Structural basis of EHEP-mediated offense against phlorotannin-induced defense from brown algae to protect *aku*BGL activity

Xiaomei Sun[1,¶], Yuxin Ye[1,#a,¶], Naofumi Sakurai[1], Hang Wang[1], Koji Kato[1,#b], Jian Yu[1], Keizo Yuasa[2,#c], Akihiko Tsuji[2], Min Yao[1,*]

[1] Faculty of Advanced Life Science, Hokkaido University, Sapporo, Japan

[2] Graduate School of Bioscience and Bioindustry, Tokushima University, Tokushima, Japan

[#a] Current address: Pingshan translational medicine center, Shenzhen Bay Laboratory, Shenzhen, China

[#b] Current address: Structural Biology Division, Japan Synchrotron Radiation Research Institute (JASRI), Hyogo, Japan

[#c] Current address: Faculty of Science and Engineering, Setsunan University, Neyagawa, Japan

*Corresponding author

E-mail: yao@castor.sci.hokudai.ac.jp (MY)

¶ These authors contributed equally to this work.

Abstract

The defensive-offensive associations between algae and herbivores determine marine ecology. Brown algae utilize phlorotannin as their chemical defense against the predator *Aplysia kurodai*, which uses β-glucosidase (*aku*BGL) to digest the laminarin in algae to glucose. Moreover, *A. kurodai* employs *Eisenia* hydrolysis-enhancing protein (EHEP) as an offense to protect *aku*BGL activity from phlorotannin inhibition by precipitating phlorotannin. To underpin the molecular mechanism of this digestive-defensive-offensive system, we determined the structures of apo and tannic-acid (TNA, a phlorotannin-analog) bound form of EHEP, as well as *aku*BGL. EHEP consisted of three peritrophin-A domains formed in a triangle and bound TNA in the center without significant conformational changes. Structural comparison between EHEP and EHEP–TNA led us to find that EHEP can be resolubilized from phlorotannin-precipitation at an alkaline pH, which reflects a requirement in the digestive tract. *aku*BGL contained two GH1 domains, only one of which conserved the active site. Combining docking analysis, we propose the mechanisms by which phlorotannin inhibits *aku*BGL by occupying the substrate-binding pocket, and EHEP protects *aku*BGL against the inhibition by binding with phlorotannin to free the *aku*BGL pocket.

Introduction

Over millions of years of evolution, predators have successfully coevolved with their prey to maintain an ecological balance [1]. In marine habitats, interactions between algae and marine herbivores dominate marine ecosystems. Most algae are consumed by marine herbivores [2]. They produce secondary metabolites as a chemical defense to protect themselves against predators. The sea hare *Aplysia kurodai*, a marine gastropod, preferentially feeds on the laminarin-abundant brown algae *Eisenia bicyclis* (laminarin constitutes 20%–30% of the dry weight of *E. bicyclis* and acts as a major storage carbohydrate), releasing large amounts of glucose through hydrolysis mediated by 110 and 210 kDa β-glucosidases (*aku*BGLs). Interestingly, such a feeding strategy has attracted attention for producing glucose as a renewable biofuel source [3]. However, to protect themselves against predators, brown algae produce phlorotannin, a secondary metabolite, thereby reducing the digestion of *A. kurodai* by inhibiting the hydrolytic activity of *aku*BGLs. This inhibition has a negative impact on the application of brown algae for producing renewable biofuel sources. As the 110 kDa *aku*BGL is more sensitive to phlorotannin than the 210 kDa BGL [4], we focused on the 110 kDa *aku*BGL in this study (hereafter, *aku*BGL refers to 110 kDa *aku*BGL).

To counteract the antipredator adaptations of algae, herbivores use diverse approaches, such as detoxification, neutralization, defense suppression, and physiological adaptations [5]. *A. kurodai* inhibits the phlorotannin-defense of brown algae through *Eisenia* hydrolysis-enhancing protein (EHEP), a protein from their digestive system that protects *aku*BGL activity from phlorotannin inhibition [4]. Previous studies have shown that incubating *E. bicyclis* with *aku*BGL in the presence of EHEP results in increased glucose production because EHEP binds to phlorotannin and forms an insoluble complex [4].

The *aku*BGL–phlorotannin/laminarin–EHEP system exemplifies the digestion process of *A. kurodai* as well as the defense and antidefense strategies between *E.*

bicyclis and *A. kurodai*. Although the defense/antidefense strategy has been established, the detailed molecular mechanism of this interplay remains unknown. Further, phlorotannin inhibition hinders the potential application of brown algae as feedstocks for enzymatically producing biofuel from laminarin. Thus, understanding the underlying molecular mechanisms will be beneficial for the application of this system in the biofuel industry.

Despite the potential use of laminarin hydrolytic enzymes in the biofuel industry, only a few BGLs of glycoside hydrolases belonging to the GH3 and GH1 family are known to hydrolyze laminarin (e.g., *Talaromyces amestolkiae* BGL[6], *Ustilago esculenta* BGL [7], and V*ibrio campbellii* BGL [8] from the GH3 family and *Saccharophagus degradans* 2-40^T BGL [9] from the GH1 family). GH3 is a multidomain enzyme family characterized by N-terminal $(\beta/\alpha)_8$ (NTD) and C-terminal $(\beta/\alpha)_6$ (CTD) domains, with or without auxiliary domains [10]; the nucleophile aspartate and the acid/base glutamate residues exist in the NTD and CTD, respectively. In contrast, the members of the GH1 family generally share a single $(\beta/\alpha)_8$-fold domain (hereafter referred to as GH1 domain [GH1D]), and the two glutamic acid catalytic residues are located in the carboxyl termini of β-strands 4 and 7. Therefore, the two families may use different substrate-recognition and catalytic mechanisms for laminarin. Intriguingly, although *aku*BGL possesses laminarin hydrolytic activity and belongs to the GH1 family, its molecular weight is considerably greater than that of other GH1 members. Sequence analysis has indicated that *aku*BGL consists of ≥2 GH1Ds. Because no structural information of BGL active on polysaccharides is available, the catalytic mechanism toward laminarin is unclear.

There is limited information on EHEP, a novel cysteine-rich protein (8.2% of the amino acid content), because no structural or functional homologous protein exists in other organisms. EHEP was predicted to consist of three peritrophin-A domains (PADs) with a cysteine-spacing pattern of $CX_{15}CX_5CX_9CX_{12}CX_{5\text{-}9}C$. The PADs consist of peritrophic matrix proteins, which have been proposed to play an important role in

detoxifying ingested xenobiotics [11]. For instance, *Aedes aegypti* intestinal mucin 1 (*Ae*IMUC1) consists of a signal peptide followed by three PADs with an intervening mucin-like domain; its expression is induced by blood feeding. *Ae*IMUC1-mediated blood detoxification during digestion is completed by binding to toxic heme molecules [12]. Despite the similar domain organization of EHEP and *Ae*IMUC1, their function and binding partner are completely different, implying their different characteristics. However, the characteristics of the EHEP-phlorotannin insoluble complex remain unknown; moreover, it remains unclear why and how EHEP protects *aku*BGL from phlorotannin inhibition.

In this study, we determined the structures of apo and tannic acid (TNA, phlorotannin-analog) bound-form of EHEP, as well as *aku*BGL; all isolated from *A. kurodai*. The structure of EHEP consists of three PADs arranged in a triangle shape, with TNA bound at the surface of the triangle center. A structural comparison of EHEP and EHEP–TNA revealed no significant changes in conformation upon TNA binding, implying that EHEP maintains its structure when precipitated with TNA. Then, we found the conditions to resolubilize EHEP–TNA precipitate for EHEP recycling. The obtained *aku*BGL structure suggests that only one GH1D (GH1D2) possesses laminarin hydrolytic activity; subsequently, ligand-docking experiments demonstrated that TNA/phlorotannin has a higher binding affinity than laminarin. Our results revealed the mechanisms by which EHEP protects *aku*BGL from phlorotannin inhibition and phlorotannin inhibits the hydrolytic activity of *aku*BGL, providing structural support for the potential application of brown algae for biofuel production.

Results

Effects of TNA on akuBGL activity with or without EHEP

As a chemical defense metabolite of brown algae, phlorotannins are a type of

tannins. It is difficult to isolate a compound from phlorotannins because they are a group of polyphenolic compounds with different sizes and varying numbers of phloroglucinol units [13]. Previous studies have reported that the phlorotannin-analog TNA has a comparable inhibition effect on *aku*BGL to that of phlorotannin [4]. Hence, we used TNA instead of a phlorotannin to explore phlorotannin binding with EHEP and *aku*BGL. First, we confirmed TNA inhibition of *aku*BGL activity and clarified the protective effects of EHEP from TNA inhibition. The inhibition experiments showed that the galactoside hydrolytic activity of *aku*BGL decreased with increasing TNA concentration, indicating that TNA inhibits *aku*BGL activity in a dose-dependent manner (Fig. 1A). Approximately 70% *aku*BGL activity was inhibited at a TNA concentration of 40 μM. Moreover, protection ability analysis revealed that EHEP protects *aku*BGL activity from TNA inhibition in a dose-dependent manner, as indicated by the recovery of the inhibited *aku*BGL activity with increasing EHEP concentration (Fig. 1B). Further, approximately 80% of *aku*BGL activity was recovered at an EHEP concentration of 3.36 μM.

Overall structure of EHEP

Considering the lack of known homologous proteins of EHEP, we determined the structure of natural EHEP using the native-SAD method at a resolution of 1.15 Å, with a R_{work} and R_{free} of 0.18 and 0.19, respectively (Table 1). The residues A21–V227 of A21–K229 in purified EHEP (1–20 aa were cleaved during maturation) were appropriately visualized, whereas two C-terminal residues were disordered. The structure of EHEP consists of three PADs: PAD1 (N24–C79), PAD2 (I92–C146), and PAD3 (F164–C221), which are linked by two long loops, LL1 (Q80–N91) and LL2 (R147– G163), and arranged in a triangle shape (Fig. 2A). These three PADs share a similar structure, with a root-mean-squared difference (RMSD) of 1.065 Å over 46 Cα atoms and only ~20.3% sequence identity (Figs. 2B and 2C). The three PADs share a canonical CBM14 fold consisting of two β-sheets containing three N-terminal and two

C-terminal antiparallel β-strands. Additionally, two small α-helices were appended to the N- and C-terminus in PAD1 and PAD3 but not in PAD2 (Fig. 2B).

Although the Dali server [14] did not provide similar structures using the overall structure of EHEP as the search model, six structures showed similarities with a single PAD of EHEP. These structures were the members of the PAD family, including the chitin-binding domain of chitotriosidase (PDB ID 5HBF) [15], avirulence protein 4 from *Pseudocercospora fuligena* [*Pf*Avr4 (PDB ID 4Z4A)] [16] and *Cladosporium fulvum* [*Cf*Avr4 (PDBID 6BN0)] [17], allergen Der p 23 (PDB ID 4ZCE) [18], tachytitin (PDB ID 1DQC) [19], and allergen Blot 12 (PDB ID 2MFK), with Z-scores of 4.7–8.4, RMSD values of 1.2–2.8 Å, and sequence identity of 19%–37%. The highest sequence disparity was detected in PAD2, whereas the greatest structural differences were noted in PAD3. The $C^{No1}X_{15}C^{No2}X_5C^{No3}X_9C^{No4}X_{12}C^{No5}X_{5\text{-}9}C^{No6}$ motif (superscripts and subscripts indicate the cysteine number and number of residues between adjacent cysteines, respectively) in each PAD of EHEP formed three disulfide bonds between the following pairs: C^{No1}–C^{No3}, C^{No2}–C^{No6}, and C^{No4}–C^{No5} (Fig. 2B). Such rich disulfide bonds may play a folding role in the structural formation of EHEP, with >70% loop conformation. A similar motif with disulfide bonds was observed in tachycitin [19], *Pf*Avr4 [16], *Cf*Avr4 [17], and the chitin-binding domain of chitinase [15]. Although these proteins share a highly conserved core structure, they have different biochemical characteristics. For example, the chitin-binding domain of human chitotriosidase Avr4 and tachycitin possess chitin-binding activity, but the critical residues for chitin-binding are not conserved [15, 17, 20], indicating that they employ different binding mechanisms. In contrast, EHEP and allergen Der p 23 do not possess chitin-binding activity [4, 18]. Thus, the PAD family may participate in several biochemical functions.

Modification of EHEP

Among our structures, the apo structure2 (1.4 Å resolution) clearly showed that the cleaved N-terminus of Ala21 underwent acetylation (Fig. 3A), consistent with the

molecular weight results obtained using MALDI–TOF MS [21]. N-terminal acetylation is a common modification in eukaryotic proteins. Such acetylation is associated with various biological functions, such as protein half-life regulation, protein secretion, protein–protein interaction, protein–lipid interaction [22], metabolism, and apoptosis [23]. Further, N-terminal acetylation may stabilize proteins [24]. To explore whether acetylation affects the protective effects of EHEP on *aku*BGL, we measured the TNA-precipitating assay of recombEHEP (A21–K229) without acetylation. The results revealed that recombEHEP precipitated after incubation with TNA at a comparable level to that of natural EHEP (Fig. 3B), indicating that acetylation is not indispensable for the phlorotannin binding activity and stabilization of EHEP. Future studies are warranted to verify the exact role of N-terminal acetylation of EHEP in *A.kurodai*.

