org/10.1021/bi301021u

38   Holm, L. (2019) Benchmarking fold detection by daliLite v.5. *Bioinformatics* **35**, 5326–5327 https://doi.org/10.1093/bioinformatics/btz536

39   Yip, V.L.Y., Varrot, A., Davies, G.J., Rajan, S.S., Yang, X., Thompson, J. et al. (2004) 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 https://doi.org/10.1021/ja047632w

40   Wu, Y. (1995) Transition structures of hydride transfer reactions of protonated pyridinium ion with 1,4-dihydropyridine and protonated nicotinamide with 1,4-dihydronicotinamide. *J. Am. Chem. Soc.* **117**, 4100–4108 https://doi.org/10.1021/ja00119a026

41   Sannikova, N., Gerak, C.A.N., Shidmoossavee, F.S., King, D.T., Shamsi, S., Abadi, K. et al. (2018) Both chemical and non-chemical steps limit the catalytic efficiency of family 4 glycoside hydrolases. *Biochemistry* **57**, 3378–3386 https://doi.org/10.1021/acs.biochem.8b00117

42   Yun Yang, B. and Montgomery, R. (2001) β-Elimination of glucosyluronic residues during methylation of an acidic polysaccharide from *Erwinia chrysanthemi CU 643. Carbohydr. Res.* **332**, 317–323 https://doi.org/10.1016/S0008-6215(01)00062-3

43   Kiss, J. (1974) β-Eliminative degradation of carbohydrates containing uronic acid residues. *Adv. Carbohydr. Chem. Biochem.* **29**, 229–303 https://doi.org/10.1016/S0065-2318(08)60251-6

44   Fitch, C.A., Platzer, G., Okon, M., Garcia-Moreno, B.E. and McIntosh, L.P. (2015) Arginine: its pKa value revisited. *Protein Sci.* **24**, 752–761 https://doi.org/10.1002/pro.2647

45   Dolinsky, T.J., Nielsen, J.E., McCammon, J.A. and Baker, N.A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, 665–667 https://doi.org/10.1093/nar/gkh381

46   Di Nardo, G., Breitner, M., Bandino, A., Ghosh, D., Jennings, G.K., Hackett, J.C. et al. (2015) Evidence for an elevated aspartate pKa in the active site of human aromatase. *J. Biol. Chem.* **290**, 1186–1196 https://doi.org/10.1074/jbc.M114.595108