

# ***Rewiring Catalytic Craters: A Path for Engineering $\beta$ -Glucosidases to Improve Glucose Tolerance***

Abhishek B. Suryawanshi<sup>1</sup>, Rajiv K. Bedi<sup>1</sup>, Meera Pawar<sup>1</sup>, Santosh Noronha<sup>2</sup>, Prasenjit Bhaumik<sup>1\*</sup>

<sup>1</sup>Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai - 400076, Maharashtra, India

<sup>2</sup>Department of Chemical Engineering, Indian Institute of Technology Bombay, Powai, Mumbai - 400076, Maharashtra, India

\*To whom correspondence should be addressed:

Prasenjit Bhaumik, Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai-400076, India; Telephone: +91-22-2576-7748; E-mail: pbhaumik@iitb.ac.in

**Keywords:**  $\beta$ -glucosidase, feedback inhibition, glucose tolerance, glucose stimulation, hydrolysis mechanism, crystal structure

## Abstract

Bioethanol, a sustainable alternative to fossil fuels, is produced from cellulose via enzymatic saccharification.  $\beta$ -Glucosidase, a key enzyme in this process, hydrolyses disaccharides into glucose but is limited by feedback inhibition. This study investigated GH1  $\beta$ -glucosidase from a soil metagenome (UnBG11) to understand and overcome this limitation. We solved the near-atomic resolution crystal structure of UnBG11 at 1.15 Å as an apo enzyme and report the first high-resolution crystal structures of the enzyme in its pre-hydrolytic state as a cellobiose complex and covalent intermediate-bound state, capturing key stages of its catalytic mechanism. Structural analysis revealed three crucial glucose-binding subsites in UnBG11's crater. Glucose binding at the -1 subsite induced a 1.4 Å shift in E370, increasing the distance between catalytic glutamates and reducing enzymatic activity. The C188V variant, generated by rational engineering, significantly improved glucose tolerance to 2.5 M. Additionally, the H261W mutation at the +2 subsite enhanced kinetic properties by improving cellobiose affinity ( $K_m = 22.87 \pm 1.1$  mM) and shifted the optimal pH to 5.5 from 6.0. Comparative structural analysis with other glucose-tolerant GH1  $\beta$ -glucosidases revealed conserved residues at the -1 subsite crucial for substrate stabilization and +1 subsite residues interacting with glucose, offering targets for further optimization. Engineered UnBG11 variants retained high stability and activity on sugarcane bagasse, demonstrating their potential for industrial cellulase cocktails. These findings provide a robust framework for engineering  $\beta$ -glucosidases with enhanced glucose tolerance and catalytic efficiency, paving the way for improved bioethanol production and contributing to sustainable energy solutions.

## Introduction

The dual threat of climate change and finite fossil fuel reserves have led us to search and development of environment-friendly, renewable, carbon-neutral energy solutions. Bioethanol is one such energy source, which is derived from plant-based lignocellulose. The main benefit of employing second-generation bioethanol is that it requires the lignocellulosic biomass as primary resource, which avoids interference with food crops and can readily be grown on poor-quality marginal land with little water and fertilizers<sup>1</sup>. However, the presence of lignin and hemicellulose can hinder the process of bioethanol production. Therefore, the separation of lignin and hemicellulose from the lignocellulose is done by several physical, physicochemical, chemical and biological approaches to obtain cellulose<sup>2</sup>. Cellulose is the most abundant and renewable carbohydrate bio-polymer comprising of  $\beta$ -1,4-linked glucose monomers<sup>3</sup>. Hydrolysis of cellulose by cellulases produces glucose to produce ethanol. Cellulases are enzymes complexes, including endo-glucanases, cellobiohydrolases, exo-glucanases and  $\beta$ -glucosidases. These enzymes synergistically hydrolyse cellulose into fermentable sugars<sup>4-6</sup>. Endo-glucanase acts randomly on the amorphous area of the cellulose and generates various lengths of oligosaccharides; cellobiohydrolases and exo-glucanases catalyze the release of glucose or cellobiose from the reducing and non-reducing end of polysaccharide and finally  $\beta$ -glucosidase hydrolyses and produces glucose from released cellobiose<sup>5,7</sup>.

Cellobiose acts as an inhibitor for endo-glucanase and exo-glucanase<sup>4</sup>.  $\beta$ -glucosidase, the final enzyme in the saccharification process, catalyses the hydrolysis of cellobiose and other smaller oligosaccharides. Even if  $\beta$ -glucosidase can reduce the inhibitory effect of cellobiose, it produces glucose, which at higher concentrations acts as an inhibitor for  $\beta$ -glucosidase<sup>8,9</sup>. Increased concentration of the sugars produced during the saccharification process in the reactors is responsible for the increase in viscosity and raises the difficulty of the product diffusion out of the active site of  $\beta$ -glucosidase<sup>10</sup>. At high concentrations, glucose can either block the  $\beta$ -glucosidase active site for substrate entry or prevent product exit<sup>11,12</sup>. Moreover, it is worth mentioning that there is a correlation between glucose tolerance and a stimulating effect triggered by glucose in  $\beta$ -glucosidases. This correlation only applies to some members of the glycoside hydrolase family 1 (GH1)  $\beta$ -glucosidases<sup>12</sup>. Stimulation effect in  $\beta$ -glucosidases can be observed due to the binding of glucose to the secondary site, whereas the enzyme gets inhibited when glucose binds and blocks the active site<sup>13</sup>. A deeper understanding of feedback inhibition of  $\beta$ -glucosidases by glucose is highly essential and would contribute to the development of efficient enzymes for the industrial saccharification process.

Bioethanol producing industries require the  $\beta$ -glucosidases with higher activity and lower  $K_m$  for saccharification to reduce the burden of inhibition exerted by cellobiose<sup>14</sup> as insufficient activity leads to build-up of cellobiose during cellulose hydrolysis<sup>15,16</sup>. One of the strategies used by industry to avoid feedback inhibition is the implication of simultaneous saccharification and fermentation (SSF) method for bioethanol production where both saccharification and fermentation are performed in a single reactor. Hence, the released glucose from the saccharification process is used immediately for bioethanol production<sup>17</sup>. Generally, SSF occurs at lower temperatures as organisms like yeast that ferment sugar prefers lower temperatures. Therefore, it is essential to develop enzymes that do not undergo product inhibition and have higher activity and stability at lower temperatures. Until now, no proper comprehensive studies have been conducted that reveal the molecular basis of glucose tolerance in  $\beta$ -glucosidase<sup>18</sup>. Here we present extensive structural and biochemical studies to understand the mechanistic details of product inhibition and designing strategies for improving the glucose tolerance of GH1  $\beta$ -glucosidases.

$\beta$ -glucosidase (EC 3.2.1.21) is an enzyme that hydrolyses the compounds with  $\beta$ -D-glycosidic linkage<sup>18,19</sup>. Several studies have reported functional characterizations of this enzyme due to its universal distribution and well-defined variety of substrates<sup>20</sup>.  $\beta$ -glucosidase has been classified into different glycoside hydrolase families: GH1, GH3, GH5, GH9 and GH30. Among them, the GH1 family contains the most significant number of characterized  $\beta$ -glucosidases with the classic  $(\alpha/\beta)_8$  TIM-barrel structure<sup>21,22</sup>. GH1 family  $\beta$ -glucosidases are enzymes which catalyse hydrolysis of  $\beta$ -D-glycosidic bonds by retaining a double displacement mechanism<sup>23</sup>. The conserved carboxylic acid residues present in the  $\beta$ -strands 4 and 7 of the enzyme serve as both the general acid-base catalyst and the nucleophile during the catalytic reaction<sup>24</sup>. The catalytic acid/base is located on the TXNEP motif at the end of strand 4 (where X is a hydrophobic amino acid residue), and the catalytic nucleophile is located on the I/VTENG motif at the end of strand 7<sup>25</sup>. The catalysis of  $\beta$ -glucosidase has been studied by many methods, including pH-activity analysis, mutagenesis, suicide inactivator, isotope effects, interacting studies with alternative substrates or inhibitors and 3-Dimensional (3D) structural visualization. However, the coordinated conformational changes enabling substrate distortion and glycosidic bond cleavage have remained primarily obscured over decades of study<sup>8,26</sup>. Extensive structural understanding of substrate hydrolysis is still poorly understood; therefore, structural studies with native GH1 enzymes are necessary.

In this study, we have performed biochemical characterizations of a GH1 enzyme, UnBG11, a  $\beta$ -glucosidase identified from soil metagenome<sup>27</sup>; and solved multiple high-



resolution crystal structures of native and variants of this enzyme. Further, through our crystallographic studies, we have also demonstrated three stages of cellobiose hydrolysis catalytic mechanism showing the substrate (cellobiose), covalently linked glycosyl reaction intermediate and product (glucose) binding at the catalytic centre of UnBG11. Our study is the first report that captured the covalently bound reaction intermediate to the catalytic site of a GH1  $\beta$ -glucosidase. Apo, ligand (glucose, cellobiose, thio-glucose) complexed and glucose-bound covalent intermediate structures aided in identifying the conserved product binding sites in UnBG11. Structural analysis has provided insights into understanding the molecular basis to increase glucose tolerance and improve the kinetic properties of UnBG11. These findings can be applied to other GH1 family  $\beta$ -glucosidase to enhance their glucose tolerance.

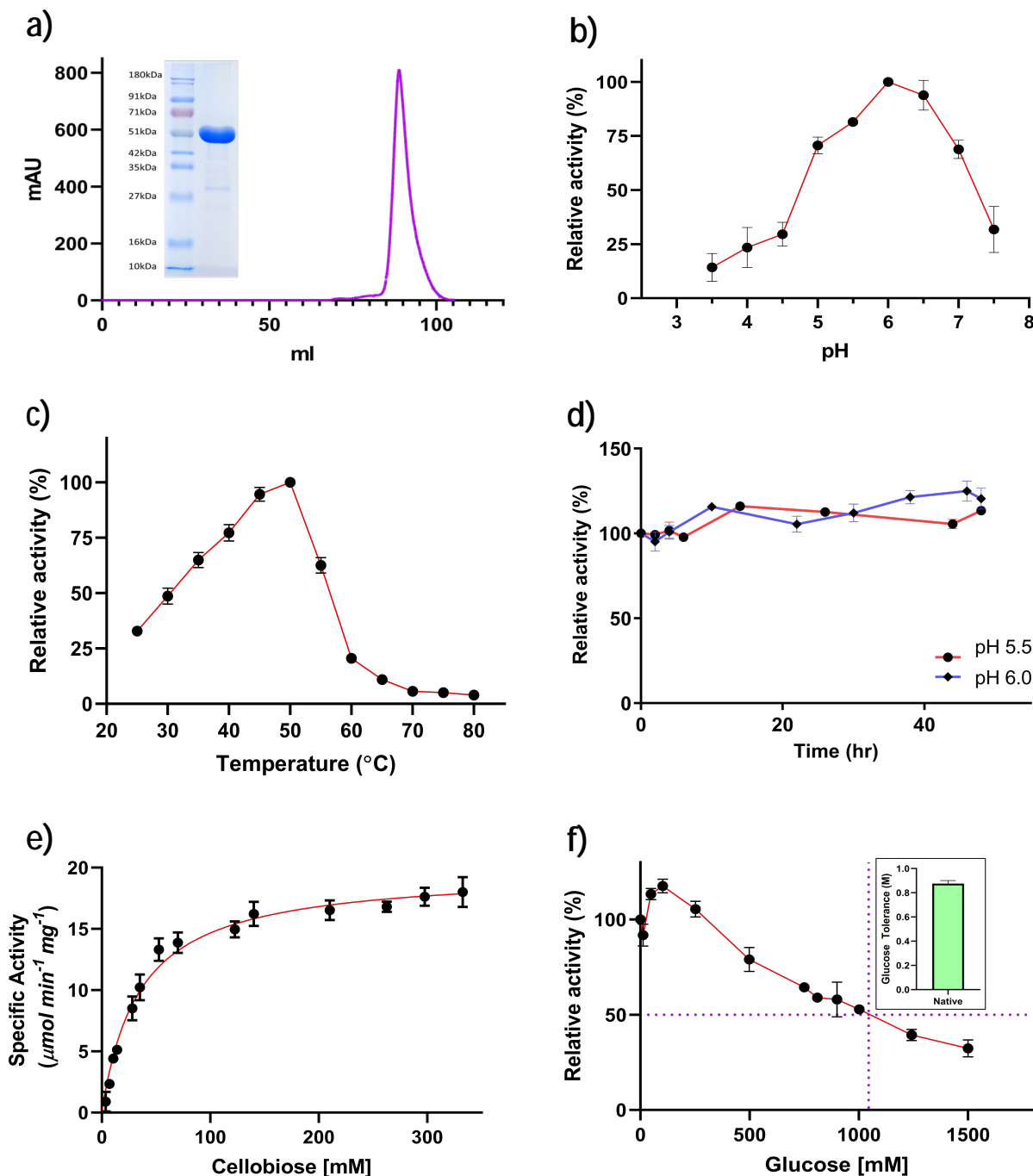
## Results

### Kinetic and biochemical properties of native UnBG11

The C-terminal His<sub>6</sub> tagged recombinant UnBG11 was overexpressed as a soluble protein by induction of 0.4 mM IPTG at 24 °C for 16 hrs. After the purification by Ni-NTA affinity and size exclusion chromatography, highly pure UnBG11 was obtained. The chromatogram (Figure 1a) showed the single elution peak, indicating the purity of the protein, which was confirmed by SDS-PAGE. Highly purified UnBG11 was further used for biochemical and kinetic characterizations and crystallization.

We performed 4-nitrophenyl β-D-glucopyranoside (pNPG) and cellobiose hydrolysis reactions at various pHs and temperatures to determine the best conditions for the catalytic reaction by UnBG11. Native UnBG11 activity was investigated throughout a pH range of 3.5 - 8.0, with reactions occurring at 55 °C for 10 minutes. At pH 6.0, the highest activity was observed for pNPG (Supplementary Figure S1a) and cellobiose hydrolysis (Figure 1b). In a pH range of 5.5 - 7.0, the enzyme showed more than 70 % of its highest activity. Below pH 5.0 and above pH 7.5, activity was less than 50 % of its maximum value. Native UnBG11 showed increased activity in the temperature range of 45 °C - 55 °C, with maximum activity at 55 °C for pNPG hydrolysis (Supplementary Figure S1b) and 50 °C for cellobiose hydrolysis (Figure 1c). The enzyme activity dropped to less than 30 % at temperatures above 55 °C. Purified native UnBG11 is stable at pH 5.5 and pH 6.0 for 48 hrs. and retains almost 100% of its activity after 48 hrs. (Figure 1d).