TNA binding to EHEP

To understand the mechanism by which TNA binds to EHEP, we determined the structure of EHEP complexed with TNA (EHEP–TNA) using the soaking method. In the obtained structure, both $2F_o–F_c$ and $F_o–F_c$ maps showed an electron density blob of 1,2,3,4,6-pentagalloylglucose, a core part of TNA missing the five external gallic acids (Fig. 4A, Fig. S1A). Previous studies have shown that acid catalytic hydrolysis of TNA requires a high temperature of 130°C [25]; even with a polystyrene-hollow sphere catalyst, a temperature of 80°C is required [26]. Therefore, the five gallic acids could not be visualized in the EHEP–TNA structure most likely due to the structural flexibility of TNA.

The apo EHEP and EHEP–TNA structures were extremely similar, with an RMSD value of 0.283 Å for 207 Cα atoms (Fig. S1B). However, the superposition of the two structures showed a decrement of the loop part in EHEP–TNA, indicating that EHEP is more stable when bound to TNA. TNA binding caused a slight increase in the α-helix and β-sheet contents of PAD2 and PAD3 (Fig. S1B). In the EHEP–TNA structure, the residues C93–Y96 of PAD2 folded into an α-helix and each β-sheet of the first β-strand

in PAD3 elongated by incorporating one residue in the first (G^{176}) and second β-sheets (S^{186}) and three residues in the third β-sheet ($H^{197}MP^{199}$).

The EHFP–TNA structure revealed that TNA binds to the center of the triangle formed by the three PADs, a positively charged surface (Fig. 4A and Fig. S1C). The binding pocket on EHEP surface was formed by the C-terminal α-helix of PAD1, the N-terminal α-helix of PAD2, and the middle part (loop) of PAD3 assisted by two long linker loops (LL1 and LL2). TNA was primarily bound to EHEP via hydrogen bonds and hydrophobic interactions (Fig. 4B). Gallic acid1, 4, and 6 interacted with EHEP via hydrogen bonds and additional hydrophobic contacts, whereas gallic acid2 and 3 only hydrophobically interacted with EHEP. The 3-hydroxyl groups of gallic acid1 and 6 individually formed a hydrogen bond with the main chain of G74 and the side chain of N75 in PAD1. The backbone carbonyl of Y96 and P199 in PAD2 and PAD3, respectively, formed a hydrogen bond with the 3,5-hydroxyl groups of gallic acid4. Additionally, some hydrogen bonds were formed between TNA and water molecules. TNA binding was also stabilized by hydrophobic interactions between the benzene rings of gallic acid and EHEP. For instance, gallic acid4 and 6 showed alkyl-π interaction with P77 and P201, respectively; moreover, gallic acid3 and 4 formed amide-π stackings with P199.

The EHEP–TNA structure clearly showed that TNA binds to EHEP without covalent bonds and the binding does not induce significant structural changes; thus, we attempted to recover EHEP from EHEP–TNA precipitates by adjusting the pH. As hypothesized, re-solubilization of the EHEP–phlorotannin precipitate is pH-dependent (Fig. 4C). The EHEP–TNA precipitate did not resolubilize at pH 7.0; however, after incubating for >1 h at pH 7.5, the precipitate started resolubilizing. Most of the precipitate rapidly resolubilized at an alkaline pH (≥8.0) after incubation for 15 min. Further, the resolubilized EHEP had the same elution profile as that of the natural EHEP (Fig. 4D) in SEC, suggesting that resolubilized EHEP maintained the native structure and its phlorotannin–precipitate activity (Fig. S1E).

Two domains of *aku*BGL

To reveal the structural basis of *aku*BGL recognition of laminarin and inhibition by TNA, we attempted to determine its structure. We soaked crystals in TNA as well as various substrate solutions but finally obtained the optimal resolution using crystal soaking in TNA. There was no blob of TNA in the electron density map of the obtained structure; thus, we considered this structure as the apo form of *aku*BGL.

Two *aku*BGL molecules were observed in an asymmetrical unit (MolA and MolB), without the N-terminal 25 residues (M1–D25), as confirmed using N-terminal sequencing analysis of purified natural *aku*BGL. This N-terminal fragment was predicted to be a signal peptide using the web server SignalP-5.0. The residues L26–P978 of L26–N994 were constructed in both MolA and MolB with glycosylation, whereas the remaining C-terminal residues (A979–M994) could not be visualized as they were disordered. The electron density map of Fo–Fc revealed N-glycosylation at three residues, i.e., N113, N212, and N645 (Figs. S2A, B). N-glycosylation of GH enzymes prevents proteolysis and increases thermal stability [27, 28]. Additionally, a study on β-glucosidase *Aspergillus terreus* BGL demonstrated that N-glycosylation of N224 affected the folding stability, even when it is located close to a catalytic residue [29]. In *aku*BGL, all N-glycosylation sites were present on the surface, far away from the catalytic site. Therefore, we speculate that *aku*BGL glycosylation does not affect its activity. Except for the difference in visualized glycans resulting from glycosylation, MolA and MolB were similar, with a RMSD value of 0.182 for 899 Ca atoms; therefore, we used MolA for further descriptions and calculations.

The structure of *aku*BGL consisted of two independent GH1 domains, GH1D1 (L26–T494) and GH1D2 (D513–P978), linked by a long loop (D495–Y512) (Fig. 5A). There was little interaction between GH1D1 and GH1D2, only in a buried surface area comprising 2% of the total surface (708.9 Å^2) (Fig. S2C). GH1D1 and GH1D2 have a sequence identity of 40.47% and high structural similarity with an RMSD value of 0.59

252 Å for 371 Cα atoms (Fig. S3A up).

253 Glucosidases of the GH1 family utilize the retaining mechanism with two glutamic
254 acids for catalyzing glucoside hydrolysis. In general, the distance between the two
255 catalytic oxygen atoms of the side chain of two glutamic acids is approximately 5 Å [30].
256 Sequence and structure alignment of GH1D1 and GH1D2 of *aku*BGL with other
257 members of the GH1 family revealed that the second glutamate is conserved (E404),
258 but the first glutamate is replaced by D192 in GH1D1. The oxygen atoms of the side
259 chains between D192 and E404 of GHD1 were 8.4 Å apart. In contrast, GH1D2
260 conserved two glutamic acids (E675 and E885) at the carboxyl termini of β-strands 4
261 and 7; the distance between oxygen atoms of E675 and E885 side chains was 5.1 Å
262 (Fig. S3A down), similar to that of *Neotermes koshunensis* BGL (*Nk*BGL) [31],
263 *Nannochloropsis oceanica* BGL (*No*BGL) [32], and *Spodoptera frugiperda* BGL (3.9–
264 4.9Å) [33]. Furthermore, regarding the two other conserved essential regions for β-
265 glucosidase activity, namely, glycone-binding site (GBS) and catalysis-related residues
266 (CR), GH1D1 conserved neither GBS nor CR, whereas GH1D2 conserved both (Fig.
267 5B). Altogether, we suggest that GH1D1 does not possess catalytic activity. We
268 expressed and purified the recombinant GH1D1, which did not show any hydrolytic
269 activity toward O-PNG (Figs. S2D, E), although we could not rule out the effect of N-
270 glycosylation.

271 A multi-GH1D assembly has been reported in β-glucosidase *Cj*CEL1A of *Corbicula*
272 *japonica* and glycosidase *LpM*DGH1 of the shipworm *Lyrodus pedicellatus*.
273 *Lp*MDGH1 has both exo- and endo-glucanase activity and is possibly implicated in
274 cellulose and hemicellulose digestion. *Cj*CEL1A has two tandem GH1Ds with a
275 sequence identity of 43.41% [34]. Two catalytic glutamic acids and the residues related to
276 substrate binding are conserved in the second GH1D, whereas the first GH1 domain
277 lacks these conserved residues and may play a role in folding the catalytic domain.
278 *LpM*DGH1 consists of six GH1Ds, among which GH1D2, 4, 5, and 6 contain the
279 conserved residues for activity, whereas others do not contain these residues and might

be involved in protein folding or substrate interactions [35].

Structural comparison of GH1D2 with other BGLs, including *Nk*BGL, rice (*Oryza sativa L.)* BGL (*Os*BGL), and microalgae *No*BGL, revealed the characteristics of each active pocket (Fig. S3B). *Os*BGL and *No*BGL have a deep, narrow, and straight pocket, whereas GH1D2 and *Nk*BGL have a broad and crooked pocket. Such active pocket shapes reflect the substrate preferences of *Os*BGL and *No*BGL; they hydrolyze laminaribiose with no detectable activity toward laminaritetraose [32, 36]. Furthermore, the difference in the features of large active pockets between *Nk*BGL and GH1D2, wherein GH1D2 often possesses an auxiliary site with several aromatic residues bound to the carbohydrate via CH-π interactions [37], may explain their substrate specificity. *Nk*BGL efficiently hydrolyzes laminaribiose and cellobiose but has weak hydrolytic activity toward laminarin [38]. In contrast, the GH1D2 of *aku*BGL has similar activity levels toward cellobiose and laminarin [39]. Therefore, the GH1D2 of *aku*BGL may recognize larger substrates than that of other BGLs. Laminarin typically has a curved conformation; accordingly, narrow- and straight-shaped pockets are incompatible for binding. Furthermore, we docked GH1D2 with laminaritetraose, wherein the four glucose units formed extensive contacts with GH1D2. Hydrogen bonds involved the catalytic residues E675 and E885. In addition, several aromatic residues, such as F677, F689, Y819, W857, and W935, formed π-π stacking (Fig. S4). Some interacting residues belonged to GBS and CR sites, such as E675, Y819, E885, and W935. Additionally, the docking structure revealed that the +4 glucose of laminaritetraose is located at the auxiliary binding site and that atom O1 of the +4 glucose is positioned outside the pocket (Fig. S4), implying that the auxiliary binding site with several aromatic residues (F677, W681, and F689) of GH1D2 facilitates laminarin binding.

Inhibitor binding of *aku*BGL

As we could not obtain the complex structure of *aku*BGL with TNA, we performed docking calculations of *aku*BGL GH1D2 with TNA to explore the inhibition

307 mechanism. The docking model of *aku*BGL–TNA showed that seven gallic acid rings
308 of TNA formed an extensive hydrogen bond network with *aku*BGL in the binding
309 pocket (Fig. 6). The hydroxyl groups of TNA formed hydrogen bonds with the residues
310 N552, E675, D735, K739, K759, Q840, T844, D852, and K859 of GH1D2. Moreover,
311 benzene rings showed hydrophobic interactions with several hydrophobic residues. In
312 particular, stable π-π stacking was observed between TNA and residues F547, W631,
313 F689, Y846, W857, and W935. Among these residues, the conserved E675 was the
314 catalytic residue, and W631, W935, and E934 contributed to GBS and CR sites.
315 In addition, we performed a docking calculation of GH1D2 with the characteristic
316 inhibitors eckol and phloroglucinol [40]. The binding mechanisms of eckol and
317 phloroglucinol were similar to those of TNA but with different contact residues (Fig.
318 S5). For eckol, the six hydroxyl groups formed hydrogen bonds with residues E675,
319 D735, E737, K759, E885, and E934. Additionally, residues W631, F677, F689, Y819,
320 W857, W935, F943, and W927 formed π-π stacking with eckol. For phloroglucinol, the
321 three hydroxyl groups formed hydrogen bonds with E675, E885, and E934, whereas
322 residues W631, F689, Y819, W857, W935, and W927 formed π-π stacking with the
323 benzene ring.
324 In summary, the three inhibitors inhibited *aku*BGL activity through similar binding
325 mechanisms to occupy the substrate-binding site, suggesting a reversible competitive
326 inhibition mechanism. The docking scores of the inhibitors TNA, eckol, and
327 phloroglucinol were −8.8, −7.3, and −5.7 kcal/mol, respectively, whereas the substrate
328 laminaritetraose had a docking score of −6.6 kcal/mol. TNA binding and phloroglucinol
329 had the highest and lowest negative docking score, respectively, indicating that TNA
330 has a higher binding affinity to *aku*BGL. This finding is consistent with that of a
331 previous study showing that phloroglucinol binding has a weaker inhibitory activity
332 than TNA [4].

Discussion

In marine habitats, the ecological interactions between brown algae and herbivores dominate marine ecosystems [41]. The *aku*BGL–phlorotannin/laminarin–EHEP system represents the feeding defense-offense associations between *A. kurodai* and brown algae. We focused on this system to understand the molecular mechanism at the atomic level. In contrast to most GH1 BGLs containing one catalytic GH1 domain, *aku*BGL consists of noncatalytic GH1D1 and catalytic GH1D2. The noncatalytic GH1D1 may act as a chaperone of GH1D2, as we successfully overexpressed GH1D1 but failed to do the same for GH1D2. A similar function has been suggested in *Cj*CEL1A [34] and *LpM*DGH1 [35].

BGLs have different substrate preferences in the degree of polymerization and type of glycosidic bond. In general, BGLs prefer to react with mono-oligo sugars over polysaccharides. For instance, *Os*BGL, *No*BGL, and *Nk*BGL hydrolyze disaccharides (cellobiose and laminaribiose) but display no or weak activity toward polysaccharides (cellulose and laminarin) [32, 36, 38]. The structure of GH1D2 explained the substrate preference for the polysaccharide laminarin. GH1D2 contains an additional auxiliary site composed of aromatic residues (Fig. S3B) in the substrate entrance pocket, which enables it to accommodate a long substrate, contributing to *aku*BGL activity toward laminarin, as supported by docking calculations (Fig. S4). In addition, docking analysis of *aku*BGL GH1D2 with inhibitors (TNA, phlorotannin, eckol, and phloroglucinol) revealed that these inhibitors bound to the substrate-binding site via hydrogen bonds and hydrophobic interactions similar to laminarin. Such binding mechanisms suggest the presence of competitive inhibition to occupy the binding site, consistent with previous research [4].

EHEP, expressed in the midgut of *A. kurodai*, was identified as an antidefense protein, protecting the hydrolysis activity of *aku*BGL from phlorotannin inhibition [4]. Such an ecological balance also exists between plants and their predator mammals and

insects. Similar to brown algae, plants use the toxic secondary metabolite tannins as their defense mechanism against predators, which constitute 5%–10% of dry weight of leaves. In vertebrate herbivores, tannins reduce protein digestion. In phytophagous insects, tannins may be oxidized at an alkaline pH of insect midgut and cause damage to cells. The evolution of plant-herbivore survival competition has led to the development of remarkably unique adaptation strategies. Mammals feeding on plants that contain tannin may overcome this defense by producing tannin-binding proteins, proline-rich proteins, and histatins [42]. Proline constitutes at least 20% of the total amino acid content in proline-rich proteins; for some species, the proportion of proline reaches 40%. Histidine constitutes 25% of total amino acid content in histatins. Both proline-rich proteins and histatins are nonstructural proteins in solution. In caterpillars, the oxidation damage of tannin is reduced by the low oxygen level. Some insects use the peritrophic membrane to transport tannins into the hemolymph, where they are excreted [43]. Additionally, the peritrophic envelop protects insects from tannins by forming an impermeable barrier to tannins [44]. *A. kurodai* uses a similar strategy with mammals by secreting the tannin-binding protein EHEP. Although EHEP has a completely different amino acid composition with proline-rich proteins and histatins, EHEP also binds to phlorotannin. Therefore, EHEP may be a specific counteradaptation that allows *A.kurodai* to feed on brown algae, as there are no homologous proteins in other organisms.

The three PADs of EHEP are arranged in a triangle shape, forming a large cavity on the surface at the triangle center to provide a ligand-binding site. Interestingly, EHEP–TNA crystal packing revealed that each TNA simultaneously binds to three EHEP molecules and crosslinks them together (Fig. S1D); this may be responsible for EHEP precipitation by TNA. EHEP has a positively charged surface at a pH of <6.0, whereas the surface becomes negatively charged at a pH of >7.0 (Fig. S1C). Meanwhile, TNA has a pKa of 4.9–8 [45-47], showing minor negative charges at an acidic pH and the highest negative charge at a pH of >7.0 [48]. Therefore, TNA binds to EHEP at a pH of <6.0 (pH

388 of crystallization = 4.5), but it shows charge repulsion with EHEP at a pH of >8.0.
389 Altogether, TNA is protonated and behaves as a hydrogen bond donor when the pH is
390 below its pKa, whereas when the pH is above its pKa, TNA is deprotonated and the
391 hydrogen bonding cannot be maintained. As losing hydrogen bonds and increasing
392 repulsive forces at a pH >8.0, the precipitated EHEP–TNA could not dissolve in the
393 buffer of pH > 8.0. This pH-induced reversible interaction also occurred in other
394 proteins, such as BSA, pepsin, and cytochrome C [49]. The phlorotannin members share
395 a similar structure with TNA; thus, we speculate that the EHEP–phlorotannin complex
396 also exhibits a pH-induced reversible interaction. *In vivo*, the pH of the digestive fluid
397 of *A. kurodai* is approximately 5.5, which favors the binding of EHEP to phlorotannin.
398 In the alkaline hindgut [50], the EHEP–phlorotannin disassociates (Fig. 7), and the
399 phlorotannin is subsequently excreted from the anus.

400 Based on the EHEP–TNA structure and docking models of *aku*BGL-
401 inhibitor/substrate, we proposed a mechanism of phlorotannin inhibition on *aku*BGL
402 activity and EHEP protection from phlorotannin inhibition (Fig. 7). Because laminarin
403 lacks the benzene rings essentially to forming π-stacking interactions with EHEP, the
404 EHEP can be considered no binding with the laminarin. In the absence of EHEP,
405 phlorotannin occupies the substrate-binding site of *aku*BGL, inhibiting the substrate
406 from entering the activity pocket and resulting in no glucose production. In the presence
407 of EHEP, it competitively binds to phlorotannin, freeing the *aku*BGL pocket. Then, the
408 substrate can enter the active pocket of *aku*BGL and glucose can be produced normally.
409 The digestive fluid of *A. kurodai* contains EHEP at a high concentration (>4.4 μM) [4],
410 which is slightly higher than the concentration of EHEP (3.36 μM) that protects
411 *aku*BGL activity (Fig. 1B). The high concentration of EHEP allows *A. kurodai* feeding
412 of phlorotannin-rich brown algae. The balance between phlorotannin inhibition and
413 protection is controlled by the concentrations of phlorotannin and EHEP *in vivo*.

414 The *aku*BGL–phlorotannin/laminarin–EHEP system is the digestive-defensive-
415 offensive associations between algae and herbivores. Our study presented the molecular

mechanism of this system at the atomic level, providing a molecular explanation for how the sea hare *A. kurodai* utilizes EHEP to protect *aku*BGL activity from phlorotannin inhibition. Further, such a feeding strategy has attracted attention for producing glucose as a renewable biofuel source, so our studies provide a molecular basis for the biofuel industry applications of brown algae.

Materials and Methods

EHEP and *aku*BGL preparation

Natural EHEP (22.5 kDa) and *aku*BGL (110 kDa) were purified from *A. kurodai* digestive fluid as described previously [4]. For crystallization, we further added one step purification of EHEP using size exclusion chromatography (HiLoad 16/60 Superdex 75, GE Healthcare), for which the column was equilibrated with 20 mM MES–NaOH buffer (pH 6.5). Obtained EHEP was then concentrated to 15–25 mg/mL using Vivaspin-4 10K columns (Sartorius, Göttingen, Germany). About *aku*BGL, we exchanged buffer from 20 mM Tris-HCl pH 7.0 to 20 mM Bis-tris pH 6.0 and concentrated it to 11 mg/mL using Amicon with a cutoff of 50 kDa.

To verify whether chemical modifications which was indicated by previous study (13) affect the function of EHEP, we prepared recombinant EHEP (recombEHEP) without the N-terminal signal peptide (1–20 aa) and chemical modifications [21]. EHEP cDNA was obtained via reverse transcription–polymerase chain reaction (RT–PCR) using the total RNA of *A. kurodai* as a template. The reamplified fragment was digested and ligated to a plasmid derived from pET28a (Novagene, Darmstadt, Germany). We

transformed the plasmid containing recombEHEP into *E. coli* B834(DE3) pARE2 and expressed it with a C-terminal hexahistidine-tag. The cells were cultured in lysogeny broth (LB) medium with the antibiotics kanamycin (25 mg/L) and chloramphenicol (34 mg/L) until the optical density at 600 nm (OD_{600}) reached 0.6. Subsequently, overexpression was induced by adding 0.5 mM isopropyl b-D-L-thiogalactopyranoside for 20 h at 20 ℃. After harvesting by centrifugation, the cells were resuspended in a buffer containing 50 mM Tris–HCl pH 7.4, 300 mM NaCl, DNase, and lysozyme and were disrupted via sonication. The insoluble part was removed by centrifugation for 30 min at 40000 ×*g* at 4 ℃. We loaded the supernatant onto a 5-mL Histrap HP column and the recombEHEP was eluted using increasing concentrations of imidazole (0–500 mM). The purified proteins were dialyzed against a solution containing 50 mM Tris–HCl pH 7.4 and 50 mM NaCl and subsequently loaded onto a Hitrap Q column and eluted by a linear gradient of a solution containing 50 mM Tris–HCl and 1 M NaCl. Fractions containing recombEHEP were concentrated and further purified using a gel filtration column (Hiload 16/60 superdex 75 pg) equilibrated with 20 mM sodium acetate pH 6.0 and 100 mM NaCl. We collected the fractions containing recombEHEP and concentrated them to 2.1 mg/mL using Amicon (Merck, American).

TNA binding assay for recombEHEP

We measured the binding activity of recombEHEP using precipitation analysis method,

as described previously [4]. Briefly, recombEHEP was incubated with TNA at 25 ℃ for 90 min and centrifuged for 10 min at 12000 ×*g* at 4 ℃. Then, we washed the precipitates twice and resuspended them in an SDS–PAGE loading buffer for binding analysis.

Effects of TNA on *aku*BGL activity with or without EHEP

Ortho-nitrophenyl-β-galactoside (ONPG) was used as a substrate to measure *aku*BGL activity. The reaction system (100μL) included 2.5 mM ONPG, 49 nM *aku*BGL, and different TNA concentrations in a reaction buffer (50 mM CH_3COONa pH 5.5, 100 mM NaCl, and 10 mM $CaCl_2$). After incubation for 10 min at 37 ℃, 100 μL of methanol was added to each sample to terminate the reaction. Then, the mixture was centrifuged for 10 min at 15000 ×*g* at 4 ℃ and the supernatant was used for analyzing *aku*BGL activity via HPLC. To measure the protective effect of EHEP on *aku*BGL, we added different amounts of EHEP to the reaction system.

Resolubilization of the EHEP–eckol precipitate

A mixture of 2 mg of EHEP and 0.4 mg of eckol was incubated at 37 °C for 1 h, followed by centrifugation at 12000 ×g for 10 min, and the supernatant was removed. The sediment was dissolved in 50 mM Tris–HCl buffer at different pH (7.0–9.0) and the absorbance at 560 nm was measured over time.

Crystallization and data collection

The crystallization, data collection, and initial phase determination of EHEP were

476 described previously [21]. As EHEP precipitates when bound to TNA, we could not
477 cocrystallize EHEP with TNA. Therefore, we used the soaking method to obtain the
478 EHEP–TNA complex. Owing to the poor reproducibility of EHEP crystallization, we
479 used a co-cage-1 nucleant [51] to prepare EHEP crystals for forming the complex with
480 TNA. Finally, we obtained high-quality EHEP crystals under the reservoir solution
481 containing 1.0 M sodium acetate, 0.1 M imidazole (pH 6.5) with co-cage-1 nucleant.
482 Subsequently, we soaked the EHEP crystals in a reservoir solution containing 10 mM
483 TNA at 37 ℃ for 2 days; further, they were maintained at 20 ℃ for 2 weeks. Next, we
484 soaked the EHEP crystals in a reservoir solution containing 10 mM phloroglucinol. For
485 data collection, the crystal was soaked in a cryoprotectant solution containing 20% (v/v)
486 glycerol along with the reservoir solution. Diffraction data were collected under a cold
487 nitrogen gas stream at 100 K using Photon Factory BL-17 (Tsukuba, Japan) or Spring
488 8 BL-41XU (Hyogo, Japan).
489 For *aku*BGL crystallization, the initial crystallization screening was performed using
490 the sitting-drop vapor-diffusion method with Screen classics and Classics II
491 crystallization kits (Qiagen, Hilden, Germany) and PACT kits (Molecular Dimensions,
492 Anatrace, Inc.) at 20 ℃. Crystallization drops were set up by mixing 0.5 µL of protein
493 solution with an equal volume of the reservoir solution. The initial crystals were
494 obtained under condition no. 41 (0.1 M sodium acetate pH 4.5 and 25% polyethylene
495 glycol [PEG] 3350) of Classics II, no. 13 (0.1 M MIB buffer [25 mM sodium malonate
496 dibasic monohydrate, 37.5 mM imidazole, and 37.5 mM boric acid], with pH 4.0 and

25% PEG 1500), and no. 37 (0.1 M MMT buffer [20 mM DL-malic acid, 40 mM MES monohydrate, and 40 mM Tris], with pH 4.0 and 25% PEG 1500) of PACT. After optimization by varying the buffer pH and precipitant concentration and adding co-cage1 nucleant, the optimal crystals were obtained using 0.1 M sodium acetate pH 4.5, and 20% PEG 3350 with a co-cage1 nucleant at a protein concentration of 5.4 mg/mL. Diffraction data were collected under a cold nitrogen gas stream at 100 K using Photon Factory BL-1A (Tsukuba, Japan) after cryoprotection by adding glycerol to a 20% final concentration into the reservoir solution. The optimal resolution of diffraction data was obtained by soaking a crystal with 5 mM TNA in the reservoir buffer at 37 ℃ for 4 h. All datasets were indexed, integrated, scaled, and merged using *XDS/XSCALE* program [52]. Statistical data collection and process are summarized in Table 1.

Structure determination and refinement

For EHEP structure determination, after initial phasing via the native-SAD method[21, 53], the model was obtained and refined with *auto-building* using *Phenix.autobuil* of *Phenix* software suite [54]. The obtained native-SAD structure was used as a model for rigid body refinement using *phenix.refine* [55] of *Phenix* software suite with a native data at high resolution of 1.15 Å. The structure of EHEP was automatically rebuilt using *Phenix.autobuil* of the *Phenix* software suite again [54]. Several rounds of refinement were performed using *Phenix.refine* of the *Phenix* software suite [54], alternating with manual fitting and rebuilding using *COOT* program [56]. The final refinement statistics

and geometry are shown in Table 1.