The kinetic properties of native UnBG11 were initially determined with its artificial substrate pNPG, which, upon hydrolysis, forms p-nitrophenol (pNP), which is a yellow colour in the alkaline pH. The substrate saturation assays were performed to determine the kinetic parameters ( $K_m$ ,  $V_{max}$ ) of UnBG11. The native enzyme followed Michaelis-Menten kinetics, with  $V_{max}$  value of  $36.14 \pm 0.50 \mu\text{mol min}^{-1}\text{mg}^{-1}$ ,  $K_m$  value of  $0.92 \pm 0.05 \text{ mM}$  and catalytic efficiency ( $k_{cat}/K_m$ ) value of  $34.36 \pm 11 \text{ mM}^{-1} \text{ s}^{-1}$  (Supplementary Figure S1c). We also determined the kinetic parameters of native UnBG11 for hydrolysis of cellobiose, a natural substrate of β-glucosidase. For cellobiose hydrolysis, UnBG11 attained the  $V_{max}$  of  $19.74 \pm 0.37 \mu\text{mol min}^{-1}\text{mg}^{-1}$  with  $K_m$  of  $34.61 \pm 2.4 \text{ mM}$  and  $k_{cat}/K_m$  of  $0.47 \pm 0.171 \text{ mM}^{-1} \text{ s}^{-1}$  (Figure 1e). The glucose tolerance level of the native UnBG11 was determined. It was observed that native UnBG11 loses its 50 % activity in the presence of around 1.0 M (Figure 1f) glucose concentration.



**Figure 1: Biochemical properties of native UnBG11.** **a)** Size exclusion chromatogram (purple colour line) indicating the elution of purified native UnBG11 from Superdex 200 16/600 column. Single band near 51 kDa on SDS-PAGE (inset) represents the purity of the enzyme. **b, c)** Graphs illustrate the optimum pH of 6.0 and temperature of 50 °C for cellobiose hydrolysis, respectively. Data points are shown in black solid circles and connected with the red line. **d)** stability of native UnBG11 at pH 5.5 (red) and pH 6.0 (purple) at 30°C. **e)** Kinetics profile of native UnBG11 showing the substrate saturation for cellobiose hydrolysis. **f)** The graph depicts UnBG11 glucose tolerance level expressed as concentration of glucose that causes loss of 50 % enzyme activity. Inset is a bar graph representation of glucose tolerance. All experiments were performed in triplicates ( $n = 3$ ) with error bar representing  $\pm$  SEM.

## Substrate, product and reaction intermediate binding in the native UnBG11

The near atomic-resolution crystal structure of the native UnBG11 was solved as apo-enzyme at 1.15 Å (PDB ID. 9JLZ) (Table 1: part I). The structure has been well refined with an acceptable range of  $R_{\text{work}}$  and  $R_{\text{free}}$ . This 469 residue-containing enzyme has a spherical shape. The overall dimension of UnBG11 structure is approximately  $57.5 \times 50.5 \times 68.3$  Å. This enzyme belongs to the GH1 family and has a well-characterized typical  $(\alpha/\beta)_8$  TIM-barrel fold (Figure 2a, Supplementary movie S1). The structure of UnBG11 was compared with the structure of its close well-characterized homologous (50.8 % sequence identity) (Supplementary Figure S2a)  $\beta$ -glucosidase from *Humicola insolens* (HiBG) (PDB ID. 4MDO)<sup>13</sup> and root mean square deviation (r.m.s.d.) value was 0.695 Å. (Supplementary Figure S2b). Structural alignment of these two enzymes helped to identify the catalytic crater and some gatekeeper residues of UnBG11. E185 (acid-base) and E370 (nucleophile) are the catalytic residues (Figure 2b). H140, N184, and W417 are the supporting active site residues, and E424, W343, F313, C188, L187, Y312, I192, and H261 (Supplementary Figure S2c) are residues that form the crater of the UnBG11. The crater of apo UnBG11 is filled with water molecules. In the catalytic crater of the apo enzyme structure, we also observed two glycerol molecules which might have appeared from the cryo-protectant solution (Supplementary Figure S2d).

**Table 1, part I: Data collection and refinement statistics of UnBGI1 variants.**

<b>Data collection statistics<sup>a</sup></b>	<b>UnBGI1 Apo</b>	<b>UnBGI1 Glucose</b>	<b>UnBGI1 Cellobiose</b>	<b>UnBGI1 Covalent Intermediate</b>	<b>UnBGI1 Thio-glucose</b>	<b>UnBGI1 C188V Apo</b>
Wavelength (Å)	0.97893	1.5418	1.5418	1.5418	1.5418	0.97949
Space group	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Unit cell constants						
<i>a</i> , <i>b</i> , <i>c</i> (Å)	51.84, 65.99, 138.18	51.91, 66.18, 138.17	51.47, 66.51, 138.20	51.20, 65.80, 137.90	51.15, 65.71, 137.66	51.88, 66.27, 138.25
$\alpha$ , $\beta$ , $\gamma$ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	40.0-1.15 (1.25-1.15)	40.0-1.73 (1.80-1.73)	40.0-1.73 (1.8-1.73)	30.0-2.0 (2.1-2.0)	40.0-1.73 (1.83-1.73)	30.0-1.3 (1.4-1.3)
No. of reflections	1161293 (217836)	349373 (32373)	343914 (30879)	225378 (29989)	325795 (42715)	549033 (97401)
Unique reflections	167030 (35537)	50247 (5305)	50020 (5356)	31988 (4234)	49212 (7451)	116123 (23019)
Completeness (%)	99.1 (96.3)	99.4 (94.9)	99.3 (96.2)	99.1 (98.1)	99.9 (99.2)	98.6 (99.2)
Multiplicity	6.95 (6.12)	6.95 (6.10)	6.87 (5.76)	7.51 (7.08)	6.62 (5.73)	4.72 (4.2)
<i>R</i> <sub>merge</sub> (%)	7.4 (55.4)	2.5 (7.0)	4.8 (45.8)	30.3 (90.3)	3.5 (16.7)	5.1 (43.9)
Mean <i>I</i> / $\sigma$ < <i>I</i> >	13.61 (3.17)	56.01 (23.33)	22.91 (3.5)	9.54 (2.80)	35.02 (9.99)	16.62 (3.4)
CC <sub>1/2</sub> (%)	99.9 (85.2)	100 (99.7)	99.9 (90.9)	95.7 (76.8)	100.0 (98.1)	99.9 (88.0)
<b>Refinement statistics</b>						
Resolution (Å)	39.10-1.15	37.8-1.73	37.87-1.73	29.69-2.0	38.73-1.73	28.76-1.30
Working set: no. of reflections	158699	47733	47521	30390	46751	110314
<i>R</i> <sub>factor</sub> (%)	12.58	12.82	16.36	18.56	14.69	16.80
Test set: no. of reflections	8353	2513	2502	1600	2461	5806
<i>R</i> <sub>free</sub> (%)	14.70	15.57	20.04	23.15	18.06	18.77
<b>Number of atoms</b>						
Protein atoms	3728	3638	3625	3584	3601	3660
Water molecules	636	558	381	224	524	593
<b>Ligands</b>						
Glucose	0	3	1	3	1	0
Cellobiose	0	0	1	0	0	0
Bond lengths (Å)	0.012	0.012	0.008	0.006	0.011	0.012
Bond angles (°)	1.849	1.903	1.596	1.541	1.881	1.914
<b>Ramachandran Plot<sup>c</sup> (%)</b>						
Favoured region	96.9	96.9	96.5	96.9	97.1	96.9
Allowed region	3.1	3.1	3.5	3.1	2.9	3.1
Disallowed region	0	0	0	0	0	0
PDB code	<b>9JLZ</b>	<b>9J3W</b>	<b>9J3Y</b>	<b>9J4Q</b>	<b>9J4B</b>	<b>9J4I</b>

<sup>a</sup>Values in parentheses indicate statistics in highest resolution shell.

<sup>b</sup> $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$  where *I* is the observed integrated intensity,  $\langle I \rangle$  is the average integrated intensity obtained from multiple measurements, and the summation is overall observed reflections.

<sup>c</sup>Values are obtained from PROCHECK analysis.

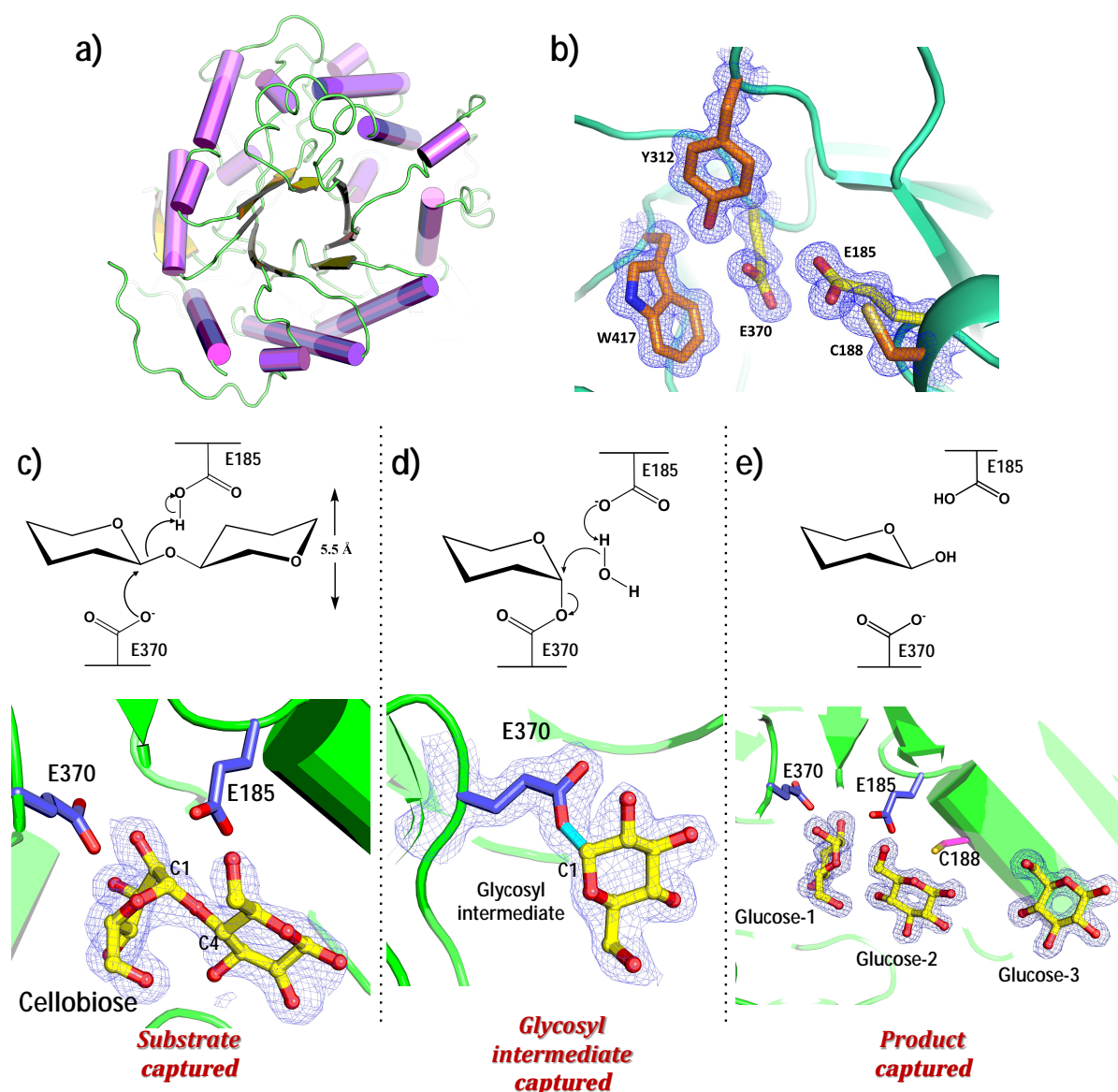
**Table 1, part II: Data collection and refinement statistics of UnBGI1 variants.**

<i>Data collection statistics<sup>a</sup></i>	UnBGI1 C188V Glucose	UnBGI1 H261W Apo	UnBGI1 H261W Glucose	UnBGI1 H261W Cellobiose	UnBGI1 C188V_ H261W Apo	UnBGI1 C188V_ H261W Glucose
Wavelength (Å)	1.5418	0.97777	1.5418	0.97777	0.97777	0.97777
Space group	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Unit cell constants						
<i>a</i> , <i>b</i> , <i>c</i> (Å)	52.11, 66.98, 138.64	51.66, 66.14, 137.33	51.47, 65.87, 137.35	50.53, 65.36, 136.15	51.75, 66.38, 137.49	52.00, 68.10, 136.00
$\alpha$ , $\beta$ , $\gamma$ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	40.0-1.75 (1.85-1.75)	40.0-1.8 (1.9-1.8)	65.87-1.78 (1.81-1.78)	40.58-1.70 (1.73-1.70)	30.0-1.55 (1.65-1.55)	30.0-2.0 (2.1-2.0)
No. of reflections	339608 (48264)	324512 (48061)	323344 (14972)	372926 (18249)	515477 (72060)	239076 (32459)
Unique reflections	48288 (7023)	44406 (6535)	45575 (2183)	50499 (2496)	69348 (11661)	33409 (4463)
Completeness (%)	96.9 (93.0)	99.9 (99.8)	100.0 (99.0)	100 (99.8)	99.7 (99.3)	99.9 (100)
Multiplicity	7.03 (6.87)	6.75 (7.35)	7.09 (6.85)	7.4 (7.3)	7.43 (6.17)	7.1 (7.2)
<i>R</i> <sub>merge</sub> (%)	7.1 (28.7)	24.9 (101.1)	23.9 (78.2)	14.7 (70.3)	10.0 (82.7)	22.6 (87.1)
Mean <i>I</i> / $\sigma$ < <i>I</i> >	19.31 (6.64)	9.38 (2.93)	8.8 (2.7)	12.2 (2.3)	14.42 (2.45)	7.44 (2.30)
CC <sub>1/2</sub> (%)	99.8 (95.3)	99.2 (79.9)	98.7 (30.8)	99.3 (82.0)	99.8 (76.7)	99.0 (76.0)
Refinement statistics						
Resolution (Å)	38.04-1.75	39.03-1.8	59.34-1.78	40.58-1.70	29.89-1.55	28.49-2.0
Working set: no. of reflections	45873	42185	43195	47867	65880	31738
<i>R</i> <sub>factor</sub> (%)	15.37	13.96	17.4	14.16	13.88	16.17
Test set: no. of reflections	2415	2221	2309	2561	3468	1671
<i>R</i> <sub>free</sub> (%)	19.11	16.78	20.74	18.00	16.30	21.7
Number of atoms						
Protein atoms	3614	3634	3701	3662	3742	3580
Water molecules	574	518	41	410	523	338
Ligands						
Glucose	2	0	2	1	0	1
Cellobiose	0	0	0	2	0	0
Bond lengths (Å)	0.01	0.009	0.008	0.009	0.011	0.015
Bond angles (°)	1.602	1.58	1.505	1.652	1.679	2.078
Ramachandran Plot <sup>c</sup> (%)						
Favoured region	96.5	97.4	97.2	97.2	97.0	97.3
Allowed region	3.5	2.6	2.8	2.8	3.0	2.7
Disallowed region	0	0	0	0	0	0
PDB code	<b>9J42</b>	<b>9J43</b>	<b>9J44</b>	<b>9J45</b>	<b>9J49</b>	<b>9J4A</b>

<sup>a</sup>Values in parentheses indicate statistics in highest resolution shell.

<sup>b</sup> $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$  where *I* is the observed integrated intensity, <*I*> is the average integrated intensity obtained from multiple measurements, and the summation is overall observed reflections.

<sup>c</sup>Values are obtained from PROCHECK analysis.



**Figure 2: Structural fold, active site architecture and cellobiose hydrolysis mechanism of UnBG11.** **a)** Overall structure of UnBG11 with  $(\alpha/\beta)_8$  TIM-barrel fold is represented as cartoon. Helices are shown in purple whereas sheets are shown in yellow colour. **b)** The  $2F_o - F_c$  map (blue mesh) contoured at  $1\sigma$  level showing the fully satisfied electron density around active site residues. Catalytic glutamates are presented as yellow stick and other active site residues are represented as orange sticks.

**Mechanism of cellobiose hydrolysis captured in the active site of UnBG11.** The catalytic glutamates are shown in purple stick, cellobiose and glucose are represented as ball (red) and stick (yellow) model. **c)** The  $2F_o - F_c$  map contoured at  $1\sigma$  level representing the electron density (slate mesh) of cellobiose near the catalytic glutamates suggesting that substrate has been captured just before nucleophilic attack. **d)** The  $2F_o - F_c$  map contoured at  $1\sigma$  level illustrating the electron density (slate mesh) of glycosyl intermediate covalently linked to catalytic glutamate (E370). **e)** The  $2F_o - F_c$  map contoured at  $1\sigma$  level showing the electron density (slate mesh) of three glucose molecules at the active site crater of UnBG11.

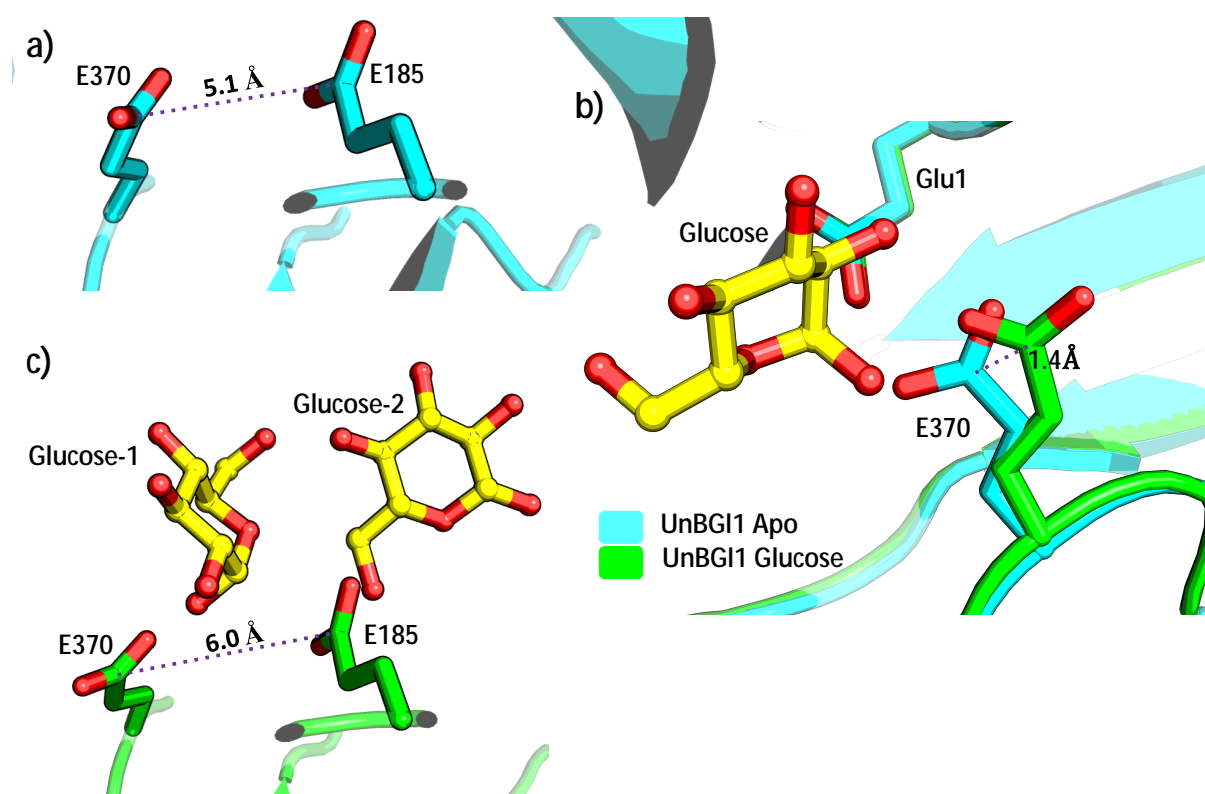
Understanding the hydrolysis mechanism of any enzyme in its native state is challenging. However, we aimed to visualize cellobiose hydrolysis by UnBG11 and understand the molecular details of substrate, intermediate and product binding in this GH1 enzyme active site. To demonstrate the enzymatic activity of UnBG11 in the crystal state, the crystals were

soaked with pNPG substrate. Within a short time of incubation, the surrounding solution started turning yellow (Supplementary movie S2) due to the formation of a yellow product pNP upon hydrolysis of pNPG by UnBG11 present in the crystal. These observations confirmed that UnBG11 was enzymatically active in the crystal form and could hydrolyse its substrate cellobiose to glucose. We performed various soaking experiments using native UnBG11 crystals with multiple concentrations of cellobiose and varied soaking periods to capture substrate, reaction intermediate and glucose in the enzyme's active site. Cellobiose complexed structure (PDB ID. 9J3Y) was obtained from a short soaking experiment of native UnBG11 crystals in cellobiose. Upon solving the structure, we found out that the UnBG11 active site has the substrate cellobiose (Supplementary Figures S3a, S3b and S3c) and the hydrolysis product glucose molecule (Supplementary Figures S3b and S3c). The cellobiose molecule is bound near the catalytic glutamates (E185, E370) (Figure 2c, Supplementary Figure S3a, Supplementary movie S3). At the catalytic crater, aromatic residues like W141, W343, F313, W417 and Y312 dominate the interaction with glucose units of the substrate. In the process of capturing the substrate-bound state of the UnBG11, we also captured one of the intermediate states of the enzyme reaction. We determined a crystal structure of UnBG11 complexed to the reaction intermediate at 2.0 Å resolution (PDB ID. 9J4Q) (Table 1: part I). In this structure, we observed that nucleophilic glutamate (E370) formed a covalent bond with the reaction intermediate, which was produced by breaking of  $\beta$ -1,4 glycosidic linkage (Figure 2d, Supplementary Figure S3d) of cellobiose by UnBG11. We could also see the product glucose in the crater and the reaction intermediate formation during the hydrolysis. Capturing the reaction intermediate in the native and active state of the enzyme through crystallography is one of the complex tasks. Until now, there is no report of capturing the glycosyl intermediate through crystallography using an active  $\beta$ -glucosidase. Our study is the first report to present a high-resolution structure of a covalently bound glycosyl intermediate formed by the action of active native UnBG11 on cellobiose.

$\beta$ -glucosidases get inhibited by its product glucose due to the feedback inhibition. We solved the 1.73 Å resolution UnBG11 glucose complexed structure (PDB ID. 9J3W) (Table 1: part I) that provided insights into crucial residues which facilitate feedback inhibition by the product. Three glucose molecules were discovered during the refinement of the structure (Figure 2e, Supplementary Figure S3e). The glucose bound and cellobiose bound structures did not show any significant difference to the apo structure of native UnBG11 and have root mean square deviation (r.m.s.d.) of 0.136 Å and 0.148 Å, respectively (Supplementary Figure S3f), based on the overall structural superposition.



For the GH1 family  $\beta$ -glucosidases that follow a retaining reaction mechanism, the typical separation (measured by the distance between CD atoms) of the two carboxylate groups of the catalytic glutamates is  $\sim 5.5$  Å<sup>23,28</sup>. In the apo UnBG11 structure, this distance is 5.1 Å (Figure 3a). However, at high glucose concentrations, the sugar molecules bind to all the conserved glucose binding sites. Notably, the carboxylate group of the nucleophilic glutamate (E370) showed 1.4 Å displacement from its original position that is observed for apo or cellobiose bound UnBG11 (Figure 3b). In the glucose complexed UnBG11 structure, it is observed that the nucleophilic glutamate (E370) side chain is a bit displaced from its position, and the distance between two catalytic glutamates (E185, E370) increased to 6.0 Å (Figure 3c). Such wide distance between two catalytic glutamates is not optimal to facilitate catalysis. The glucose molecules are stabilized by multiple interactions in the active site crater of UnBG11.



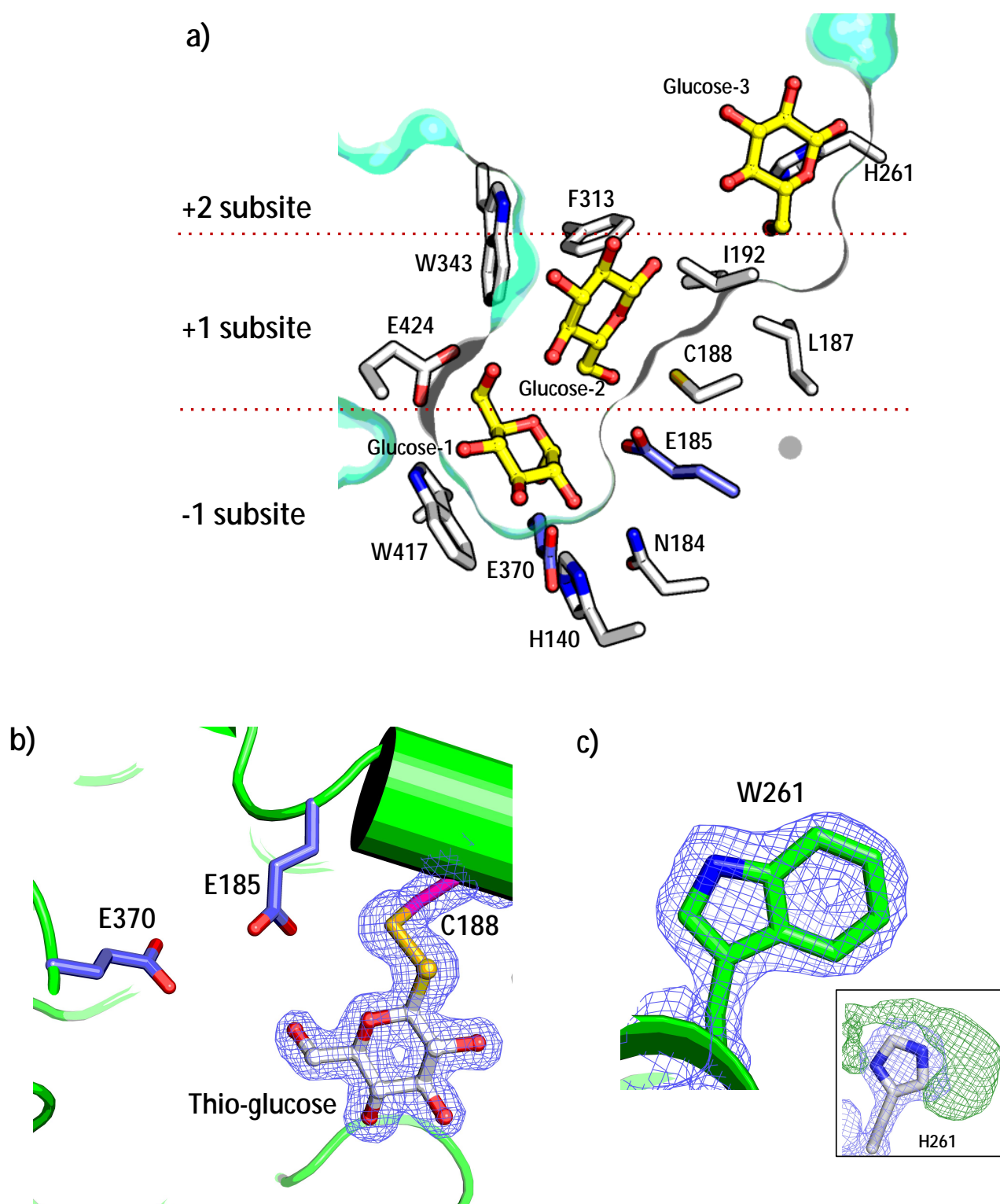
**Figure 3: Structural difference at the active site of apo and glucose bound native UnBG11.** The structure of native apo UnBG11 (cyan) is compared with the glucose complexed structure (green). **a)** Distance between the CD atoms of catalytic glutamates of apo UnBG11 is shown in dotted line. **b)** Catalytic glutamate (E370) showing 1.4 Å displacement from its original place due to binding of glucose. **c)** Distance between the CD atoms of catalytic glutamates of glucose complexed UnBG11 is shown in dotted line.

Our data present the first report of successfully captured three distinct stages of cellobiose hydrolysis reaction: (a) substrate (cellobiose) at the catalytic site (Figure 2c), (b) covalently bound glycosyl intermediate to the catalytic glutamate (E370) (Figure 2d) and (c) product (glucose) in the active site crater (Figure 2e) of an active GH1 enzyme, UnBG11. These

high-resolution crystal structures of UnBG11 provided quite detailed insights into the substrate binding and the conserved glucose binding sites.

### **Identification of conserved glucose binding sites in UnBG11 and rational engineering**

From the analysis of the glucose-complexed native UnBG11 structures, we identified the specific sites where glucose molecules bind in the catalytic crater of this enzyme. These binding sites are designated as -1 subsite, +1 subsite, and +2 subsite (Figure 4a). The cellobiose binds near the catalytic glutamates where the substrate hydrolysis occurs (Figure 2c, Supplementary Figure S3a), and this site is referred as the -1 subsite. Also, in a glucose complexed structure, one glucose molecule is seen at -1 subsite, that contains catalytic residues. Any structural hindrance at -1 subsite may adversely affect the enzyme's activity, so it is better not to make any structural perturbation at -1 subsite. A second glucose molecule was found at +1 subsite located at the middle of the crater. The glucose molecule is well stabilized at the +1 subsite through interaction with the side chains of residues N241, E185, C188, I192, W343, F313. To understand the importance of residues involved in stabilizing the substrate at both -1 and +1 subsite, a thio-cellobiose soaking experiment with native UnBG11 crystal was performed. Thio-cellobiose acts as a moderate competitive inhibitor of  $\beta$ -glucosidase and has a similar structural resemblance to its substrate cellobiose<sup>22,29</sup>. To our surprise, no ligand was observed near the catalytic glutamates in the active site of this structure. Instead, a thio-glycosidic enzyme intermediate covalently bound to 188<sup>th</sup> cysteine (C188) residue at +1 subsite was formed due to which the substrate entry channel is blocked (PDB ID. 9J4B) (Figure 4b, Supplementary Figure S4a). Thio-glucose bound structure suggests possible importance of C188 facilitating feedback inhibition of UnBG11. We also hypothesized, introducing the hydrophobic interactions in the crater may restrict the entry of glucose to the active site<sup>5</sup>. Therefore, to test our hypothesis cysteine at 188<sup>th</sup> position at +1 subsite was mutated to valine (C188V) (Supplementary Figure S4b) to increase hydrophobic interactions and further biochemical as well as structural studies of the variant were performed.