The structure of the EHEP–TNA complex was determined via the molecular replacement (MR) method using the EHEP structure as a search model with *Phaser* of *Pheni*x software suite [57]. The electron density block of TNA was clearly shown in both $2F_o–F_c$ and $F_o–F_c$ maps. Subsequently, TNA structure was manually constructed, followed by several rounds of refinement using *Phenix.refine* [54], with manual fitting and rebuilding using *COOT* [56]. We also determined the structure of phloroglucinol-soaked crystals at a resolution of 1.4 Å via the MR method using the refined EHEP structure as a search model with *Phaser*, but no electron density block of phloroglucinol was obtained. Therefore, we referred to this structure as the apo form (apo structure2). The final refinement statistics and geometry are shown in Table 1.

We determined the structure of *aku*BGL via the MR method using *Phaser* of *Phenix* software suite [57]. We used one GH domain (86–505 aa) of β-klotho (PDB entry: 5VAN) [58] as the search model. This GH domain of β-klotho shares 30% sequence identity with *aku*BGL. Four GH domains of two molecules in an asymmetric unit were found and subsequently rebuilt with *Phenix_autobuild* of *Phenix* software suite [54]. Finally, refinement of *aku*BGL structure was performed as described for EHEP.

Docking studies of *aku*BGL with phlorotanins and laminarins

We used Schrodinger Maestro program for performing docking studies [59]. First, we superimposed the structure of *Os*BGL mutant complexed with cellotetraose (PDB ID

4QLK) to that of *aku*BGL GH1D2 to define the ligand position in the ligand-binding cavity. Then, we modified the structure of the *aku*BGL GH1D2 using wizard module to remove water molecules and add hydrogen atoms for docking. The 2D structures of the inhibition ligands, including TNA, phloroglucinol, and eckol, were downloaded from Pubchem [60] and further converted to 3D structures using the LigPrep module of Schrodinger Maestro program. The structure of the substate laminaritetraose was extracted from the *Zobellia galactanivorans* β-glucanase–laminaritetraose complex structure (PDB ID: 4BOW) [61]. Then, a receptor grid was constructed in the center of the ligand-binding cavity. We performed docking using the Glide standard precision mode without any constraints. The optimal binding pose was determined using the lowest Glide score, and docked structures were analyzed using PyMol.

Data Availability

The atomic coordinates were deposited in the PDB with the accession codes as follows: EHEP with 1.15 Å resolution (8IN3), EHEP with 1.4 Å resolution (8IN4), EHEP complexed with tannic acid (8IN6), *aku*BGL(8IN1).

Acknowledgments

This work was supported in part by Grant-in-Aid for Scientific Research (B) (Grant Number 21H01754 To M. Y) and Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from Japan Agency for Medical Research and

558 Development (AMED) under Grant Number JP18am0101071 and JP19am0101083.
559 We are grateful to the Photon Factor and SPring-8 (No. 2017B2545, 2017A2551,
560 2018B2538) for beam time and the beamline staff for their assistance for data collection.
561

562 # References

563 1. Becklin, K.M. A Coevolutionary Arms Race: Understanding Plant-Herbivore
564 Interactions. *The American Biology Teacher* **70**, 288-292 (2008).
565 2. Jormalainen, V. & Honkanen, T. in Algal Chemical Ecology 57-89 (2008).
566 3. Enquist-Newman, M. et al. Efficient ethanol production from brown
567 macroalgae sugars by a synthetic yeast platform. *Nature* **505**, 239-243 (2014).
568 4. Tsuji, A., Kuwamura, S., Shirai, A. & Yuasa, K. Identification and
569 Characterization of a 25 kDa Protein That Is Indispensable for the Efficient
570 Saccharification of Eisenia bicyclis in the Digestive Fluid of Aplysia kurodai.
571 *PLoS One* **12**, e0170669 (2017).
572 5. Erb, M. & Reymond, P. Molecular Interactions Between Plants and Insect
573 Herbivores. *Annu Rev Plant Biol* **70**, 527-557 (2019).
574 6. Mendez-Liter, J.A., de Eugenio, L.I., Prieto, A. & Martinez, M.J. The beta-
575 glucosidase secreted by Talaromyces amestolkiae under carbon starvation: a
576 versatile catalyst for biofuel production from plant and algal biomass.
577 *Biotechnol Biofuels* **11**, 123 (2018).
578 7. Nakajima, M., Yamashita, T., Takahashi, M., Nakano, Y. & Takeda, T.

579 Identification, cloning, and characterization of beta-glucosidase from Ustilago
580 esculenta. *Appl Microbiol Biotechnol* **93**, 1989-1998 (2012).
581 8. Wang, Z. et al. A novel Vibrio beta-glucosidase (LamN) that hydrolyzes the
582 algal storage polysaccharide laminarin. *FEMS Microbiol Ecol* **91**, fiv087 (2015).
583 9. Kim, D.H., Kim, D.H., Lee, S.H. & Kim, K.H. A novel beta-glucosidase from
584 Saccharophagus degradans 2-40(T) for the efficient hydrolysis of laminarin
585 from brown macroalgae. *Biotechnol Biofuels* **11**, 64 (2018).
586 10. Mohsin, I., Poudel, N., Li, D.C. & Papageorgiou, A.C. Crystal Structure of a
587 GH3 beta-Glucosidase from the Thermophilic Fungus Chaetomium
588 thermophilum. *Int J Mol Sci* **20** (2019).
589 11. Hegedus, D., Erlandson, M., Gillott, C. & Toprak, U. New Insights into
590 Peritrophic Matrix Synthesis, Architecture, and Function. *Annual Review of*
591 *Entomology* **54**, 285-302 (2009).
592 12. Devenport, M. et al. Identification of the Aedes aegypti peritrophic matrix
593 protein AeIMUCI as a heme-binding protein. *Biochemistry* **45**, 9540-9549
594 (2006).
595 13. Cassani, L. et al. Seaweed-based natural ingredients: Stability of phlorotannins
596 during extraction, storage, passage through the gastrointestinal tract and
597 potential incorporation into functional foods. *Food Res Int* **137**, 109676 (2020).
598 14. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic*
599 *Acids Res* **38**, W545-549 (2010).

15. Fadel, F. et al. X-Ray Crystal Structure of the Full Length Human Chitotriosidase (CHIT1) Reveals Features of Its Chitin Binding Domain. *PLoS One* **11**, e0154190 (2016).

16. Kohler, A.C. et al. Structural Analysis of an Avr4 Effector Ortholog Offers Insight into Chitin Binding and Recognition by the Cf-4 Receptor. *Plant Cell* **28**, 1945-1965 (2016).

17. Hurlburt, N.K., Chen, L.H., Stergiopoulos, I. & Fisher, A.J. Structure of the Cladosporium fulvum Avr4 effector in complex with (GlcNAc)6 reveals the ligand-binding mechanism and uncouples its intrinsic function from recognition by the Cf-4 resistance protein. *PLoS Pathog* **14**, e1007263 (2018).

18. Mueller, G.A. et al. Serological, genomic and structural analyses of the major mite allergen Der p 23. *Clin Exp Allergy* **46**, 365-376 (2016).

19. Suetake, T. et al. Chitin-binding proteins in invertebrates and plants comprise a common chitin-binding structural motif. *J Biol Chem* **275**, 17929-17932 (2000).

20. Madland, E., Crasson, O., Vandevenne, M., Sorlie, M. & Aachmann, F.L. NMR and Fluorescence Spectroscopies Reveal the Preorganized Binding Site in Family 14 Carbohydrate-Binding Module from Human Chitotriosidase. *ACS Omega* **4**, 21975-21984 (2019).

21. Sun, X. et al. Crystallographic analysis of Eisenia hydrolysis-enhancing protein using a long wavelength for native-SAD phasing. *Acta Crystallographica Section F-Structural Biology Communications* **76**, 20-24 (2020).

621 22. Silva, R.D. & Martinho, R.G. Developmental roles of protein N-terminal
622 acetylation. *Proteomics* **15**, 2402-2409 (2015).
623 23. Hollebeke, J., Van Damme, P. & Gevaert, K. N-terminal acetylation and other
624 functions of Nalpha-acetyltransferases. *Biol Chem* **393**, 291-298 (2012).
625 24. Lange, P.F. & Overall, C.M. TopFIND, a knowledgebase linking protein termini
626 with function. *Nat Methods* **8**, 703-704 (2011).
627 25. Jie Fu, J.x., Zhang, Y. & Lu, X. A Greener Process for Gallic Acid Production
628 from Tannic Acid Hydrolysis with Hydrochloric Acid. *Asian Journal of*
629 *Chemistry* **27**, 3328-3332 (2015).
630 26. Luo, Q. et al. A novel green process for tannic acid hydrolysis using an
631 internally sulfonated hollow polystyrene sphere as catalyst. *RSC Advances* **8**,
632 17151-17158 (2018).
633 27. Amore, A. et al. Distinct roles of N- and O-glycans in cellulase activity and
634 stability. *Proc Natl Acad Sci U S A* **114**, 13667-13672 (2017).
635 28. Han, C. et al. Improvement of the catalytic activity and thermostability of a
636 hyperthermostable endoglucanase by optimizing N-glycosylation sites.
637 *Biotechnol Biofuels* **13**, 30 (2020).
638 29. Wei, W. et al. N-glycosylation affects the proper folding, enzymatic
639 characteristics and production of a fungal beta-glucosidas. *Biotechnol Bioeng*
640 **110**, 3075-3084 (2013).
641 30. Hayashi, Y. et al. Klotho-related protein is a novel cytosolic neutral beta-

642 glycosylceramidase. *J Biol Chem* **282**, 30889-30900 (2007).

643 31. Jeng, W.Y. et al. High-resolution structures of Neotermes koshunensis beta-
644 glucosidase mutants provide insights into the catalytic mechanism and the
645 synthesis of glucoconjugates. *Acta Crystallogr D Biol Crystallogr* **68**, 829-838
646 (2012).

647 32. Dong, S. et al. Structural insight into a GH1 beta-glucosidase from the
648 oleaginous microalga, Nannochloropsis oceanica. *Int J Biol Macromol* **170**,
649 196-206 (2021).

650 33. Tamaki, F.K. et al. Using the Amino Acid Network to Modulate the Hydrolytic
651 Activity of beta-Glycosidases. *PLoS One* **11**, e0167978 (2016).

652 34. Sakamoto, K., Uji, S., Kurokawa, T. & Toyohara, H. Molecular cloning of
653 endogenous beta-glucosidase from common Japanese brackish water clam
654 Corbicula japonica. *Gene* **435**, 72-79 (2009).

655 35. Sabbadin, F. et al. Uncovering the molecular mechanisms of lignocellulose
656 digestion in shipworms. *Biotechnol Biofuels* **11**, 59 (2018).

657 36. Opassiri, R. et al. Characterization of a rice β-glucosidase highly expressed in
658 flower and germinating shoot. *Plant Science* **165**, 627-638 (2003).

659 37. Hudson, K.L. et al. Carbohydrate-Aromatic Interactions in Proteins. *J Am Chem*
660 *Soc* **137**, 15152-15160 (2015).

661 38. Ni, J., Tokuda, G., Takehara, M. & Watanabe, H. Heterologous expression and
662 enzymatic characterization of .BETA.-glucosidase from the drywood-eating

termite, Neotermes koshunensis. *Applied Entomology and Zoology* **42**, 457-463 (2007).

39. Tsuji, A., Tominaga, K., Nishiyama, N. & Yuasa, K. Comprehensive enzymatic analysis of the cellulolytic system in digestive fluid of the Sea Hare Aplysia kurodai. Efficient glucose release from sea lettuce by synergistic action of 45 kDa endoglucanase and 210 kDa ss-glucosidase. *PLoS One* **8**, e65418 (2013).

40. Jung, H.A., Oh, S.H. & Choi, J.S. Molecular docking studies of phlorotannins from Eisenia bicyclis with BACE1 inhibitory activity. *Bioorg Med Chem Lett* **20**, 3211-3215 (2010).

41. Amsler, C.D. & Fairhead, V.A. in Advances in Botanical Research, Vol 43: Incorporating Advances in Plant Pathology, Vol. 43. (ed. J.A. Callow) 1-91 (2006).

42. Shimada, T. Salivary proteins as a defense against dietary tannins. *J Chem Ecol* **32**, 1149-1163 (2006).

43. War, A.R. et al. in Co-Evolution of Secondary Metabolites 795-822 (2020).

44. Barbehenn, R.V. & Peter Constabel, C. Tannins in plant-herbivore interactions. *Phytochemistry* **72**, 1551-1565 (2011).

45. Lin, D. et al. The effect of ionic strength and pH on the stability of tannic acid-facilitated carbon nanotube suspensions. *Carbon* **47**, 2875-2882 (2009).

46. Ge, D., Yuan, H., Shen, Y., Zhang, W. & Zhu, N. Improved sludge dewaterability by tannic acid conditioning: Temperature, thermodynamics and

684 mechanism studies. *Chemosphere* **230**, 14-23 (2019).

685 47. Yi, Z. et al. Green, effective chemical route for the synthesis of silver nanoplates
686 in tannic acid aqueous solution. *Colloids and Surfaces A: Physicochemical and*
687 *Engineering Aspects* **392**, 131-136 (2011).

688 48. Dultz, S., Mikutta, R., Kara, S.N.M., Woche, S.K. & Guggenberger, G. Effects
689 of solution chemistry on conformation of self-aggregated tannic acid revealed
690 by laser light scattering. *Sci Total Environ* **754**, 142119 (2021).