**Figure 4: Understanding the molecular details of sugar-binding in the catalytic crater of UnBG11.** a) The crater is shown as surface (colour: greencyan) and lining representation, glucose molecules are shown in ball and stick (yellow carbon). Two catalytic glutamates are shown in purple sticks and all other residues are shown in white sticks. The -1, +1, +2 subsites are divided by dotted red lines. b) The  $2F_o - F_c$  electron density map (slate mesh) contoured at  $1\sigma$  level around the thio-glycosidic enzyme intermediate is shown. Catalytic glutamates are shown in purple stick and cysteine residue is shown as magenta stick whereas thio-glucose is shown in ball and stick representation. c) The  $2F_o - F_c$  electron density map (blue mesh) contoured at  $1\sigma$  level indicates presence of tryptophan at 261<sup>st</sup> position. Inset represents the omit  $F_o - F_c$  electron density map (green mesh) suggesting the presence of tryptophan residue at 261<sup>st</sup> position.

The third glucose molecule observed at the crater's entry site was identified to present as the same site of glucose captured in cellobiose complexed structure (Supplementary Figure S3c), suggesting that this is a conserved glucose binding site. This third glucose binding site is assigned as the +2 subsite, which has close contact with the histidine residue at the 261<sup>st</sup> (H261) position. Figure 4a shows all residues in the crater at different glucose binding subsites. In order to improve the glucose tolerance level of a GH1  $\beta$ -glucosidase, it is essential to explore the product binding sites for modifying them to reduce the glucose accumulation in the crater at high glucose concentrations in the bulk solvent. Increasing the hydrophobicity of the surface of the crater can obstruct the entry as well as binding of glucose to the active site<sup>5,13</sup>. It's a well-known fact that the aromatic residues play important role in the recognition of sugar substrates and may facilitate to increase the specificity of the enzyme<sup>30,31</sup>. Also, the aromatic residues at the entrance of the crater are responsible for substrate binding<sup>32</sup>. With this structural rational, a variant of UnBG11 was created by replacing H261 with tryptophan and the crystal structure of UnBG11\_H261W variant in apo form was determined at 1.8 Å resolution (PDB ID: 9J43) (Table 1: part II). The electron density confirms the mutation as the side chain of W261 is well defined in the map (Figure 4c). Further the kinetic and biochemical characterizations, as well as structural studies of these variants as glucose bound complex were performed.

### **Glucose tolerance and kinetic properties of UnBG11 variants**

Several variants of UnBG11, were created targeting the residues in +1 and +2 subsites in order to improve the glucose tolerance of this enzyme. Biochemical properties of the variants were determined using the pNPG hydrolysis following the same protocol as reported for the native UnBG11. UnBG11\_C188V did not change biochemical properties and had an optimal pH of 6.0 with an optimum temperature of 55 °C (Supplementary Figure S4c, S4d). However, UnBG11\_C188V showed difference in its kinetic properties (Table 2) (Supplementary Figure S4e) and has reduced affinity towards pNPG with  $K_m = 7.35 \pm 3.0$  mM. Maximum hydrolysis of cellobiose was seen at pH 6.5, and almost the same amount of activity was observed at pH 6.0 (Supplementary Figure S4f). Kinetic properties for cellobiose hydrolysis were determined at its optimum temperature 45 °C (Supplementary Figure S4g). As saturation of the UnBG11\_C188V mutant was not achieved (Supplementary Figure S4h), the kinetic properties of the enzymes for cellobiose hydrolysis could not be determined.

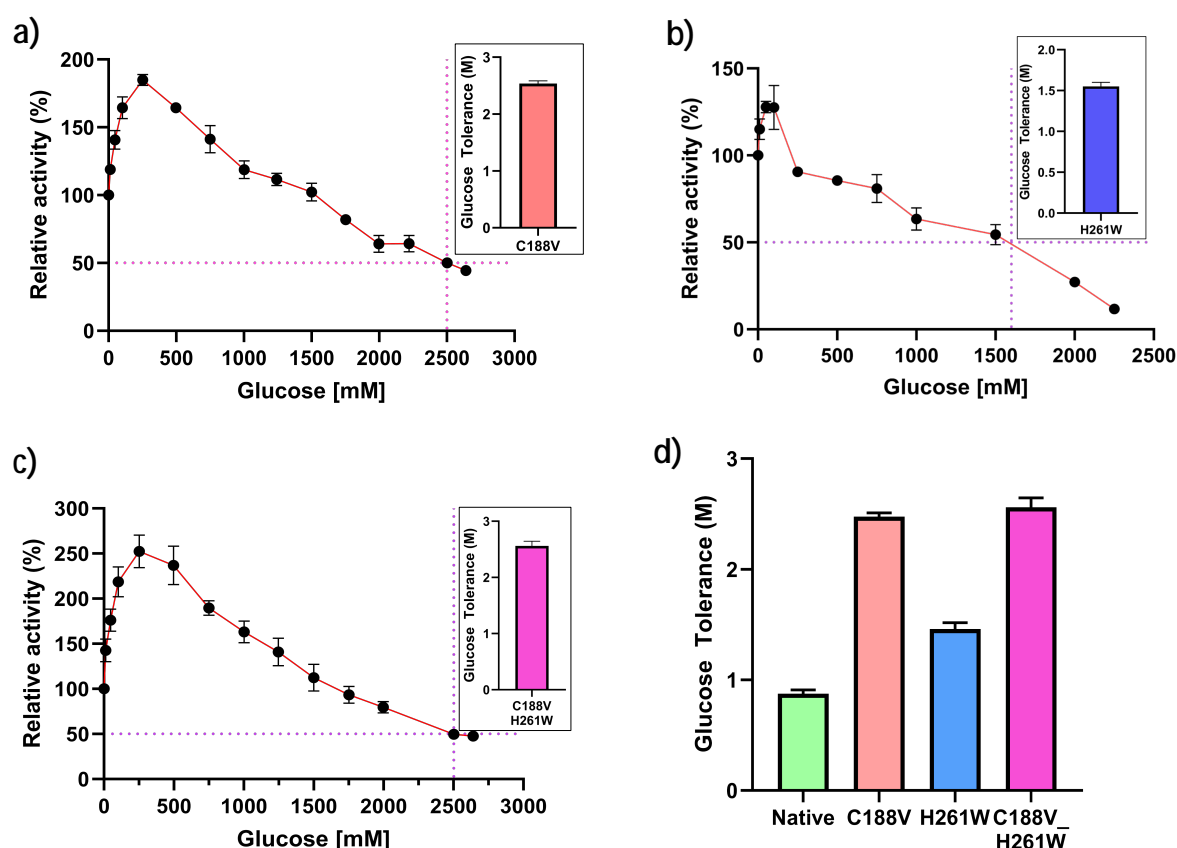
As mentioned earlier, we created another variant UnBG11\_H261W, by introducing mutation at +2 subsite, where glucose complexed structure showed alternate conformation of

W261 (Supplementary Figure S5a and S5b). UnBG11\_H261W showed activity in a wide range of pH from pH 5.0 to 7.0 and retained almost 80 % of its ability to hydrolyse pNPG, considering the 100 % activity at pH 6.5 (Supplementary Figure S5c). UnBG11\_H261W showed maximum hydrolysis of pNPG at 55 °C and almost the same activity at 50 °C (Supplementary Figure S5d) with  $K_m$  of  $0.65 \pm 0.04$  mM (Supplementary Figure S5e) (Table 2). Similarly, for cellobiose hydrolysis this variant enzyme showed activity within a short range of pH from pH 5.5 to 6.5. Activity at pH 5.5 was maximum and considered as 100 %, whereas pH 6.0 and 6.5 the enzyme has 90 % activity (Supplementary Figure S5f). This variant has same temperature optimum as of native UnBG11 (Supplementary Figure S5g). UnBG11\_H261W has improved kinetic properties than native and UnBG11\_C188V. UnBG11\_H261W has shown increased affinity towards cellobiose with  $K_m$  of  $22.87 \pm 1.1$  mM (Supplementary Figure S5h). The kinetic parameters of all variants of UnBG11 are mentioned in the Table 2.

**Table 2: Kinetic parameters of the native and variants of UnBGl1 towards hydrolysis of pNPG and cellobiose.**

Enzyme	$V_{\max}$ ( $\mu\text{mol min}^{-1}\text{mg}^{-1}$ )		$K_m$ (mM)		$k_{\text{cat}}$ ( $\text{s}^{-1}$ )		$k_{\text{cat}}/K_m$ ( $\text{mM}^{-1}\text{s}^{-1}$ )		Glucose Tolerance (M)
	<u>pNPG</u>	<u>Cellobiose</u>	<u>pNPG</u>	<u>Cellobiose</u>	<u>pNPG</u>	<u>Cellobiose</u>	<u>pNPG</u>	<u>Cellobiose</u>	
<i>Native UnBGl1</i>	$36.14 \pm 0.50$	$19.74 \pm 0.37$	$0.92 \pm 0.05$	$34.61 \pm 2.4$	$31.62 \pm 0.57$	$17.27 \pm 0.42$	$34.36 \pm 11$	$0.47 \pm 0.17$	1.0
<i>UnBGl1_C188V</i>	$17.23 \pm 1.5$	ND	$7.35 \pm 3.0$	ND	$15.07 \pm 1.31$	ND	$2.05 \pm 0.43$	ND	2.5
<i>UnBGl1_H261W</i>	$89.18 \pm 1.36$	$62.98 \pm 1.1$	$0.65 \pm 0.04$	$22.87 \pm 1.1$	$78.03 \pm 1.19$	$55.10 \pm 0.96$	$120.05 \pm 29$	$2.40 \pm 0.87$	1.6
<i>UnBGl1_C188V_ H261W</i>	ND	ND	ND	ND	ND	ND	ND	ND	2.5

ND = Not Determined



**Figure 5: Glucose tolerance level of all the UnBG11 Variants.** Data points of the relative activity are shown in black solid circle and connected with red line. The concentration of the glucose that causes 50 % loss of the relative activity, is considered as glucose tolerance (indicated by purple dotted lines). **a)** The graph depicts glucose tolerance level of UnBG11\_C188V. Inset is another representation of glucose tolerance through bar graph. **b)** The graph depicts glucose tolerance level of UnBG11\_H261W. Inset is another representation of glucose tolerance through bar graph. **c)** The graph represents the glucose tolerance level of UnBG11\_C188V\_H261W. Inset is a bar representation of glucose tolerance indicating highest glucose tolerance. **d)** Bar graph of comparison of glucose tolerance level of native UnBG11 with variants. All experiments were performed in triplicates (n = 3) with error bar representing  $\pm$  SEM.

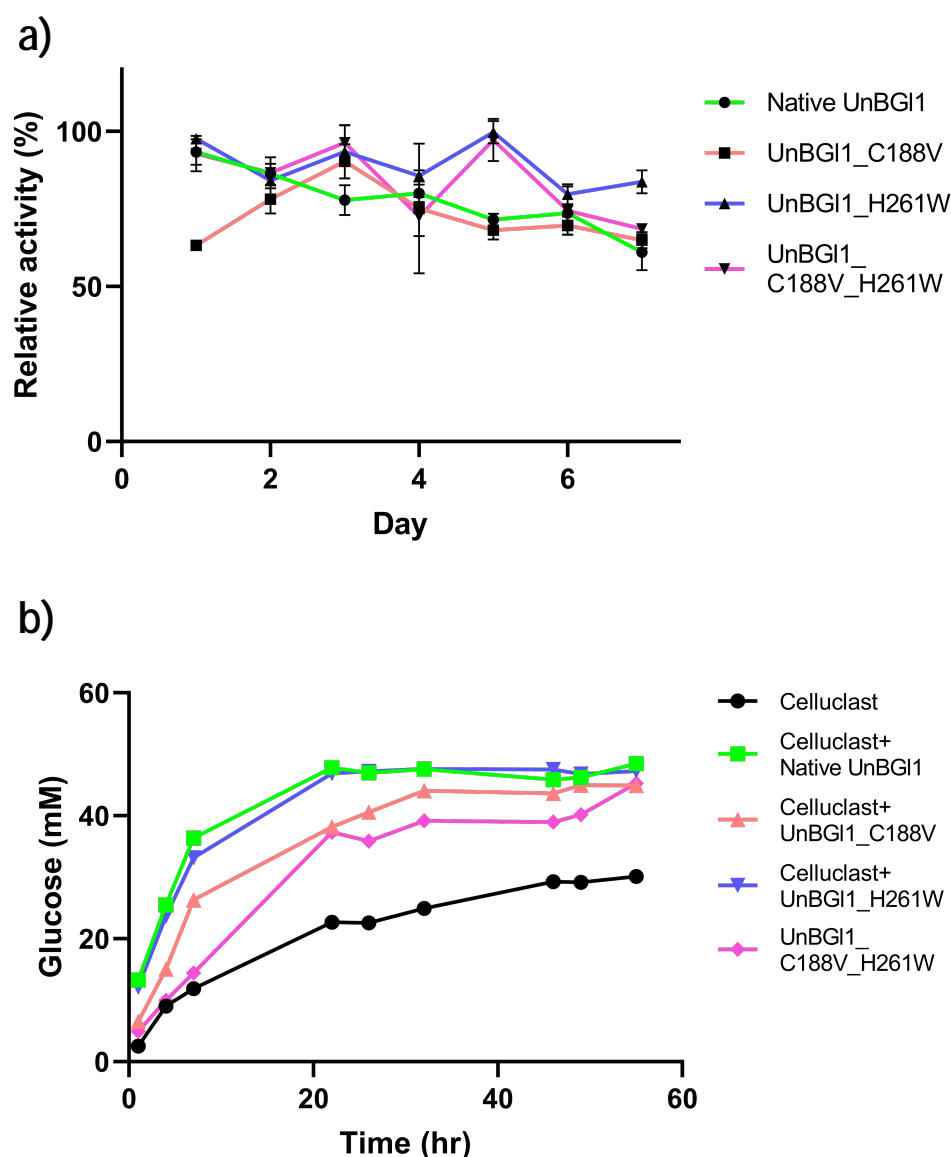
UnBG11\_C188V showed increased activity with a glucose tolerance level of about 2.5 M (Figure 5a). However, the kinetic properties were not improved (Table 2). Meanwhile, the UnBG11\_H261W variant exhibited improved kinetic properties for hydrolysing pNPG and Cellobiose (Table 2). The glucose tolerance of this variant is increased to 1.6 M (Figure 5b). As +1 subsite modification showed improved glucose tolerance and +2 modification showed improvement in kinetic properties therefore, a new variant of UnBG11, referred as UnBG11\_C188V\_H261W, was created by combining the modifications of these subsites (Supplementary Figure S6a). This double mutant gets stimulated by less concentration of glucose. Figure 5c illustrates an increase in enzyme activity with an increase in glucose concentration. UnBG11\_C188V\_H261W has 2.5-fold higher activity at 250 mM of glucose, and at further glucose concentration, the enzyme started to show reduced activity (Figure 5c). Notably, compared to all variants of UnBG11, the UnBG11\_C188V\_H261W has the highest

glucose tolerance of 2.5 M (Figure 5d). Meanwhile, its biochemical properties remain the same as those observed for the native UnBG11 using pNPG hydrolysis (Supplementary Figure S6b, S6c). However, the kinetic parameters for hydrolysing pNPG could not be determined as pNPG is not soluble at higher concentrations (Supplementary Figure S6d). For cellobiose hydrolysis, pH optimum has shifted to 6.5, retaining more than 80% of activity at pH 6.0 (Supplementary Figure S6e) and reduced the optimum temperature for hydrolysis to 45 °C (Supplementary Figure S6f) the kinetic properties for cellobiose hydrolysis were also not determined as cellobiose is not soluble at higher concentration (Supplementary Figure S6g).

### **Evaluation of UnBG11 variants for their industrial use**

Simultaneous saccharification and fermentation (SSF) is widely used method for bioethanol production. The enzymatic hydrolysis and microbial fermentation occur in a single process and so saves the cost of extra vessel. Preparation of soluble fermentable sugars and fermentation of these sugar generally happens at a temperature of 25 °C - 30 °C for 6 - 72 hrs.<sup>33</sup>, therefore it is essential to have active and stable  $\beta$ -glucosidase for bioethanol production.





**Figure 6: Evaluation of the properties of UnBG11 variants as per the industrial requirements. a)** Assessment of the stability and activity of UnBG11 variants over the period of 7 days in cell lysate. **b)** Glucose release profiles of the UnBG11 variants after blending the enzymes with industrially used  $\beta$ -glucosidase Celluclast.

Further enzyme stability from sugarcane bagasse was monitored in the crude extract (cell lysate of expressed UnBG11 variants cells). All the variants are stable at room temperature in pH 6.0 for 7 days (Figure 6a). On the 7<sup>th</sup> day native UnBG11, UnBG11\_C188V, UnBG11\_H261W, UnBG11\_C188V\_H261W retained 61%, 65%, 84%, 68% of activity respectively (Figure 6a). Although all the variants have different activity towards cellobiose, they are stable for 7 days. Also, all the UnBG11 variants were active when blended with a cellulase cocktail from *T. reesei* (Celluclast). Native UnBG11 and UnBG11\_H261W, when blended with Celluclast, showed higher amount of glucose release of around 48 mM and UnBG11\_C188V, UnBG11\_C188V\_H261W released 45 mM of glucose after 55 hrs. by hydrolysing sugarcane bagasse. Whereas, Celluclast only showed release of glucose of around

30 mM (Figure 6b). Further structural studies of UnBG11 variants were done to understand the mechanism behind the improvement in glucose tolerance.

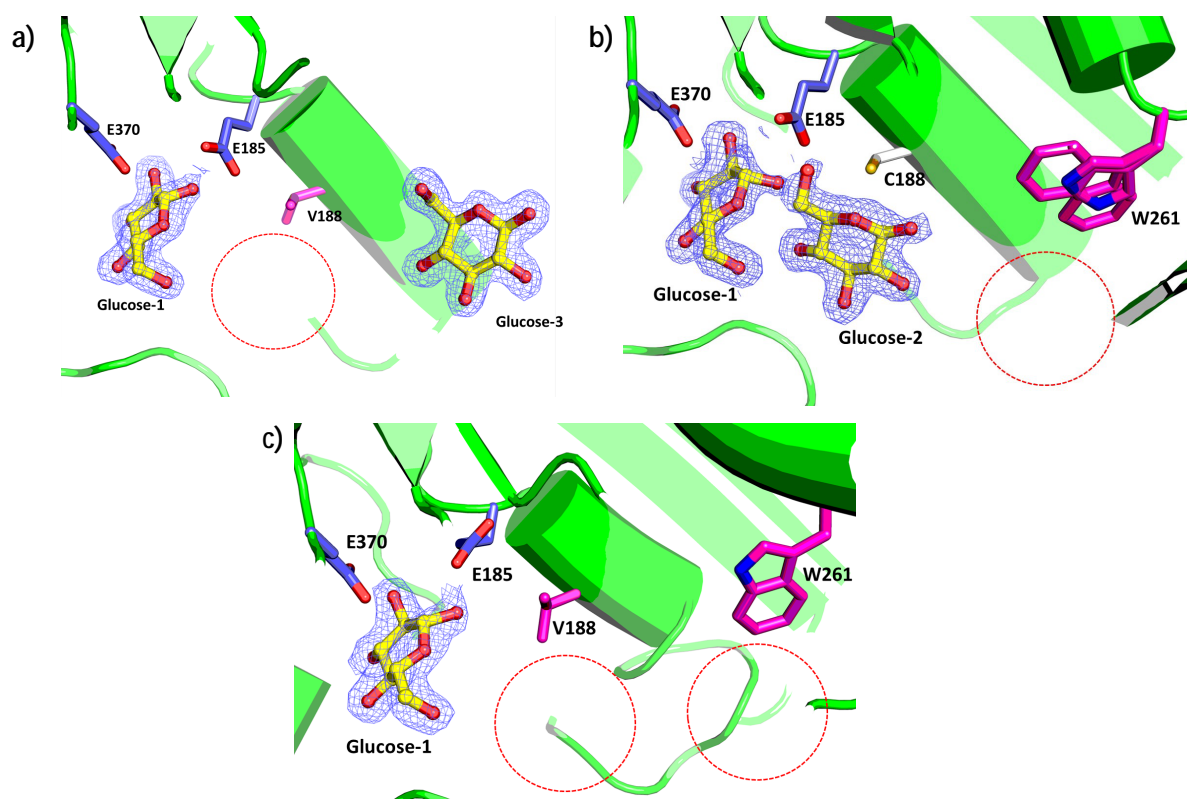
### **Structural relevance with the glucose tolerance**

We determined multiple high-resolution crystal structures (Table 1: Part I and Part II) of the UnBG11 variants as apo and complexed with glucose (Figure 7) to understand the molecular basis of improved glucose tolerance of these variants. We also solved the cellobiose bound structure of UnBG11\_H261W. The glucose complexed native UnBG11 and UnBG11\_C188V structures do not show any overall structural difference, and their superposition produces a r.m.s.d value of 0.108 Å. The UnBG11\_C188V structure has only two glucose molecules in its crater, one at the -1 subsite and another at the +2 subsite (PDB ID. 9J42) (Figure 7a). As expected, glucose at the +1 subsite is absent in the UnBG11\_C188V structure (Figure 7a). This might be primarily because of loss of the interactions between glucose at the +1 subsite and the cysteine due to its replacement by a valine. Loss of interactions of glucose in the +1 subsite aided in reducing the glucose binding ability of the crater leading in reduction of glucose accumulation, leading to the increase of the glucose tolerance level (2.5 M) (Figure 5a) of UnBG11\_C188V as compared to the native UnBG11 (0.9 M) (Figure 1f).

The glucose tolerance of the UnBG11\_H261W was observed to be 1.6 M (Figure 5b) which is not greater than that of UnBG11\_C188V but much higher than the native enzyme. However, the kinetic properties of UnBG11\_H261W have improved as compared to the native and C188V variant (Table 2). The high-resolution glucose complexed UnBG11\_H261W structure (PDB ID. 9J44) provided insights into the active site crater to understand the basis for increased glucose tolerance level of this variant. The structure showed absence of a glucose molecule at the +2 subsite (Figure 7b). Further electron density analysis revealed the double confirmation of W261 (Supplementary Figure S5a, S5b). Flexible nature of the bulky hydrophobic side chain of W261 might be one of the reasons for the absence of a glucose at the +2 subsite. Decrease of glucose affinity due to the presence of a hydrophobic residue at the entry site (+2 subsite) of UnBG11\_H261W crater contributed to increase glucose tolerance level of this enzyme.

Structure of glucose complexed UnBG11\_C188V\_H261W double mutant was refined at 2.0 Å (PDB ID. 9J4A) (Table 1: part II). Double mutant showed the absence of glucose molecules at +1 and +2 subsites and a single molecule of glucose was present only at -1 subsite (Figure 7c). This suggest that it requires high concentration of glucose to get trapped into the crater of double mutant UnBG11. Our structural and biochemical data clearly indicate that incorporation

of hydrophobic residues at +1 and +2 subsites of UnBG11 has aided to improve glucose tolerance of this enzyme.



**Figure 7: Molecular basis of improvement of the glucose tolerance of UnBG11 variants.** Zoomed in view of the active sites of the UnBG11 variants showing the glucose binding. Catalytic glutamates are shown in purple stick, mutated residues (V188, W261) are shown in magenta stick, glucose molecules are shown as ball (red) and stick (yellow) representation. Absence of glucose molecules in the active site crater of these mutants as compared to the native enzyme (Figure 2e) is marked with dotted red circle. **a)**  $2F_o - F_c$  map contoured at  $1\sigma$  level illustrating the electron density (blue mesh) around two bound glucose molecules in UnBG11\_C188V variant structure. The glucose molecule near V188 is absent in this mutant. **b)**  $2F_o - F_c$  map contoured at  $1.0\sigma$  illustrating the electron density (blue mesh) around two bound glucose molecules in UnBG11\_H261W variant structure. The glucose molecule near W261 is absent, no electron density for glucose was observed near W261. **c)** The  $2F_o - F_c$  map contoured at  $1\sigma$  level showing an electron density (blue mesh) around a single glucose molecule near active site and absence of electron density of 2<sup>nd</sup> and 3<sup>rd</sup> glucose molecule.

## Discussion

GH1  $\beta$ -glucosidases play pivotal role in process of degradation of cellulose in smaller fermentable glucose units by hydrolysing  $\beta$ ,1-4 linkage<sup>18</sup>. These enzymes possess feedback inhibition property, emphasizing the necessity for  $\beta$ -glucosidases with enhanced glucose tolerance to ensure efficient and continuous cellulose breakdown<sup>34</sup>. In this investigation, a  $\beta$ -glucosidase from soil metagenome (UnBG11) has been studied extensively. We have determined 12 high-resolution crystal structures (including apo, substrate-bound, product-

bound, and inhibitor-bound forms), of UnBG11 and its variants. Taken together our findings, the strategies and important structural features to be considered for the rational design to improve glucose tolerance of GH1 family  $\beta$ -glucosidases are discussed.

### ***Understanding cellobiose hydrolysis mechanism through crystallographic snapshots***

For native  $\beta$ -glucosidases, no structure is reported so far that shows substrate binding and presence of a covalently linked glycosyl intermediate in the active site. Here, we present the first-ever crystal structure of  $\beta$ -glucosidase in its native state complexed with the substrate, cellobiose, precisely at the pre-hydrolytic state (Figure 2c). In case of UnBG11, E185 and E370 are catalytic residues that carry overall hydrolytic reaction. The binding and hydrolysis of cellobiose have been studied through computational techniques. The mode of cellobiose binding seen in UnBG11 catalytic pocket is similar to that observed previously by the quantum mechanics/molecular mechanics (QM/MM) studies on hydrolysis of laminaribiose ( $\beta$ -1,3 linked) and cellobiose by a plant  $\beta$ -glucosidase (BGul1)<sup>35,36</sup>. During the initial binding phase, the enzyme's active site adjusts to the binding of cellobiose, making it easy for the substrate to be catalytically processed. Substrate binding further induces conformational changes in UnBG11, aligning cellobiose appropriately within the catalytic site. QM/MM calculations have demonstrated how the enzyme dynamically adapts the positioning of substrates to optimize conditions for nucleophilic attack, thus underlining structural adaptability during catalysis. The same orientation of cellobiose is seen in the UnBG11 active site. This state is stabilized by various hydrogen bond interactions between cellobiose and active site residues. Hydrogen bonding interactions (Supplementary movie S3) seen in the crystal structure of cellobiose bound UnBG11 are comparable to those reported for BGul1. The correct orientation of cellobiose is necessary to achieve transition states and further formation of glycosyl intermediate. This study is the first report of capturing a glycosyl intermediate in native and active  $\beta$ -glucosidase. So far, no experimental structural evidence of pre-hydrolytic state of substrate and glycosyl intermediate has been reported for  $\beta$ -glucosidases. The structural snapshot of cellobiose complexed UnBG11 elucidates critical substrate-enzyme interactions and sheds light on the initial steps of the enzymatic reaction. Also, the structure of a reaction intermediate where the substrate forms a covalent bond with the nucleophilic glutamate residue in the active site, offers a deeper understanding of the transition-state stabilization during catalysis.

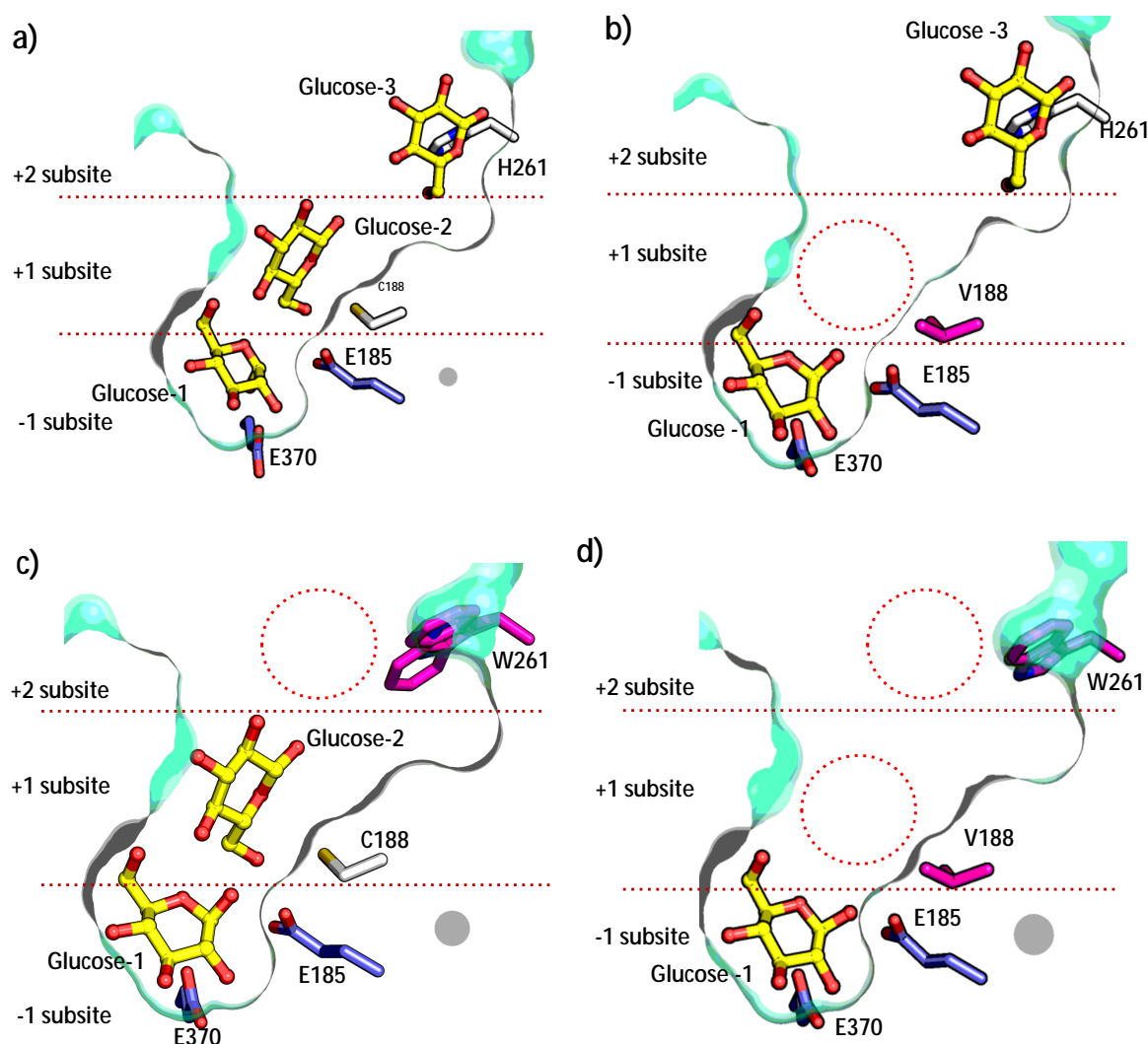
## ***Glucose binding subsite mutations and corresponding modulation in kinetic properties in UnBGII***