691 49. Han, Y. et al. Polyphenol-Mediated Assembly of Proteins for Engineering
692 Functional Materials. *Angew Chem Int Ed Engl* **59**, 15618-15625 (2020).

693 50. Lemke, T., Stingl, U., Egert, M., Friedrich, M.W. & Brune, A. Physicochemical
694 conditions and microbial activities in the highly alkaline gut of the humus-
695 feeding larva of Pachnoda ephippiata (Coleoptera: Scarabaeidae). *Appl Environ*
696 *Microbiol* **69**, 6650-6658 (2003).

697 51. Yao, M. & Li, L. (NATIONAL UNIVERSITY CORPORATION
698 HOKKAIDO UNIVERSITY 2022).

699 52. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132 (2010).

700 53. Yu, J., Shinoda, A., Kato, K., Tanaka, I. & Yao, M. A solution-free crystal-
701 mounting platform for native SAD. *Acta Crystallogr D Struct Biol* **76**, 938-945
702 (2020).

703 54. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for
704 macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**,

213-221 (2010).

55. Afonine, P.V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-367 (2012).

56. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).

57. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).

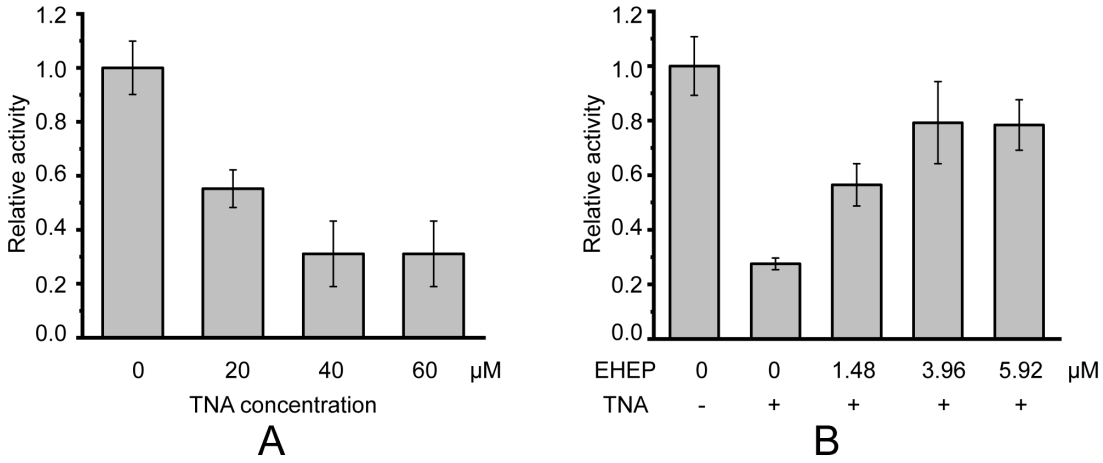
58. Lee, S. et al. Structures of beta-klotho reveal a 'zip code'-like mechanism for endocrine FGF signalling. *Nature* **553**, 501-505 (2018).

59. Sastry, G.M., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design* **27**, 221-234 (2013).

60. Wang, Y.L. et al. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Research* **37**, W623-W633 (2009).

61. Labourel, A. et al. The beta-glucanase ZgLamA from Zobellia galactanivorans evolved a bent active site adapted for efficient degradation of algal laminarin. *J Biol Chem* **289**, 2027-2042 (2014).

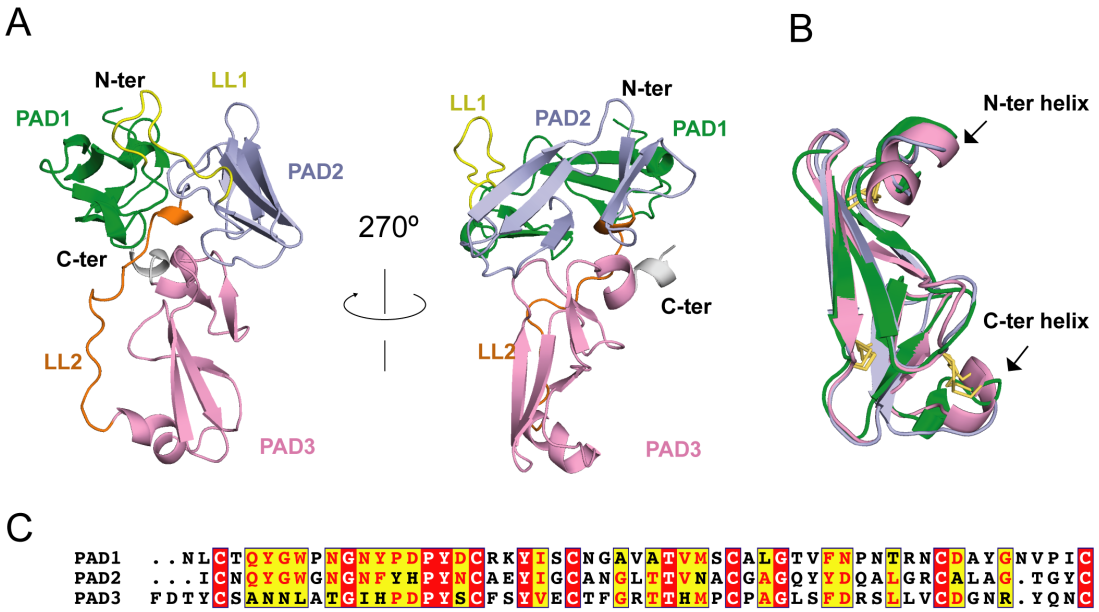
725 Fig 1



726

727 **Figure 1.** Galactoside hydrolytic activity of *aku*BGL toward ortho-Nitrophenyl-β-
728 galactoside. (A) The hydrolytic activity of akuBGL with TNA at different
729 concentrations. (B) The hydrolytic activity of akuBGL (0.049 μM) with 40 μM TNA
730 and EHEP at different concentrations. The average and standard deviation of the
731 relative activity were estimated from three independent replicates (N = 3).

732 Fig 2



733

734 **Figure 2.** EHEP structure. (A) Cartoon representation of EHEP. The three PAD

735 domains are colored green, light blue, and pink, respectively. Linker long loop1 (LL1)

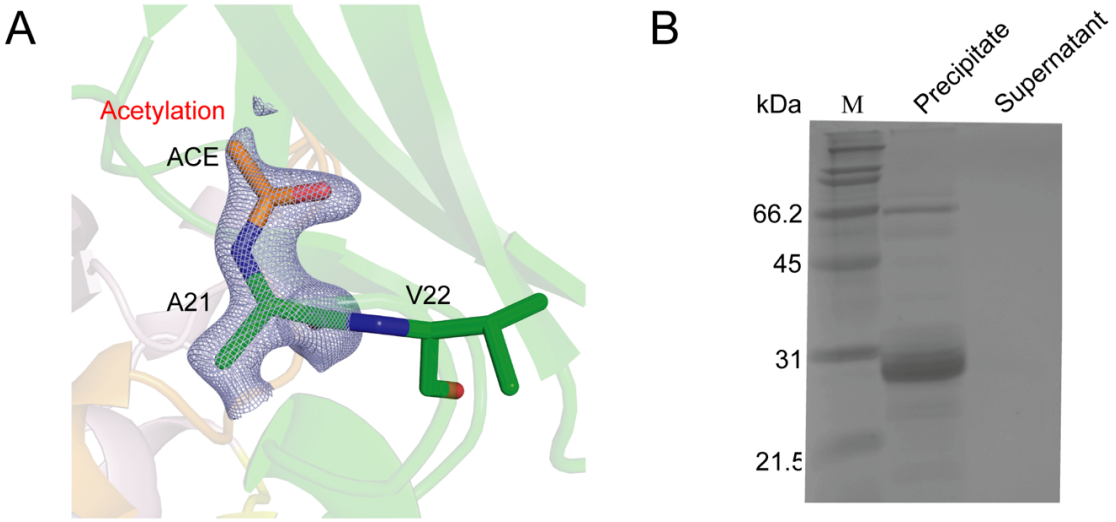
736 and loop2 (LL2) are colored yellow and blue. (B) Structural superposition of the three

737 PAD domains of EHEP. The three domains are colored as in (A). The disulfide bonds

738 are shown as yellow sticks. (C) Sequence alignment of three PAD domains. Alignment

739 was performed by CLUSTALW and displayed with ESPript3.

740 Fig 3



741

742 **Figure. 3**. (A) Acetylation modification on the N-terminal residue A21. The structure

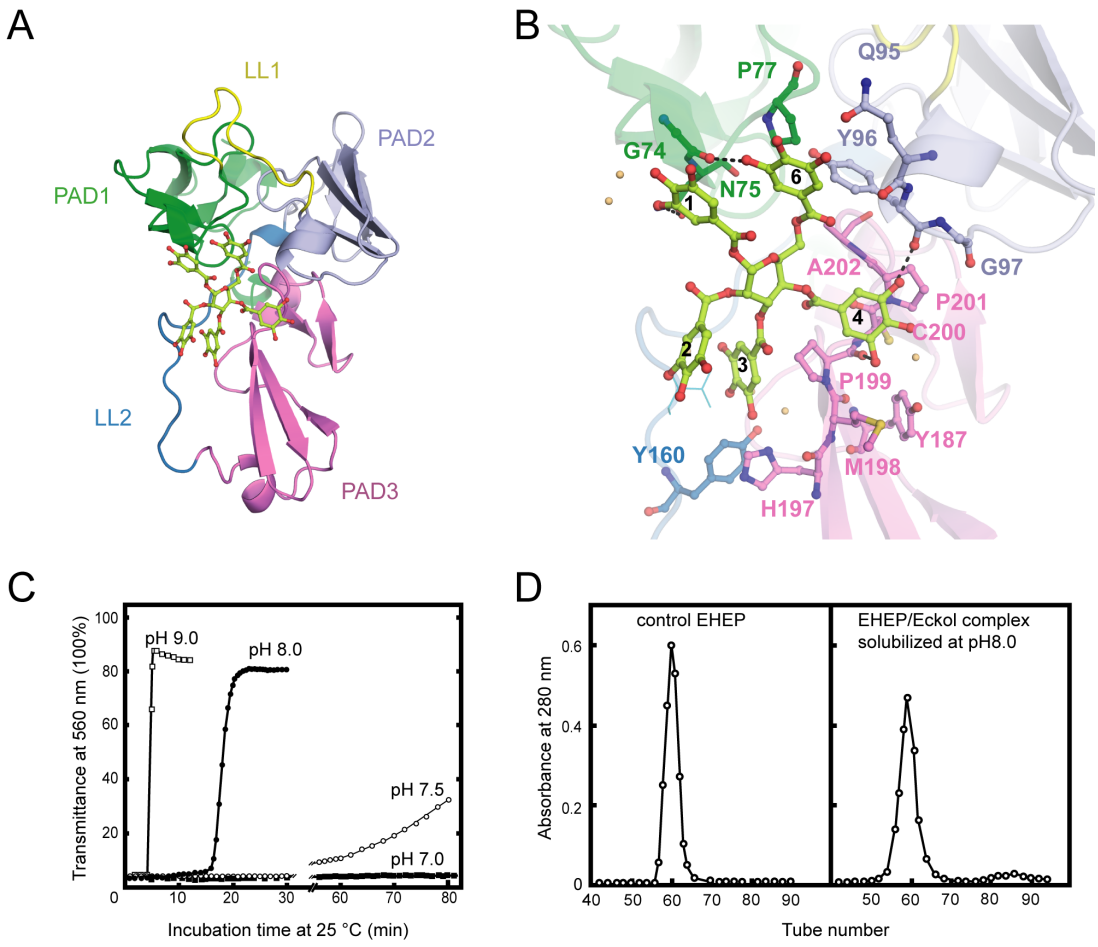
743 is shown as sticks with an omitted map of A21-ace at a 3.3 σ level (blue-white). (B)

744 TNA binding activity of recombEHEP. SDS–PAGE was run using a mixture of

745 recombEHEP and TNA.

746

747 Fig 4



748

749 **Figure 4.** Structure of EHEP-TNA. (a) The overall structure of EHEP-TNA. (A) the

750 overall structure of EHEP-TNA. EHEP and TNA are shown by the cartoon and stick

751 model, respectively. EHEP is colored as in Fig. 1. The C and O atoms of TNA are

752 colored lemon and red, respectively. (B) Interaction of TNA (ball-stick in same color

753 as (A)) with EHEP (cartoon in same color as (A)) in EHEP-TNA structure. The residues

754 of EHEP in contact are labeled and shown by a ball-stick with N, O, and S atoms in

755 blue, red, and brown, respectively. The C and O atoms of TNA are colored the same as

756 (A), lemon and red, respectively. Dashed lines show hydrogen bonds. The water

757 molecules stabilizing TNA was shown as light orange spheres. (C) Effect of pH on

758 resolubilization of an EHEP–eckol precipitate. Buffers with pH 9.0, 8.0, 7.5, and 7.0

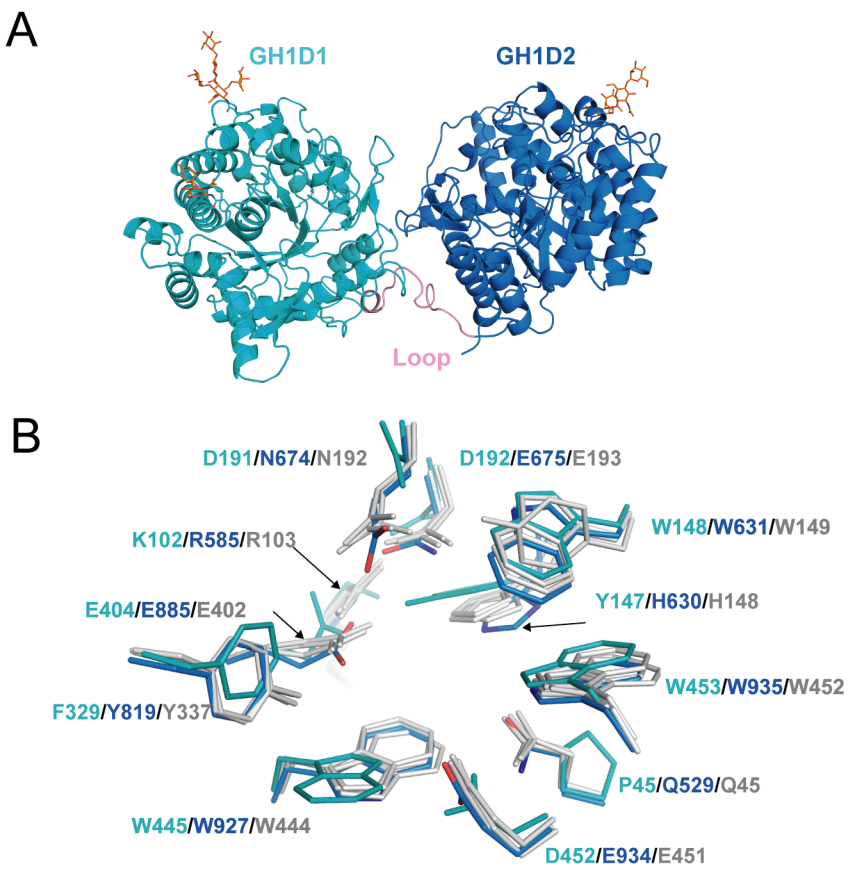
759 are presented as hollow square, solid circle, hollow circle, and solid square, respectively.

760 (D) The EHEP–eckol precipitate was dissoloved in 50 mM Tris–HCl (pH 8.0) and

761 analyzed by Sephacryl S-100.

762

763 Fig 5

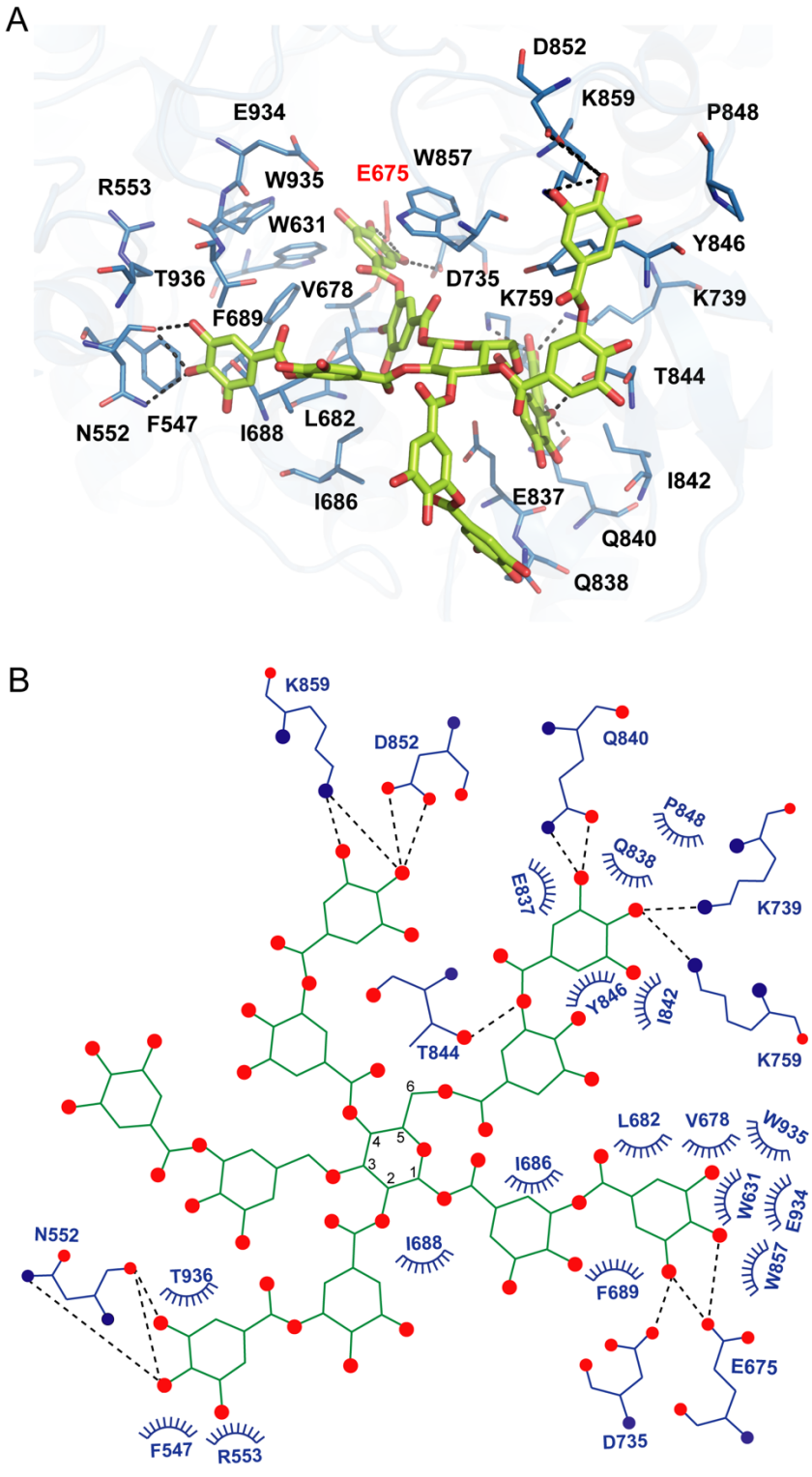


764

765 **Figure 5.** Structure of *aku*BGL. (A) Overall structure. The GH1D1 (light blue) and
766 GH1D2 (cyan) domains are linked by a long loop (linker-loop) colored in pink. The N-
767 linked glycans were shown in the orange stick. (B) Residues superposition of the GBS
768 and CR sites of the domains GH1D1 (cyan), GH1D2 (light blue) with β-glucosidase
769 structures from termite *Neotermes koshunensis* (*Nk*BGL, grey), β-glucosidase from rice
770 (*Os*BGL, gray), β-glucosidase from *Bacillus circulans* sp. *Alkalophilus* (grey). Only the
771 residues numbers of GH1D1 (cyan), GH1D2 (light blue), and *Nk*BGL (grey) are shown
772 for clarity.

773

774 Fig 6



775

776 **Figure 6.** Docking model of *aku*BGL with TNA. (A) Detailed interaction between

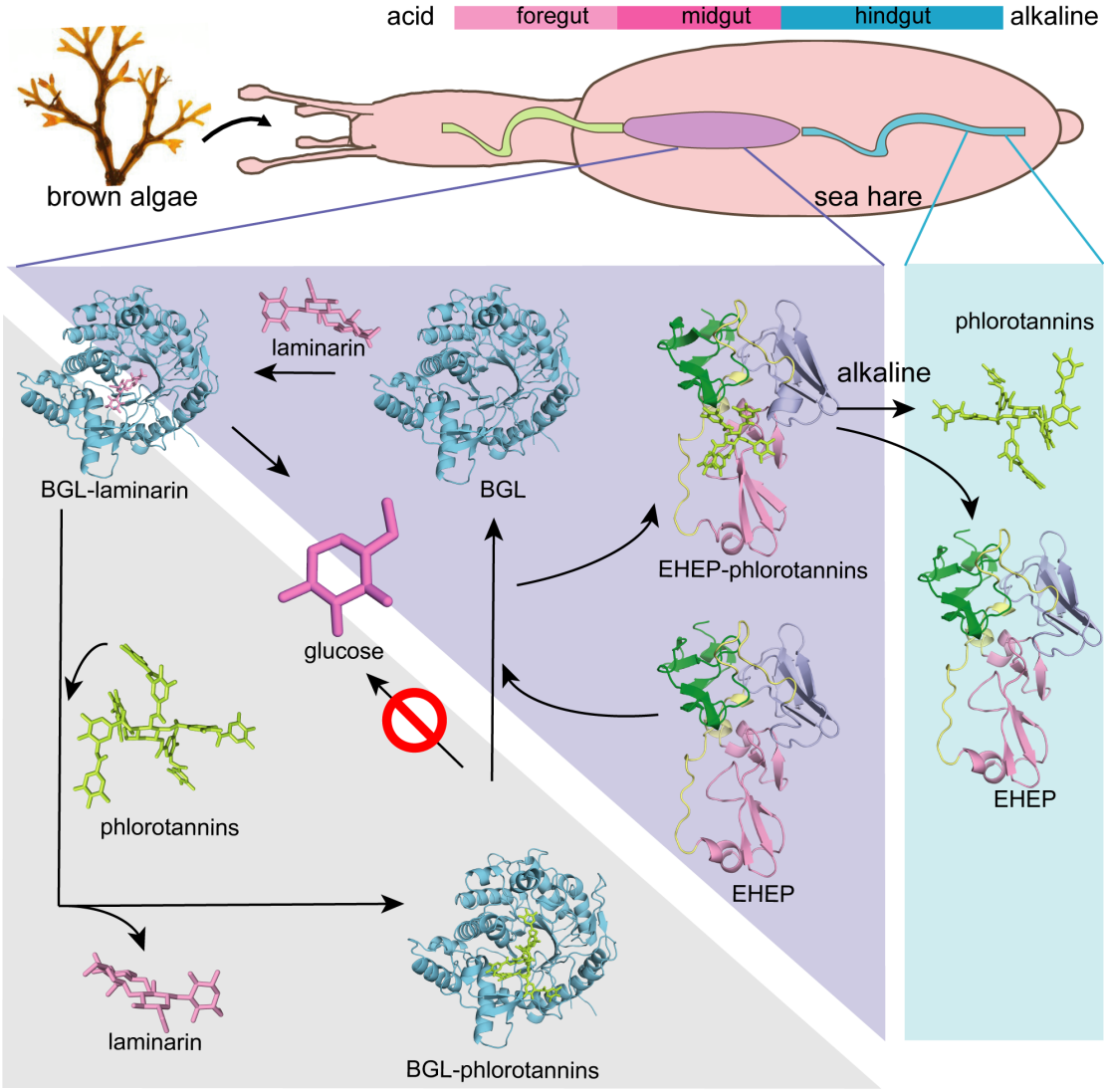
777 *aku*BGL and TNA in the docking model. TNA is shown in the green stick model. The

778 hydrogen bonds are shown as dashed lines. (B) A 2D diagram of the interaction between

779 *aku*BGL and TNA shown in (A). The hydrogen bonds are shown as dash lines, and the

780 hydrophobic contacts as circular arcs.

781 Fig 7



782

783 **Figure 7.** Proposed molecular mechanisms. Proposed molecular mechanisms of TNA

784 inhibition of *aku*BGL activity and EHEP's protective effects of *aku*BGL in *aku*BGL–

785 phlorotannin/laminarin–EHEP system (light purple triangle). The digestive tract of

786 *A.kurodai* consists of foregut (blue), midgut (purple), and hindgut (blue) (top). The bar

787 chart above depicts the pH of the digestive tract, with pink denoting acid and blue

788 denoting alkalinity.

789

790 **Table 1** X-ray data collection and structure-refinement statistics.

	Native-SAD of EHEP	EHEP1	EHEP2	EHEP-TNA	*aku*BGL
Data Collection					
Beamline	PF BL17A	PF BL17A	Spring 8 BL-41XU	PF BL17A	PF BL1A
Wavelength (Å)	2.1000	0.9800	1.0000	0.9800	1.0000
Resolution range (Å)	46.61-2.48 (2.54-2.48)	46.56–1.15 (1.20–1.15)	47.02- 1.4 (1.45-1.4)	46.84–1.9 (1.97–1.9)	49.65-2.7 (2.80-2.7)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P6_2$
Unit-cell parameters *a, b, c* (Å)	42.2, 65.3, 66.5	42.2, 65.3, 66.5	40.6, 65.6, 67.5	42.5, 65.4, 67.2	191.7, 191.7, 112.6
Completeness (%)	97.7 (87.5)	93.4 (75.8)	99.30 (97.98)	99.9 (99.7)	99.9 (99.1)
Redundancy	99.7 (38.9)	6.6 (6.1)	6.4 (5.8)	6.4 (6.4)	10.7 (10.9)
Average I/σ(I)	105.1 (74.4)	19.28 (1.97)	14.34 (3.41)	15.39 (1.83)	10.69 (2.57)
R_{meas}(%)[a]	6.9 (14.2)	7.3 (90.5)	8.6 73.7(55.3)	8.9 (89.4)	19.4 (83.0)
$CC^{1/2}$ (%)	100 (99.8)	99.9（70.4）	99.8 (86.2)	99.9（73.7）	99.5 (84.2)
Ano/Sig	3.1 (1.7)				
Molecules/ asymmetric unit	1	1	1	1	2
Refinement					
R_{work}[b]/R_{free}[c] (%)		18.19/18.91	16.57/18.39	19.87/23.54	18.39/21.98
No. of atoms		1955	1861	1732	15595
No. of residues		1600	1573	1580	15256
No. of water molecules		343	273	87	96
No. of Ligands		12	15	67	243
RMSD from ideality					
bond length (Å)		0.005	0.006	0.008	0.004
bond angle (°)		0.84	0.84	0.91	0.65
Ramachandran plot (%)					

Favoured	99.02	98.54	98.52	96.16
Allowed	0.98	1.46	1.48	3.74
Outliers	0.00	0.00	0.00	0.11
PDB accession code	8IN3	8IN4	8IN6	8IN1

791 The highest resolution shell is shown in parentheses.

792 [a]$R_{meas} = \Sigma_{hkl}\{N(hkl) / [N(hkl) -1] \}^{1/2} \Sigma_i|I_i(hkl)-< I(hkl)> | /\Sigma_{hkl} \Sigma_i|I_i(hkl)$, where where

793 $I_i(hkl)$ is the ith observation of the intensity of reflection *hkl* and $\langle I(hkl)\rangle$ is the mean

794 over n observations.