Insufficient  $\beta$ -glucosidase activity is the reason for the accumulation of cellobiose in the enzymatic cellulose hydrolysis process, which results in the inhibition of endo- and exo-glucanases<sup>16</sup>. Therefore, identifying the critical residues that can enhance the efficiency of  $\beta$ -glucosidases activity is highly needed. Determination of the kinetic parameters of  $\beta$ -glucosidases is essential for evaluating their efficiency and potential industrial applications. Here, we investigated the kinetic properties of all the variants of UnBGII against their enzyme property to hydrolyse pNPG and cellobiose (Table 2). Most of the  $\beta$ -glucosidase enzymes have an optimum temperature above 40 °C and exhibit the maximum activity in the pH range of 4.0 - 6.5; as documented by Teugjas et al.<sup>37</sup>. UnBGII and all of its variants have the same property with an optimum temperature range of 45 - 55 °C and pH range of 4.0 - 6.5 for both pNPG and cellobiose hydrolysis. Efficient industrial cellulose conversion to simple sugar like glucose requires highly stable and active enzymes. A process like simultaneous saccharification and fermentation (SSF) uses active and stable enzymes at lower temperatures<sup>38</sup> of around 30 °C - 40 °C<sup>39</sup>. UnBGII satisfies the requirements of SSF process with enzyme stability of 48 hrs at 30 °C in pH 5.5 and 6.0 (Figure 1d). All the UnBGII variants were stable in the cell lysate for 7 days at room temperature (Figure 6a), and therefore the use of these variants in the industry will save a lot of protein purification and storage expenses. When kinetic parameters were compared of native enzyme with +2 subsite H261W variant, native UnBGII showed more affinity and catalytic efficiency towards both pNPG and cellobiose. Meanwhile, the +1 subsite C188V variant drastically changed the kinetic behaviour of the enzyme and could not achieve complete substrate saturation. A similar observation has already been reported where a  $\beta$ -glucosidase from *Thermococcus sp.* (O08324) was modified by introducing various mutations at +1 and +2 subsite<sup>40</sup>. H0HC94, a  $\beta$ -glucosidase from *Agrobacterium tumefaciens* 5A showed lower affinity after L178E mutation was introduced at +1 subsite towards both pNPG and cellobiose<sup>41</sup>. Based on our structural analysis, we believe that +1 subsite plays an important role in maintaining the kinetic properties of GH1  $\beta$ -glucosidases. Extreme care and through structure as well as sequence analysis have to be done if there is a need to modify the +1 subsite. It is important to determine the enzyme's activity when it is used in cellulase cocktails. For that we used Celluclast (an enzyme cocktail from *T. reesei* that does not have  $\beta$ -glucosidase) blended with variants of UnBGII and used to hydrolyse bagasse. UnBGII variants can release

more glucose from bagasse when added to the cellulase cocktail (Figure 6b). Therefore, our data suggest that these variants of UnBG11 are suitable candidates for the SSF process for ethanol production.

### ***Structural basis of the glucose stimulation and tolerance in GH1 $\beta$ -glucosidases***

Based on the effect of glucose concentration,  $\beta$ -glucosidases are classified into four groups that: (i) get inhibited at lower concentrations of glucose, (ii) are glucose tolerant, (iii) get stimulated at low concentrations of glucose, (iv) do not undergo inhibition at high concentration of glucose<sup>42,43</sup>. In this study, we have selected UnBG11, a  $\beta$ -glucosidase that shows glucose stimulation at lower glucose concentrations (Figure 1f, 5a, 5b, 5c). Our biochemical studies indicate that the glucose can bind to the stimulatory sites of the enzyme active site crater up to a specific concentration, which is 100 mM for the native and UnBG11\_H261W, whereas in the case of UnBG11\_C188V and UnBG11\_C188V\_H261W, it has increased to 250 mM. However, beyond that stimulatory glucose concentration, all variants start to show inhibition. Compared to the native enzyme, all the variants have showed improved glucose tolerance. Although, UnBG11\_H261W with +2 subsite mutation did not show any effects on the glucose stimulation activity, both single and double mutant variants having cysteine to valine substitution at +1 subsite exhibited an increase in the stimulatory glucose concentration and the stimulatory effect. It is evident that the +1 subsite plays a vital role in the glucose stimulation, but the mechanism of such effect is still unclear. It has been postulated that this effect may arise from the geometric configuration of specific binding sites, where bound glucose enhances substrate cleavage activity through transglycosylation or other mechanisms, such as allosteric modulation<sup>44</sup>. Our extensive structural data and analysis of multiple ligand complexed UnBG11 structures, two predominant allosteric glucose binding sites (+1 and +2 subsites) away from the catalytic centre have been identified. However, modifying those two allosteric sites (in UnBG11\_C188V and UnBG11\_H261W) did not diminish the glucose stimulatory effect. Therefore, the allosteric binding of glucose may not be the reason for stimulatory effect of  $\beta$ -glucosidases.



**Figure 8: Effect of mutations on the conserved glucose binding sites in UnBG11 variants.** Crater of all the variants is shown in surface representation, catalytic glutamates are in purple stick and other native residues are in white sticks; whereas mutated residues are shown in magenta sticks. All glucose molecules are illustrated as ball and stick. Absence of glucose molecules in the active site crater of these mutants as compared to the native enzyme is marked with dotted red circle. **a)** Three glucose molecules bound to the subsites of native UnBG11 crater. **b)** Absence of second glucose from middle of the crater in UnBG11\_C188V due to C188V mutation at +1 site. **c)** Absence of third glucose from entry site of the crater in UnBG11\_H261W due to H261W mutation at +2 site. **d)** Absence of two glucose molecules at +1 and +2 subsites in UnBG11\_C188V\_H261W.

Glucose influences the active site environment by altering the water matrix and steric geometry in the substrate channel or at the catalytic centre, subsequently affecting substrate binding. Additionally, glucose binding outside the active site, potentially in the middle of the substrate channel or at its entrance, can increase glucose tolerance<sup>44</sup>. However, our structural and biochemical data indicate that loss of interaction of glucose from +1 and +2 subsites increases glucose tolerance in UnBG11 variants. All three variants, UnBG11\_C188V, UnBG11\_H261W and UnBG11\_C188V\_H261W, that showed loss of interaction of glucose from the middle and entry site of the crater exhibited significant improvement in glucose

tolerance to 2.5 M, 1.6 M and 2.5 M, respectively. These values are higher than the glucose tolerance reported for many  $\beta$ -glucosidases<sup>37</sup> so far. The high-resolution crystal structures clearly indicate that glucose binding sites are very specific in the case of UnBG11. That proves, glucose binding in the UnBG11 crater is not non-specific. Glucose binding to +1 and +2 subsites are responsible for low glucose tolerance of the enzyme due to blockage of the entry channel resulting in non-accessibility of the substrate to catalytic glutamates (E185, E370) at the -1 subsite for hydrolysis. Removal of the interactions of glucose from the +1 and +2 subsites in the crater is essential to facilitate the entry of the substrate cellobiose to the catalytic centre. Modifications at the +1 subsite by introducing valine at the 188<sup>th</sup> position relieved glucose from +1 subsite (Figure 8b), leading to increased glucose tolerance. Introduction of a hydrophobic residue at +2 subsite has also shown loss of interactions with glucose at +2 subsite (Figure 8c) leading to increase in glucose tolerance. Creating a double mutant by combining +1 and +2 subsite modifications has led to further increase in glucose tolerance (Figure 8d).

### ***Role of residues in catalytic crater towards increasing glucose tolerance of GH1 $\beta$ -glucosidases***

As GH1  $\beta$ -glucosidases share a highly similar structural fold, the findings from our structural studies on UnBG11 can be applicable to other enzymes within the same family. Through structure-based sequence alignment with glucose-tolerant  $\beta$ -glucosidases from the Betagdb database<sup>42</sup> and few other  $\beta$ -glucosidases, we have identified residues that could be modified to boost glucose tolerance, as well as those that should remain untouched in GH1  $\beta$ -glucosidases. The supplementary movie S4 illustrates the residues involved in formation of the catalytic core of UnBG11, which are highly conserved across most  $\beta$ -glucosidases. We believe altering these amino acids could negatively impact enzyme activity due to their proximity to the catalytic residues. Characteristic NEP and ENG motifs, which are essential features of GH1  $\beta$ -glucosidases, are also part of this core. R96 interacts with E370 (salt bridge) and N184 (hydrogen bond), which are part of the catalytic core and contribute to the stability of the active site (Supplementary Figure S7a) in UnBG11. Whereas Q39 forms hydrogen bond with glucose or cellobiose, as observed in crystal structure (Supplementary Figure S7b, S7c). These hydrogen bonds are essential for suitable conformation of cellobiose in the active site for hydrolysis. H140 is located at the crater's base and closes the catalytic crater as shown in the supplementary movie S4. Additionally, W141, W417 and W425 surround the catalytic core of the enzyme and maintain the hydrophobic environment. Y312 and E424 residues from +1 subsite are present at hydrogen bonding distance from glucose and cellobiose present at -1



subsite (Supplementary Figure S7d, S7e). The residues Q39, R96, H140, W141, N184, E185, P186, E370, W417 and W425 are present at -1 subsite, Y312 and E424 present at +1 subsite are conserved across all glucose tolerant  $\beta$ -glucosidases (Supplementary Figure S8a). Modifying these residues could adversely affect cellobiose binding and enzyme activity, making such alterations unwise for improving glucose tolerance in GH1  $\beta$ -glucosidases.

Our structural and biochemical studies have underscored the critical role of C188 and H261 residues in improving glucose tolerance. Multiple sequence alignment of glucose tolerant GH1  $\beta$ -glucosidases showed the presence of aromatic residues at 187<sup>th</sup> site in +1 subsite. Most of the  $\beta$ -glucosidases have tryptophan residue at this site so introducing tryptophan at 187<sup>th</sup> position can increase the glucose tolerance. This was demonstrated by mutating G168 to tryptophan in  $\beta$ -glucosidase from *Acetivibrio thermocellus* (AtGH1), which increased in glucose tolerance to 820 mM<sup>45</sup>. G168W variant of AtGH1 structurally aligns with L187, therefore L187 to tryptophan mutation can increase glucose tolerance of UnBG11. Similarly, at 192<sup>nd</sup> and 199<sup>th</sup> position Ile and Met are present in UnBG11 whereas most glucose tolerant  $\beta$ -glucosidases have Leu and His at this position (Supplementary Figure S8b). Generating I192L and M199H variants of UnBG11 would increase the glucose tolerance. Based on our structural and bioinformatic studies, it is evident that L187W, C188V, I192L, M199H, and H261W modifications at the +1 and +2 subsites can enhance glucose tolerance. This study lays a robust foundation for the rational design of GH1 family  $\beta$ -glucosidases, enabling the development of enzyme variants with improved performance under industrially relevant conditions. Such advancements are poised to contribute substantially to the economic viability and efficiency of bioethanol production processes.

## Materials and Methods

### Cloning, mutagenesis, protein expression and purification

The codon optimized sequence of *UnBGII* gene was synthesized from GeneArt, Invitrogen (USA) to express in *Escherichia coli* (*E. coli*) strain BL21(DE3). *UnBGII* gene was amplified by PCR using gene-specific primers with help of Q5 polymerase enzyme (NEB) and was cloned into *NdeI* and *XhoI* sites by using T4 DNA ligase enzyme (ThermoFisher Scientific) of pET43.1b to obtain C-terminal His<sub>6</sub> tag protein. The recombinant plasmid was transformed in *E. coli* BL21(DE3) for over-expression of UnBGII. These cells were grown at 37 °C for 16 hrs. In 50 ml Luria-Bertani (LB) media containing 100 mg/ml ampicillin. 20 ml of this grown culture was used to inoculate in 1.5 lit LB media with the same ampicillin concentration. *E. coli* BL21 (DE3) cells were grown in 3 lit LB medium to obtain a high amount of purified protein. Cells were grown at 37 °C till OD at 600 nm reaches 0.6 - 0.8. The cells were induced with 0.4 mM IPTG for overexpression of the protein, and cells were grown at 24 °C for 16 hrs. The media was centrifuged at 8000 rpm for 10 minutes to collect the cell pellet. After harvesting the cells, 5x volume of sodium phosphate buffer (50 mM Na-PO<sub>4</sub>, 400 mM NaCl, and pH 7.4) was added to the pellet to dissolve. Cell disruption was performed with ultrasonication at 4 °C and centrifuged at 16000g for 30 minutes. A 0.4 µm PVDF membrane was used to filter the supernatant.

The supernatant was loaded onto a 5ml His Trap FF Ni-NTA affinity column equilibrated with sodium phosphate buffer using FPLC (ÄKTAprime Plus, GE Healthcare) at a flowrate 1 ml/minute. UnBGII was eluted using equilibration buffer containing increasing concentrations of imidazole (25 mM, 75 mM, 125 mM, 187 mM and 100 mM). The maximum amount of protein was eluted with 75 mM imidazole. Eluted protein was further pooled and concentrated to 2 ml using an Amicon ultrafiltration concentrator of 10 kDa cutoff. The concentrated protein was purified by a pre-equilibrated (50 mM Na-PO<sub>4</sub>, 50 mM NaCl, pH 7.4) Superdex 200 16/600 (GE Healthcare) gel filtration column. Peak fractions were collected, and the purity of the enzyme was analysed by running SDS-PAGE.

Further for preparation of site directed mutants (SDMs), site specific primers (Supplementary Table S1) were designed. Recombinant native UnBGII plasmid was amplified using site specific primers by Q5 polymerase enzyme for introducing C188V and H261W mutation. To create UnBGII\_C188V\_H261W variant, UnBGII\_C188V plasmid was chosen as template and UnBGII\_H261W SDM primers were used for amplification by Q5 polymerase.

The expression and purification of all the variants of UnBG11 was done by following the protocol used for native UnBG11.

### **Enzyme Assay:**

Understanding the ideal reaction conditions for carrying out substrate hydrolysis is crucial. pH and temperature are the two most crucial parameters to comprehend in enzymatic activity. The enzymatic characterization of UnBG11 was done using two substrates, p-nitrophenyl- $\beta$ -D-glucopyranoside (pNPG) (Merck) and cellobiose (Sigma-Aldrich). All assays were conducted under optimal conditions to ensure reliable measurement of enzymatic activity.