795 [b]$R_{work} = \Sigma_{hkl}||F_{obs}(hkl)| - |F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$.

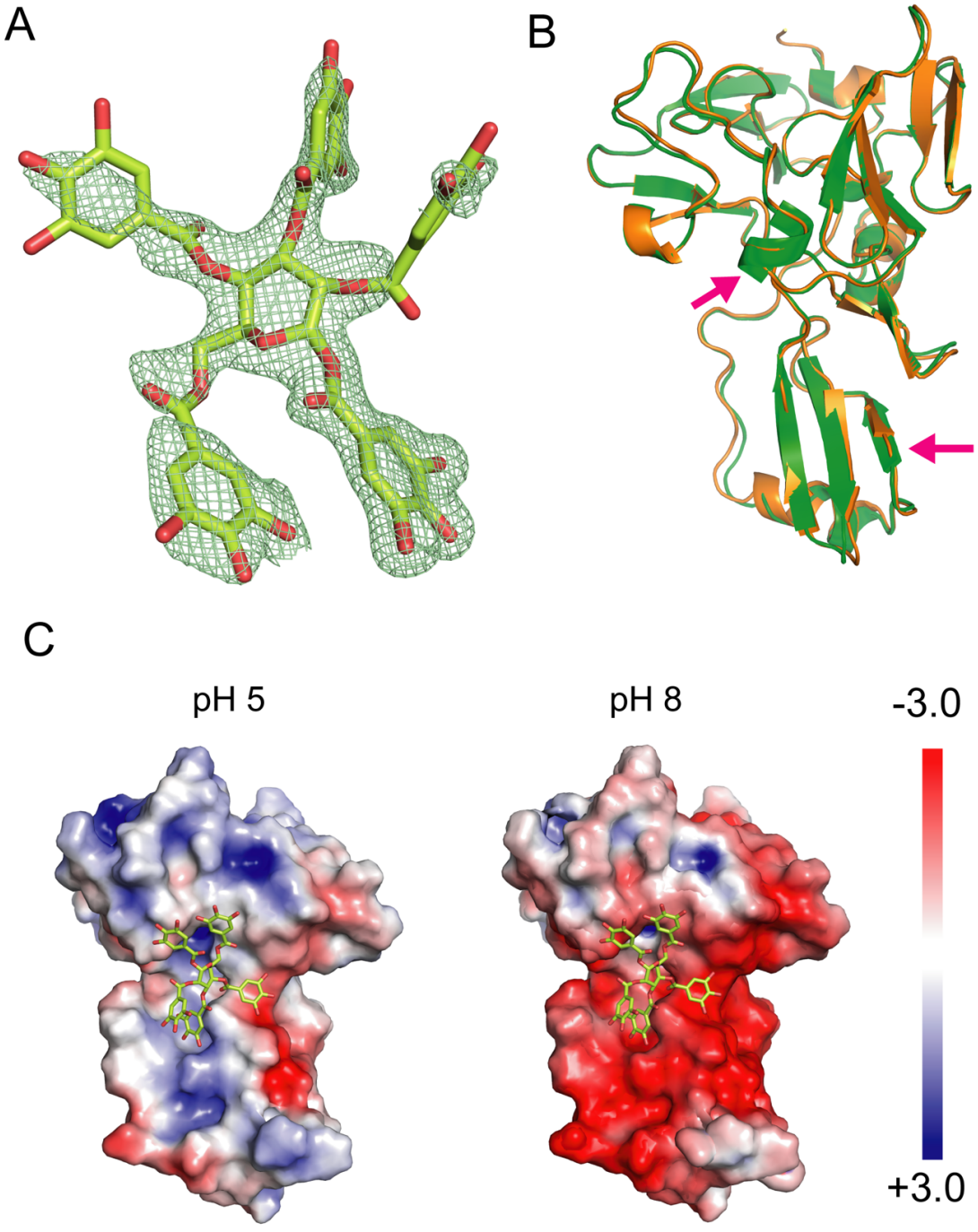
796 [c]R_{free} was calculated with an approximate 5% fraction of randomly selected reflections

797 evaluated from refinement.

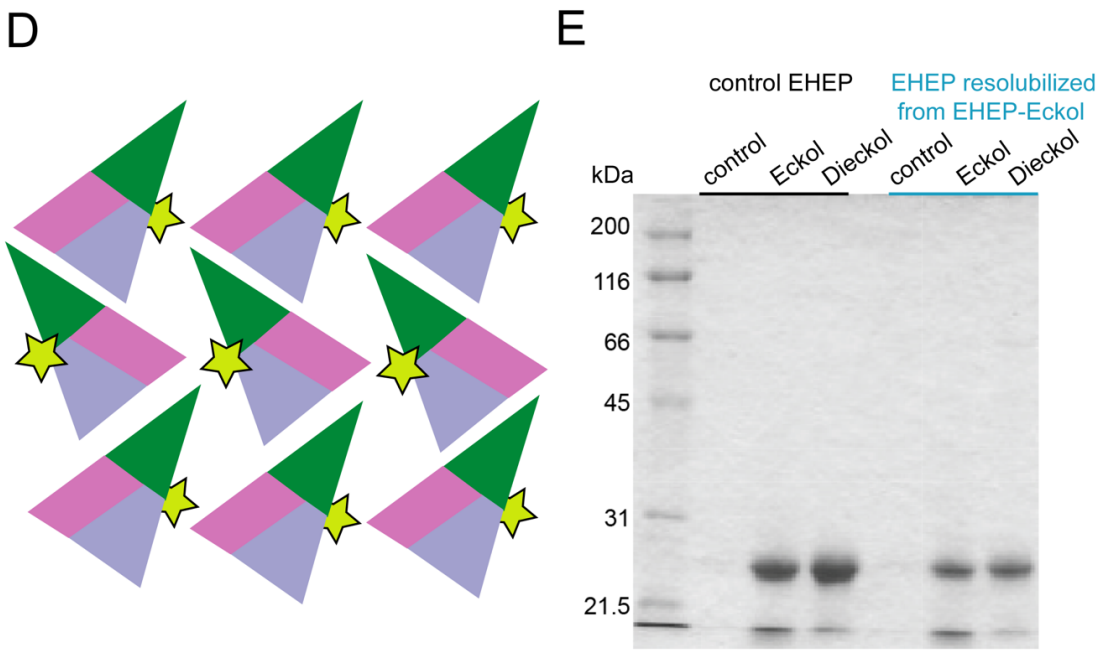
798

799 **Supporting information**

800 Fig.S1



801



802

803 **Figure S1.** EHEP–TNA structure and structural comparisons. (A) TNA structure (stick

804 model) in EHEP–TNA with an omitted map countered at the 2.0 σ level. The O and C

805 atoms are colored red and lemon, respectively. (B) Structure superposition of apo EHEP

806 (orange) with EHEP–TNA (green). Two arrows mark the conformational changes in a

807 β-sheet of PAD1 and an α-helix of PAD2, respectively. (C) Electrostatic potential of the

808 EHEP surface in the EHEP–TNA complex at different pH values. (D) Diagram of

809 EHEP–TNA packing in the crystal structure. The triangle represents the EHEP, with the

810 three colors denoting the three PAD domains, as in Fig. 3a. TNA is shown in a lemon

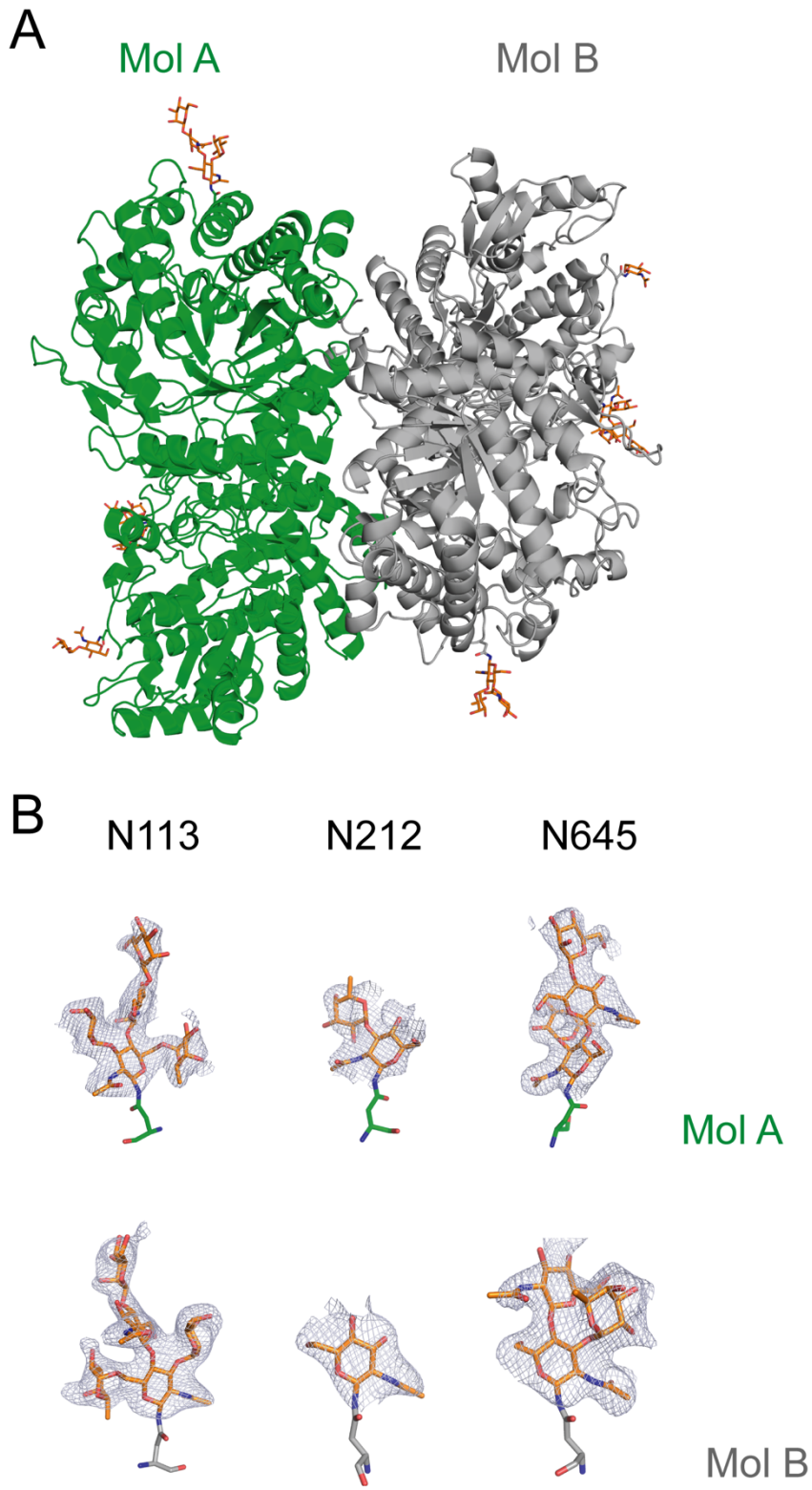
811 pentagram. (E) Activity of the resolubilized EHEP. The EHEP-eckol precipitate was

812 resolubilized in Tris–HCl (pH 8.0) and then the supernatant was purified by Sephacrys

813 S-100 column to measure eckol-binding activity.

814

815 Fig.S2



816

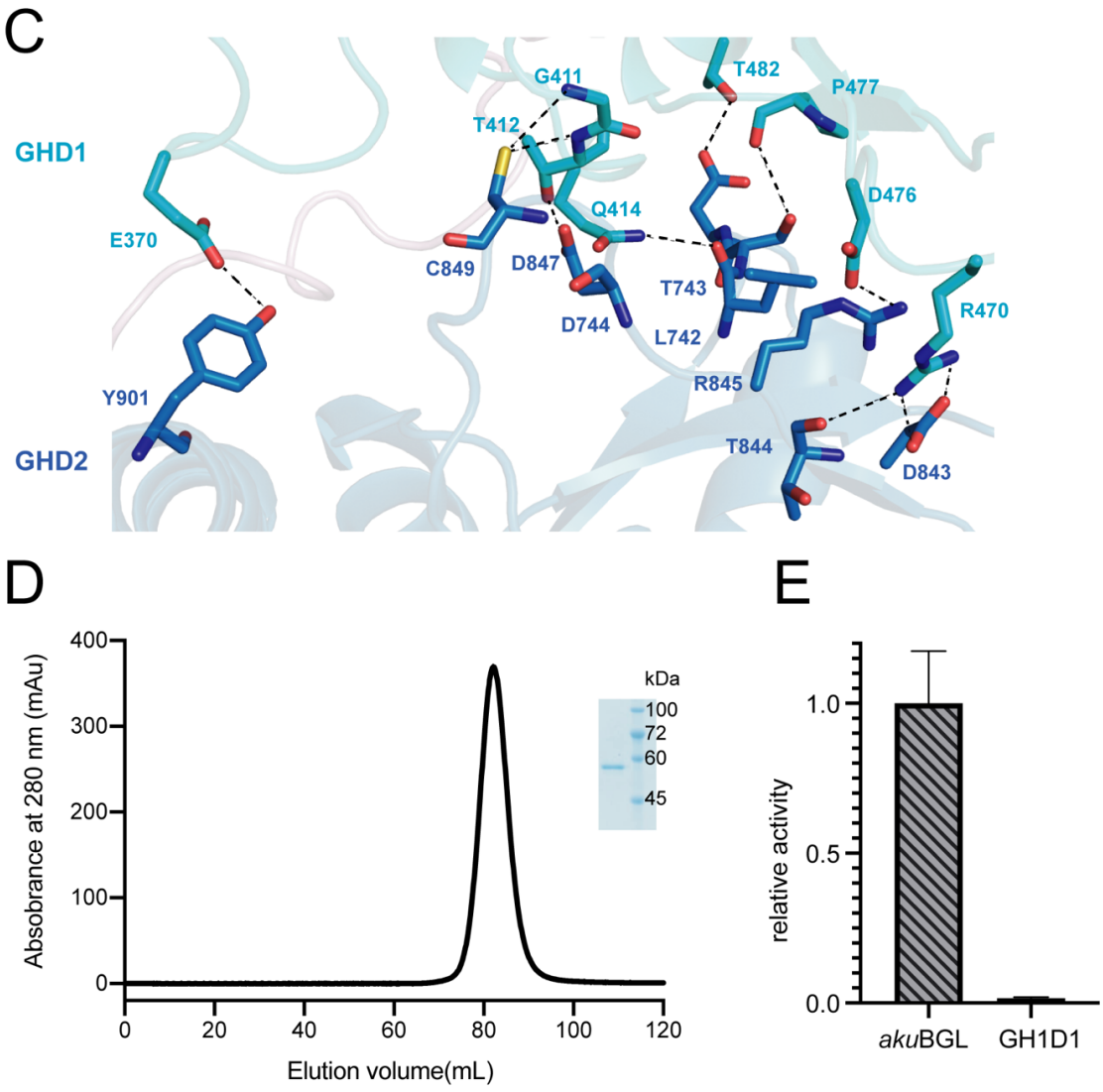
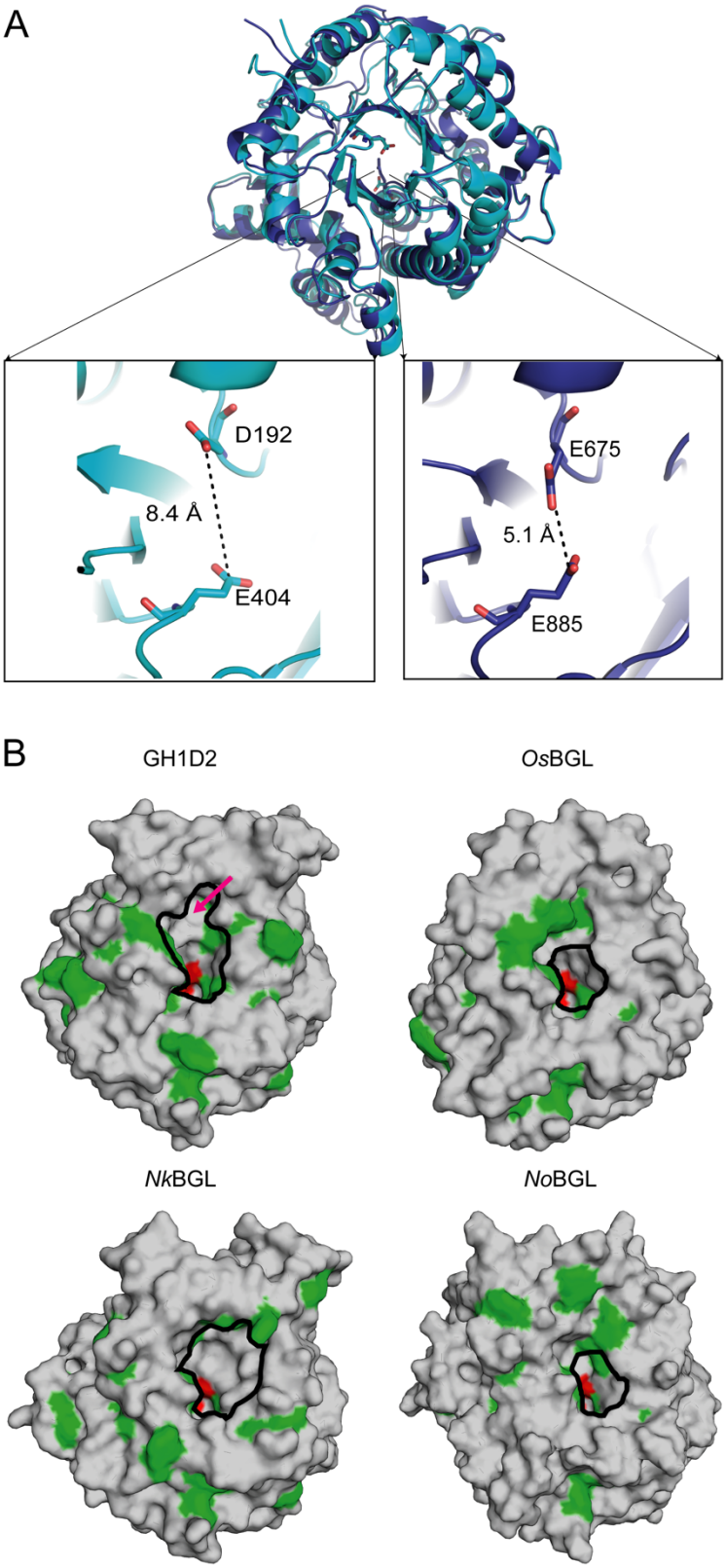


Figure S2. *aku*BGL structure. (A) Two *aku*BGL molecules in the asymmetry unit, colored green and grey, respectively. The glycosylation sites are shown as orange sticks, with O and N atoms in red and blue, respectively. (B) The glycosylation chains with omitted density maps are countered at the 2.0 σ level. (C) The interface between the GH1D1 (cyan) and GH1D2 (blue). The key interactions between the two domains are shown as black dashed lines. (D) Size exclusion chromatogram of the purified GH1D1 domain. The inset shows the SDS–PAGE analysis of the GH1D1 domain. (E) Galactoside hydrolytic activity of the GH1D1 toward ortho-Nitrophenyl-β-galactoside.

826 The average and standard deviation of the relative activity were estimated from three

827 independent replicates (N = 3).

828 Fig. S3

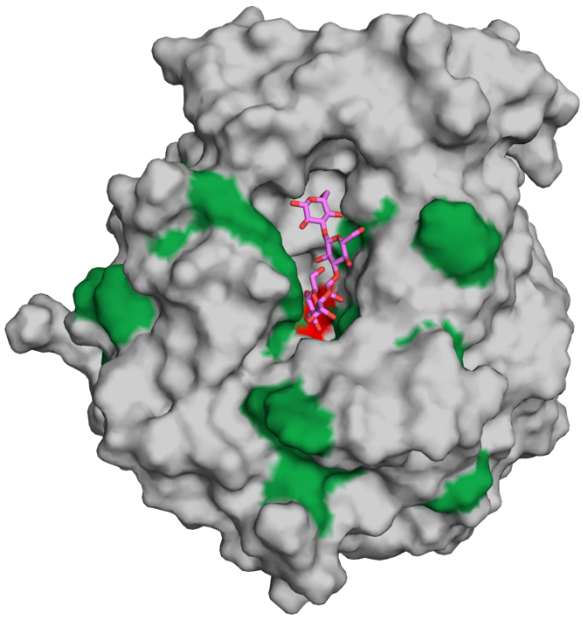


829

830 **Figure S3.** Structural comparison of BGLs. (A) Structure superposition of

831 GH1D1(cyan) and GH1D2 (blue). The enlarged picture shows the distance of
832 conceivable catalytic residues in GH1D1 and GH1D2. (B) Surface representations of
833 GH1D2, *Nk*BGL, *Os*BGL, and *No*BGL with the aromatic and catalytic residues colored
834 green and red, respectively. The red arrow indicates the location of the auxiliary site of
835 GH1D2. The active pockets are highlighted by a black circle on each surface.

836 **Fig. S4**

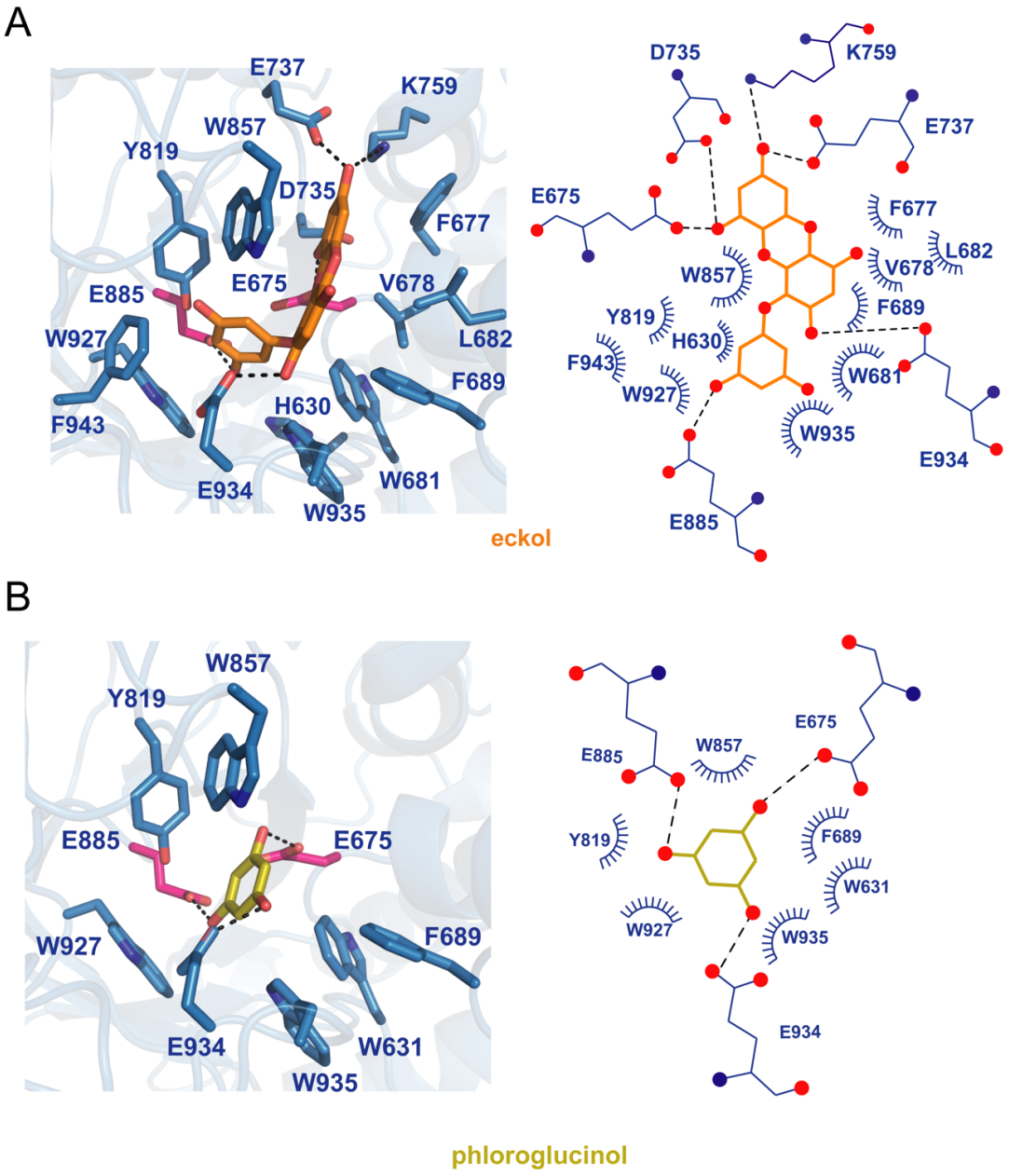


837

838 **Figure S4.** The model of GH1D2 docking with the substrate laminaritetraose. GH1D2
839 is shown as a grey surface representation and laminaritetraose as marine sticks. The
840 aromatic and catalytic residues of GH1D2 are colored green and red, respectively.

841

842 Fig. S5



843

844 **Figure S5.** The models of GH1D2 docking (A) With eckol; (B) With phloroglucinol.

845 The left panel shows the 3D structures, and the right panel shows the 2D diagrams. The

846 C, N, and O atoms of residues are colored light blue, dark blue, and red, respectively.

847 The C atoms of eckol and phloroglucinol are shown in orange and yellow, respectively.

848