For pNPG hydrolysis assay, a 40 mM stock solution was prepared in 50 mM sodium acetate and 50 mM sodium phosphate buffer of optimum pH. Reaction mixtures containing 40  $\mu$ l of appropriately diluted pNPG solution and 450  $\mu$ l of buffer were pre-incubated at optimum temperature for 5 minutes. 10  $\mu$ l of appropriately diluted UnBG11 was added to the pre-incubated reaction mixture and reaction was carried out on dry bath for 10 minutes. The reaction was terminated by addition of 500  $\mu$ l of 0.2 M sodium carbonate and release of product p-nitrophenol (pNP) was quantified spectrophotometrically by measuring absorbance at 405 nm. For cellobiose hydrolysis assay, 350 mM cellobiose stock was prepared in 50 mM sodium acetate and 50 mM sodium phosphate buffer of optimum pH. Final reaction volume was set to 100  $\mu$ l that contain 95  $\mu$ l of appropriately diluted cellobiose and 5  $\mu$ l of UnBG11. The assay was carried out at optimum temperature for 30 minutes and terminated by heating at 90 °C for 10 minutes. After cooling, 2  $\mu$ l of the reaction mixture was mixed with 200  $\mu$ l Glucose-Oxidase Peroxidase (GOD-POD) reagent (Eco Pak Glucose 500, Accurex Biomedicals Pvt. Ltd.)<sup>46</sup> in 96-well microtiter plate and incubated at 37 °C for 15 minutes. The developed pink colour, indicative of glucose production was measured spectrophotometrically at 505 nm using a microplate reader (Vantastar, BMG Labtech). Above mentioned protocol was used for characterization of all UnBG11 variants.

Specific activities for both substrates were calculated as the amount of substrate hydrolysed ( $\mu$ moles) per minute per milligram of enzyme and expressed in  $\mu$ mol/min/mg. Enzyme activities were measured in triplicates, and negative controls without enzyme were included to account for non-enzymatic hydrolysis. Standard calibration curves for pNP (405 nm) and glucose (505 nm) were used for accurate quantification of reaction products.

### **Determination of optimum pH and optimum temperature**

The reaction mixtures and substrates were prepared in 50 mM sodium acetate (pH 3.5, 4.0, 4.5, 5.0, 5.5) and sodium phosphate (pH 6.0, 6.5, 7.0, 7.5, 8.0) in these buffers. These helped a

comprehensive analysis of enzymatic performance under various pH conditions. For pNPG hydrolysis, 10  $\mu$ l of purified UnBG11 (1:200 dilution of 1 mg/ml) was used. Assay was carried at 55 °C for 10 minute and terminated by adding 0.2 M sodium carbonate. Maximum release of pNP was identified and considered as optimum pH for pNPG hydrolysis.

For cellobiose hydrolysis assay, 5  $\mu$ l of 1 mg/ml UnBG11 was used. Reaction was carried out in PCR machine at 50 °C for 30 minute and terminated at 90 °C. GOD-POD kit used to determine release of glucose, maximum glucose release was identified by spectrophotometric measurement at 505 nm. The value obtained at optimum pH was defined as 100 %.

Further for determining the optimum temperature for substrate hydrolysis, the reaction mixture and substrates were prepared in 50 mM sodium phosphate, pH 6.0 buffer. pNPG and cellobiose hydrolysis reactions were carried out in temperatures range of 25 °C - 80 °C.

For both assays, the activity at the optimal pH and temperature was normalized to 100% and served as the reference for relative activity comparisons across the tested pH and temperature range.

### **Stability of native UnBG11**

The temperature stability of the native UnBG11 was checked at 30 °C for 48 hrs. Cellobiose, the actual substrate of  $\beta$ -glucosidase, was used for performing stability tests. Experiments were performed in 50 mM sodium acetate (pH 5.5) and 50 mM sodium phosphate (pH 6.0) buffers. The reaction mixture was made such that it contained 990  $\mu$ l of 350 mM cellobiose prepared in the buffer. 10  $\mu$ l of native UnBG11 (1mg/ml) was added to the reaction system. The cellobiose hydrolysis reaction was carried out at 50 °C for 30 minutes. The reaction was inhibited by placing reaction tubes in a boiling water bath for 15 minutes. 10  $\mu$ l of an aliquot from the reaction mixture was mixed with 1 ml of GOD-POD reagent<sup>46</sup>. This assay system was incubated at 37 °C for 15 minutes. The pink colour was estimated spectrophotometrically at 505 nm for calculating enzyme activity.

Further, the cell lysate of all UnBG11 variants were incubated at room temperature for 7 days at pH 6.0 and activity of the variants against cellobiose hydrolysis was measured each day as mentioned above.

### **Measurement s of kinetic parameters**

Enzyme kinetics studies of UnBG11 variants were performed using its original substrate (cellobiose) and artificial substrate 4-nitrophenyl  $\beta$ -D-glucopyranoside (pNPG). To determine the kinetic parameters towards cellobiose hydrolysis, increasing concentration of cellobiose (3.5 mM to 332.5 mM) was incubated with the appropriately diluted enzyme at 50 °C for 30

min, and the reaction was inhibited by incubating at 90 °C for 10 minutes. Finally, to determine the amount of glucose released, 2 µl from the reaction mixture was taken, and 200 µl GOD-POD reagent was added and incubated at 37 °C for 15 min, and absorbance was measured at 505 nm. To determine the enzyme's kinetic properties for hydrolysing the pNPG, increasing concentration of pNPG (50 to 40 mM) incubated with appropriately diluted enzyme at 55 °C for 10 min and the reaction was inhibited by adding 0.5 ml of 0.2M Na<sub>2</sub>CO<sub>3</sub>. The absorbance of the produced yellow colour was measured at 405 nm.

### **Evaluation of glucose tolerance of UnBG11 variants**

The glucose tolerance level of the native and variants of UnBG11 was determined by performing a pNPG hydrolysis assay in the presence of an increasing concentration of glucose (0 mM to 2.3 M). A fixed concentration of pNPG (40 µl of 40 mM) was incubated with fixed concentration of enzyme (10 µl of 1:200 diluted 1 mg/ml enzyme) and increasing glucose concentration. The reaction was carried at 55 °C for 10 minutes and inhibited by adding 0.2 M Na<sub>2</sub>CO<sub>3</sub>, and absorbance was measured at 405 nm.

### **Monitoring activity against sugarcane bagasse**

The sugarcane bagasse was pretreated to separate the cellulose<sup>47,48</sup>. Obtained cellulose was further used for hydrolysis assay. An enzyme cocktail of cellulase from *Trichoderma reesei* (Celluclast) that does not contain β-glucosidase was choose for this assay. The Celluclast was blended with the UnBG11 variants in 1:1 ratio (Undiluted (neat) Celluclast: 2 mg/ml of UnBG11 variant) to prepare new cocktail that can hydrolyse the cellulose to glucose. 50 µl of newly prepared enzyme cocktail was used to hydrolyse 20 mg/ml of bagasse prepared in sodium phosphate buffer (pH 6.0) and incubated at 30 °C for 55 hrs. Released glucose was measured by GOD-POD kit.

### **Crystallization of UnBG11 variants and crystal soaking experiments**

Purified UnBG11 was concentrated to 10 mg/ml in 50 mM sodium phosphate buffer containing 50 mM NaCl pH 7.4, using 10 kDa Amicon centrifugal unit. The concentrated protein was subjected to crystallization screening using Phoenix crystallization robot (Art Robbins Instruments) by sitting drop vapor diffusion method in 96-well Intelli-plates available at the “Protein Crystallography Facility”, Indian Institute of Technology, Bombay, India. For the crystallizations, different commercially available crystallization screens such as PEGS suite (Qiagen), PEGRx (Hampton Research), PEG/Ion (Hampton Research), Index (Hampton Research) and JCSG-plus (Molecular Dimensions) were used. Native UnBG11 got crystallized

after 90 days at 22 °C in a condition consisting of 0.056 M Sodium phosphate monobasic monohydrate, 1.344 M Potassium phosphate dibasic, pH 8.2 (condition 19 from Index - HR2-144, Hampton research). Further optimization of crystallization was carried out by performing the hanging-drop vapor diffusion method using flat bottom 24 well polystyrene plates (Nest Biotech) and incubated at 22 °C. Different ratios of UnBG11 to crystallization buffer (mother liquor), such as 1:1, 1:2, and 2:1 were tested to obtain high-quality crystals. The crystallization drops were set on silicon coverslips and placed over the reservoir having 500 µl of mother liquor. Once crystals were obtained, they were used for seeding experiments to enhance crystal growth. Seeding experiments successfully reduced the crystal growth time to 3 - 5 days and produced better quality crystals. The same protocol was applied for crystallizing all other variants of UnBG11, ensuring consistency in crystal quality and reproducibility.

### **Soaking experiments to obtain ligand complexed structures**

As cellobiose hydrolysis is one step reaction and does not require any co-factor or activator for completion of reaction, co-crystallization is not a suitable method for obtaining substrate (cellobiose) complexed structure. Soaking experiment is often used to obtain protein-ligand complexes<sup>49</sup>, ligands can reach the binding site by diffusing through the solvent channels<sup>50</sup>. To confirm the catalytic activity of UnBG11 in its crystalline form, initial soaking experiments were conducted using 40 mM p-nitrophenyl β-D-glucopyranoside (pNPG). A crystal of UnBG11 (after washing multiple times with mother liquor) was transferred into the pNPG solution prepared in mother liquor, and the enzymatic activity was verified by observing yellow colour in the crystals and the surrounding solution.

Further soaking experiments were carried out with cellobiose at concentrations ranging from 50 mM to 350 mM. Native UnBG11 and UnBG11\_H261W crystals were harvested using cryo-loops, gently washed with mother liquor, and soaked in cellobiose solutions for varying durations (10 seconds to 15 minutes) to determine optimal binding conditions. To capture a reaction intermediate, native UnBG11 crystals were soaked in 300 mM cellobiose for 7 minutes. Additionally, native UnBG11 crystals were soaked in 1 M glucose for 3 minutes to identify conserved glucose binding sites within the active site. And thio-cellobiose soaking was performed to find inhibitory sites of UnBG11. Crystals were carefully handled to minimize physical damage during transfer and soaking procedures. Same protocol was followed to capture glucose bound structures of UnBG11 variants. The soaking conditions are mentioned in Supplementary Table S2.

## X-ray diffraction data collection and processing

Diffraction data were collected by rotation method under liquid nitrogen cryo-conditions at 100 K, for that suitable cryoprotectants were prepared in mother liquor. Cryoprotectants used for freezing the crystals are mention in Supplementary Table S2. Crystals were carefully transferred into their respective cryoprotectants, mounted on a goniometer head, and flash-cooled under a liquid nitrogen stream at 100 K.

For glucose soaked native UnBG11, UnBG11\_C188V, UnBG11\_H261W, cellobiose-soaked native UnBG11, and thio-cellobiose-soaked native UnBG11 crystals, diffraction datasets were collected using CuK $\alpha$  X-ray radiation generated by a Rigaku Micromax 007HF generator equipped with an R-Axis IV++ detector at the Protein Crystallography Facility, IIT Bombay, India. Apo native UnBG11, UnBG11\_C188V, UnBG11\_H261W and UnBG11\_C188V\_H261W, glucose soaked UnBG11\_C188V\_H261W and cellobiose soaked UnBG11\_H261W crystals were frozen in liquid nitrogen. The diffraction data of these crystals were collected using a synchrotron radiation X-ray source at beamline PX-BL21, RRCAT, Indore, India. The diffraction datasets of cellobiose soaked native UnBG11 and glucose soaked UnBG11\_H261W crystals were processed using Xia2<sup>51</sup> and Dials<sup>52–54</sup> from CCP4Cloud Remote. Diffraction datasets of all other UnBG11 variants were integrated using *XDS* and scaled using *XSCALE*<sup>55</sup>. The intensities were converted to structure factors with *F2MTZ* and *CAD* from *CCP4*<sup>54</sup>. Data collection statistics are represented in Table 1.

## Structure determination, model building and refinement

The native apo structure of UnBG11 was determined using the molecular replacement (MR) method in *PHASER*<sup>56</sup>. The structure of a  $\beta$ -glucosidase from compost metagenome, Td2F2 (PDB ID. 3WH5)<sup>57</sup> having 52.8 % amino acid sequence identity with UnBG11, was used as a search model to compute initial phases. Matthews coefficient ( $2.23 \text{ \AA}^3 \text{ Da}^{-1}$ )<sup>58</sup> indicated one UnBG11 molecule in asymmetric unit. After finding the correct solution, the template model was refined for 10 cycles using *REFMAC5*<sup>59</sup>. Initial model building was done in *COOT*<sup>60</sup>. Solvent molecules and ions were incorporated into the structure by identifying peaks in  $\sigma$ -A weighted  $F_o - F_c$  electron density maps exceeding  $3\sigma$ . These additions were carefully monitored to ensure a progressive reduction in  $R_{free}$  and improvement of the overall stereochemistry. Final refinement was done by PDB-redo which automates model rebuilding and validation for optimal stereochemical and structural parameters<sup>61</sup>.

Structures of other variants of UnBG11 were solved using apo native UnBG11 (PDB ID. 9JLZ) as a template for molecular replacement. Ligand bound structures (including

complexes with cellobiose, thio-glucose, glucose, and glycosyl intermediates) were resolved using rigid body refinement to derive initial phases. Following this, the protein components of the structures were subjected to initial refinement cycles to enhance the electron density maps of the ligands (cellobiose, thio-glucose, glucose, and glycosyl intermediates). Subsequently, the ligands were modelled into their corresponding  $\sigma$ -A weighted  $F_o - F_c$  electron density maps, followed by additional cycles of refinement. The waters and other solvent molecules were added to the structures, and alternative conformations of residues were built using *COOT*. The refinement processes were rigorously monitored by assessing the reduction in  $R_{free}$  values and ensuring that the overall stereochemical quality improved consistently throughout. The final refinement statistics, including completeness,  $R_{work}$ ,  $R_{free}$  and stereochemical parameters for all structures presented in this study, are reported in Table 1.

## Acknowledgements

We thank Dr. Ravindra Makde, Dr. Ashwani Kumar and Dr. Biplab Ghosh at the PX-BL21 beamline (BARC) at Indus-2, RRCAT, Indore, India for their support in diffraction data collection. We acknowledge the Protein Crystallography Facility at IIT Bombay. We also thank Chinmay K. Kamale and Barnava Banerjee for proof reading the manuscript. The research was supported by research funding to PB from Department of Biotechnology, Government of India (grant number: BT/PR41982/PBD/26/822/2021).

## Author contributions

ABS, RKB and PB conceived the project. ABS, RKB collected the X-ray diffraction data. ABS, PB processed and refined the data and build model. ABS, RKB and MP designed and performed the experiments. ABS and PB wrote the manuscript with inputs from RKB, MP and SN

## Competing interests

The authors declare no competing interests



## References

1. Naik, S. N., Goud, V. V., Rout, P. K. & Dalai, A. K. Production of first and second generation biofuels: A comprehensive review. *Renew. Sustain. Energy Rev.* **14**, 578–597 at <https://doi.org/10.1016/j.rser.2009.10.003> (2010).
2. Zabed, H., Sahu, J. N., Suely, A., Boyce, A. N. & Faruq, G. Bioethanol production from renewable sources: Current perspectives and technological progress. *Renew. Sustain. Energy Rev.* **71**, 475–501 at <https://doi.org/10.1016/j.rser.2016.12.076> (2017).
3. Yuan, X. & Cheng, G. From cellulose fibrils to single chains: Understanding cellulose dissolution in ionic liquids. *Phys. Chem. Chem. Phys.* **17**, 31592–31607 at <https://doi.org/10.1039/c5cp05744b> (2015).
4. Chen, H. L. *et al.* A highly efficient  $\beta$ -glucosidase from the buffalo rumen fungus *Neocallimastix patriciarum* W5. *Biotechnology for Biofuels* vol. 5 at <https://doi.org/10.1186/1754-6834-5-24> (2012).
5. Sun, J., Wang, W., Ying, Y. & Hao, J. A Novel Glucose-Tolerant GH1  $\beta$ -Glucosidase and Improvement of Its Glucose Tolerance Using Site-Directed Mutation. *Applied Biochemistry and Biotechnology* vol. 192 999–1015 at <https://doi.org/10.1007/s12010-020-03373-z> (2020).
6. Méndez-Líter, J. A. *et al.* A novel, highly efficient  $\beta$ -glucosidase with a cellulose-binding domain: Characterization and properties of native and recombinant proteins. *Biotechnology for Biofuels* vol. 10 at <https://doi.org/10.1186/s13068-017-0946-2> (2017).
7. Ariaeenejad, S. *et al.* A Novel High Glucose-Tolerant  $\beta$ -Glucosidase: Targeted Computational Approach for Metagenomic Screening. *Frontiers in Bioengineering and Biotechnology* vol. 8 at <https://doi.org/10.3389/fbioe.2020.00813> (2020).
8. Srivastava, N. *et al.* Microbial beta glucosidase enzymes: Recent advances in biomass conversation for biofuels application. *Biomolecules* **9**, 1–23 at <https://doi.org/10.3390/biom9060220> (2019).
9. Singhania, R. R., Patel, A. K., Sukumaran, R. K., Larroche, C. & Pandey, A. Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production. *Bioresour. Technol.* **127**, 500–507 at <https://doi.org/10.1016/j.biortech.2012.09.012> (2013).
10. Hodge, D. B., Karim, M. N., Schell, D. J. & McMillan, J. D. Model-based fed-batch for high-solids enzymatic cellulose hydrolysis. *Appl. Biochem. Biotechnol.* **152**, 88–107 (2009).
11. Krogh, K. B. R. M. *et al.* Characterization and kinetic analysis of a thermostable GH3  $\beta$ -glucosidase from *Penicillium brasilianum*. *Applied Microbiology and Biotechnology* vol. **86** 143–154 at <https://doi.org/10.1007/s00253-009-2181-7> (2010).
12. Yang, Y. *et al.* A mechanism of glucose tolerance and stimulation of GH1  $\beta$ -glucosidases. *Sci. Rep.* **5**, 1–12 at <https://doi.org/10.1038/srep17296> (2015).
13. De Giuseppe, P. O. *et al.* Structural basis for glucose tolerance in GH1  $\beta$ -glucosidases. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 1631–1639 at

- <https://doi.org/10.1107/S1399004714006920> (2014).
14. Chen, M. *et al.* Isolation and characterization of a  $\beta$ -glucosidase from *Penicillium decumbens* and improving hydrolysis of corn cob residue by using it as cellulase supplementation. *Enzyme Microb. Technol.* **46**, 444–449 at <https://doi.org/10.1016/j.enzmictec.2010.01.008> (2010).
  15. Gao, L., He, R., Zong, Z. & Zhang, D. A  $\beta$ -glucosidase hyperproducing strain, *Penicillium piceum*: Novel characterization of lignocellulolytic enzyme systems and its application in biomass bioconversion. *Fungal Cellulolytic Enzymes: Microbial Production and Application* 81–106 at [https://doi.org/10.1007/978-981-13-0749-2\\_5](https://doi.org/10.1007/978-981-13-0749-2_5) (2018).
  16. Bai, A., Zhao, X., Jin, Y., Yang, G. & Feng, Y. A novel thermophilic  $\beta$ -glucosidase from *Caldicellulosiruptor bescii*: Characterization and its synergistic catalysis with other cellulases. *Journal of Molecular Catalysis B: Enzymatic* vols 85–86 248–256 at <https://doi.org/10.1016/j.molcatb.2012.09.016> (2013).
  17. Singh, A. *et al.* Production of bioethanol from food waste: Status and perspectives. *Bioresour. Technol.* **360**, 127651 at <https://doi.org/10.1016/j.biortech.2022.127651> (2022).
  18. Santos, C. A. *et al.* An engineered GH1  $\beta$ -glucosidase displays enhanced glucose tolerance and increased sugar release from lignocellulosic materials. *Scientific Reports* vol. 9 at <https://doi.org/10.1038/s41598-019-41300-3> (2019).
  19. Rojas, A. & Romeu, A. A sequence analysis of the  $\beta$ -glucosidase sub-family B. *FEBS Letters* vol. 378 93–97 at [https://doi.org/10.1016/0014-5793\(95\)01412-8](https://doi.org/10.1016/0014-5793(95)01412-8) (1996).
  20. Shewale, J. G.  $\beta$ -Glucosidase: Its role in cellulase synthesis and hydrolysis of cellulose. *Int. J. Biochem.* **14**, 435–443 at [https://doi.org/10.1016/0020-711x\(82\)90109-4](https://doi.org/10.1016/0020-711x(82)90109-4) (1982).
  21. Cairns, J. R. K. & Esen, A.  $\beta$ -Glucosidases. *Cell. Mol. Life Sci.* **67**, 3389–3405 <https://doi.org/10.1007/s00018-010-0399-2> (2010).
  22. Isorna, P. *et al.* Crystal Structures of *Paenibacillus polymyxa*  $\beta$ -Glucosidase B Complexes Reveal the Molecular Basis of Substrate Specificity and Give New Insights into the Catalytic Machinery of Family I Glycosidases. *J. Mol. Biol.* **371**, 1204–1218 at <https://doi.org/10.1016/j.jmb.2007.05.082> (2007).
  23. Rye, C. S. & Withers, S. G. Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* **4**, 573–580 at [https://doi.org/10.1016/s1367-5931\(00\)00135-6](https://doi.org/10.1016/s1367-5931(00)00135-6) (2000).
  24. Dong, S. *et al.* Structural insight into a GH1  $\beta$ -glucosidase from the oleaginous microalga, *Nannochloropsis oceanica*. *Int. J. Biol. Macromol.* **170**, 196–206 <https://doi.org/10.1016/j.ijbiomac.2020.12.128> (2021).
  25. Jeng, W. Y. *et al.* Structural and functional analysis of three  $\beta$ -glucosidases from bacterium *Clostridium cellulovorans*, fungus *Trichoderma reesei* and termite *Neotermes koshunensis*. *J. Struct. Biol.* **173**, 46–56 at <https://doi.org/10.1016/j.jsb.2010.07.008> (2011).
  26. Kempton, J. B. & Withers, S. G. Mechanism of *Agrobacterium*  $\beta$ -Glucosidase: Kinetic Studies. *Biochemistry* **31**, 9961–9969 at <https://doi.org/10.1021/bi00156a015> (1992).

27. Lu, J., Du, L., Wei, Y., Hu, Y. & Huang, R. Expression and characterization of a novel highly glucose-tolerant  $\beta$ -glucosidase from a soil metagenome. *Acta Biochim. Biophys. Sin. (Shanghai)*. **45**, 664–673 at <https://doi.org/10.1093/abbs/gmt061> (2013).
28. Jayasekara, S. & Ratnayake, R. Microbial Cellulases: An Overview and Applications. in *Cellulose* vol. 4 885–892 at <https://doi.org/10.5772/intechopen.84531> (IntechOpen, 2019).
29. E. Montero *et al.* NMR studies of the conformation of thiocellobiose bound to a  $\beta$ -glucosidase from *Streptomyces* sp. *FEBS Letter* **421**, 243–248 at [https://doi.org/10.1016/S0014-5793\(97\)01571-8](https://doi.org/10.1016/S0014-5793(97)01571-8) (1998).
30. Opassiri, R. *et al.* Analysis of rice glycosyl hydrolase family 1 and expression of Os4bglu12  $\beta$ -glucosidase. *BMC Plant Biol.* **6**, 1–19 at <https://doi.org/10.1186/1471-2229-6-33> (2006).
31. Liberato, M. V. *et al.* Molecular characterization of a family 5 glycoside hydrolase suggests an induced-fit enzymatic mechanism. *Sci. Rep.* **6**, 1–16 at <https://doi.org/10.1038/srep23473> (2016).
32. Kulkarni, T. S. *et al.* Crystal structure of  $\beta$ -glucosidase 1A from *Thermotoga neapolitana* and comparison of active site mutants for hydrolysis of flavonoid glucosides. *Proteins Struct. Funct. Bioinforma.* **85**, 872–884 at <https://doi.org/10.1002/prot.25256> (2017).
33. Pinaki, D., Lhakpa, W. & Joginder, S. Simultaneous saccharification and fermentation (SSF), an efficient process for bio-ethanol production: An overview. *Biosci. Biotechnol. Res. Asia* **12**, 87–100 at <http://dx.doi.org/10.13005/bbra/1639> (2015).
34. Yi-Rui Y. *et al.* Expression and Characteristics of Two Glucose-Tolerant GH1  $\beta$ -glucosidases From *Actinomadura amylolytica* YIM 77502T for Promoting Cellulose Degradation. *Front. Microbiol.* **9**, 3149 at <https://doi.org/10.3389/fmicb.2018.03149> (2018).
35. Jinhu W. *et al.* QM-MM studies on the glycosylation mechanism of rice BGlu1  $\beta$ -glucosidase. *Journal of Molecular Graphics and Modelling* **30**, 148–152 at <https://doi.org/10.1016/j.jmgm.2011.06.012> (2011).
36. Badieyan, S., Bevan, D. R. & Zhang, C. Probing the active site chemistry of  $\beta$ -glucosidases along the hydrolysis reaction pathway. *Biochemistry* vol. 51 8907–8918 at <https://doi.org/10.1021/bi300675x> (2012).
37. Teugjas, H. & Våljamäe, P. Selecting  $\beta$ -glucosidases to support cellulases in cellulose saccharification. *Biotechnol. Biofuels* **6**, 1–13 at <https://doi.org/10.1186/1754-6834-6-105> (2013).
38. Chen, F. *et al.* A Novel Cold-Adaptive Endo-1,4- $\beta$ -Glucanase From *Burkholderia pyrrocinia* JK-SH007: Gene Expression and Characterization of the Enzyme and Mode of Action. *Frontiers in Microbiology* vol. **10** at <https://doi.org/10.3389/fmicb.2019.03137> (2020).
39. Salam, M. A. *et al.* The Effect Of Glucose , Temperature And Ph On Bioethanol Production By *Saccharomyces Cerevisiae*. *J. Popul. Ther. Clin. Pharmacol.* **31**, 656–667 at <https://doi.org/10.53555/jptcp.v31i1.4174> (2024).
40. Sinha, S. K., Prakash Reddy, K. & Datta, S. Understanding the glucose tolerance of an

- archaeon  $\beta$ -glucosidase from *Thermococcus* sp. *Carbohydrate Research* vol. **486** at <https://doi.org/10.1016/j.carres.2019.107835> (2019).
41. Goswami, S., Das, S. & Datta, S. Understanding the role of residues around the active site tunnel towards generating a glucose-tolerant  $\beta$ -glucosidase from *Agrobacterium tumefaciens* 5A. *Protein Eng. Des. Sel.* **30**, 523–530 at <https://doi.org/10.1093/protein/gzx039> (2017) (2017).
42. Mariano, D. *et al.* Glutanthase: A database for improving the rational design of glucose-tolerant  $\beta$ -glucosidases. *BMC Molecular and Cell Biology* vol. **21** at <https://doi.org/10.1186/s12860-020-00293-y> (2020).
43. Salgado, J. C. S., Meleiro, L. P., Carli, S. & Ward, R. J. Glucose tolerant and glucose stimulated B-glucosidases – A review. *Bioresour. Technol.* **267**, 704–713 at <https://doi.org/10.1016/j.biortech.2018.07.137> (2018).
44. Yang, Y. *et al.* A mechanism of glucose tolerance and stimulation of GH1  $\beta$ -glucosidases. *Sci. Rep.* **5**, 1–12 at <https://doi.org/10.1038/srep17296> (2015).
45. Kamale, C., Rauniyar, A. & Bhaumik, P. Rational design paving the way for improving glucose tolerance and catalytic properties of a  $\beta$ -glucosidase from *Acetivibrio thermocellus*. *bioRxiv* 1–42 at <https://www.biorxiv.org/content/10.1101/2024.08.05.606605> (2024).
46. Shaker, G. & Swift, C. J. Peroxidase-Coupled Glucose Method. *StatPearls* 2–8 (2023).
47. Qi, G. *et al.* Extraction and characterization of wax from sugarcane bagasse and the enzymatic hydrolysis of dewaxed sugarcane bagasse. *Prep. Biochem. Biotechnol.* **47**, 276–281 at <https://doi.org/10.1080/10826068.2016.1224246> (2017).
48. Sun, J. X., Sun, X. F., Zhao, H. & Sun, R. C. Isolation and characterization of cellulose from sugarcane bagasse. *Polym. Degrad. Stab.* **84**, 331–339 at <https://doi.org/10.1016/j.polymdegradstab.2004.02.008> (2004).
49. Hassell, A. M. *et al.* Crystallization of protein-ligand complexes. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **63**, 72–79 at <https://doi.org/10.1107/S0907444906047020> (2006).
50. Collins, P. M. *et al.* Gentle, fast and effective crystal soaking by acoustic dispensing. *Acta Crystallogr. Sect. D Struct. Biol.* **73**, 246–255 at <https://doi.org/10.1107/S205979831700331X> (2017).
51. Winter, G. Xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
52. Winter, G. *et al.* DIALS: Implementation and evaluation of a new integration package. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 85–97 at <https://doi.org/10.1107/S2059798317017235> (2018).
53. Beilsten-Edmands, J. *et al.* Scaling diffraction data in the DIALS software package: Algorithms and new approaches for multi-crystal scaling. *Acta Crystallogr. Sect. D Struct. Biol.* **76**, 385–399 at <https://doi.org/10.1107/S2059798320003198> (2020).
54. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242 at <https://doi.org/10.1107/S0907444910045749> (2011).

55. Kabsch, W. *et al.* XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 at <https://doi.org/10.1107/S0907444909047337> (2010). (2010).
56. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 at <https://doi.org/10.1107/S002188980702120> (2007).
57. Matsuzawa, T. *et al.* Crystal structure and identification of a key amino acid for glucose tolerance, substrate specificity, and transglycosylation activity of metagenomic  $\beta$ -glucosidase Td2F2. *FEBS J.* **283**, 2340–2353 (2016).
58. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497 at [https://doi.org/10.1016/0022-2836\(68\)90205-2](https://doi.org/10.1016/0022-2836(68)90205-2) (1968).
59. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 at <https://doi.org/10.1107/S0907444911001314> (2011).
60. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 at <https://doi.org/10.1107/S0907444904019158> (2004).
61. Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB-REDO server for macromolecular structure model optimization. *IUCrJ* **1**, 213–220 at <https://doi.org/10.1107/S2052252514009324> (2014).