HOKKAIDO UNIVERSITY

| | |
|---|---|
| Title | Characterization and ligand-binding manner of EHEP and BGL for producing biofuel from brown algae |
| Author(s) | 孫, 暁梅 |
| Degree Grantor | 北海道大学 |
| Degree Name | 博士(生命科学) |
| Dissertation Number | 甲第13944号 |
| Issue Date | 2020-03-25 |
| DOI | https://doi.org/10.14943/doctoral.k13944 |
| Doc URL | https://hdl.handle.net/2115/78170 |
| Type | doctoral thesis |
| File Information | SUN_XIAOMEI.pdf |

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

**Doctoral Dissertation**

# Characterization and ligand-binding manner of EHEP and BGL for producing biofuel from brown algae

褐藻からバイオ燃料を生産するタンパク質 EHEP と BGL の特性およびリガンド結合様式の研究

*By*

Xiaomei Sun

Laboratory of X-ray Structural Biology

Graduate School of Life Science, Hokkaido University

Sapporo, Japan

March, 2020

# Content

**Abstract** ..... **V**

**Chapter 1. Introduction** ..... **1**

1.1 Brown algae as promising feedstocks for producing biofuel ..... 1

1.2 β-glucosidase is responsible for the glucose production from brown algae ..... 3

1.3 EHEP protects BGL from phlorotannin inhibition ..... 7

1.4 Phlorotannin and its analogs ..... 9

1.5 Purpose ..... 11

**Chapter 2 The inhibition and protective activity of TNA and EHEP with akuBGL** ..... **13**

2.1 Material and methods ..... 13

2.1.1 Preparation of ONPG, ONG, tannic acid, EHEP and BGL ..... 13

2.1.2 The reaction system of TNA inhibits *aku*BGL activity ..... 14

2.1.3 Reaction system of EHEP protect BGL from inhibition ..... 15

2.1.4 HPLC analysis ..... 15

2.2 Results & discussion.....16

2.2.1 TNA inhibits the activity of *aku*BGL.....16

2.2.2 EHEP protects BGL from TNA inhibition.....19

**Chapter 3 Binding mechanism of EHEP with TNA.....21**

3.1 Materials and methods.....21

3.1.1 Purification of EHEP.....21

3.1.2 Crystallization and data collection.....21

3.1.3 Structure solution and refinement.....23

3.1.4 MALDI-TOF MASS spectrum analysis.....24

3.1.5 Expression and purification of recombinant EHEP.....25

3.1.6 Activity of recombinant EHEP.....26

3.1.7 Isothermal titration calorimetry (ITC).....27

3.1.8 Molecular docking.....27

3.1.9 Dissolve of precipitate of EHEP-TNA.....28

3.2 Results & discussion.....30

3.2.1 Overall structure of EHEP.....30

3.2.2 Characterization of EHEP.....36

3.2.3 The effect of acetylation modification on the activity of EHEP.....39

3.2.4 Complex structure of EHEP with TNA.....43

3.2.5 ITC result.....53

3.2.6 Molecular docking result of EHEP with phlorotannins.....55

3.2.7 Dissolution of the precipitate of EHEP-TNA.....57

**Chapter 4 Structure of *aku*BGL.....59**

4.1 Materials and methods.....59

4.1.1 Sample preparation.....59

4.1.2 Crystallization and data collection.....59

4.1.3 Structure solution and refinement.....60

4.1.4 Molecular docking study.....61

4.1 Results & discussion.....63

4.2.1 Crystallization of *aku*BGL.....63

4.2.2 Overall structure of *aku*BGL.....66

4.2.3 Molecular docking.....74

**Reference .....81**

**Acknowledgements.....87**

## Abstract

Dramatic increases in fuel demands globally have prompted a search for renewable energy sources. Biofuel is a promising alternative to fossil fuels because of its lower cost, renewable supply, and reduced greenhouse gas emissions. Brown algae are considered ideal feedstocks for producing biofuel because their many advantages. The cultivation does not need arable land, only use fresh water. The brown algae do not contain lignin, making the enzymatic hydrolysis of brown algae relatively easy.

The sea hare *Aplysia kurodai* consumes brown algae as a staple food, using *β*-glucosidases (*aku*BGLs) as catalysts for hydrolyzing the glycosidic bonds of laminarin abundant in brown algae to produce glucose, making it an excellent model for investigating the biofuel production process. However, phlorotannins which are also abundant in brown algae, inhibit the hydrolysis reaction of *aku*BGLs, bringing the serious problem in production of biofuel source glucose from brown algae. Interestingly, Eisenia hydrolysis enhancing protein (EHEP) identified in the digestive fluid of *A. kurodai*, was found recently to protect *aku*BGL from this inhibition by binding and then precipitating with phlorotannins, despite not possessing catalytic function(Tsuji *et al.*, 2017)(Tsuji *et al.*, 2017)(Tsuji *et al.*, 2017). Hence, EHEP and *aku*BGL have high potential for the application in the biofuel industry. An understanding of phlorotannin-binding manner with EHEP and *aku*BGL will lead us to elucidate protective mechanisms of EHEP and the phlorotannin-inhibitory mechanism of *aku*BGL.

Consequently, these knowledge could provide information on how increase the activity of *aku*BGL and how EHEP can be used recycled, and knowledge of the three-dimensional structures of *aku*BGL and EHEP is indispensable for this.

Because phlorotannin is very difficult to isolate from brown algae and it has been reported that tannic acid (TNA) is an analog of phlorotannins, we explore firstly the inhibitory activity of TNA for *aku*BGL and whether EHEP binds TNA to protect *aku*BGL from TNA-inhibition by analyzing the *aku*BGL activity using ortho-Nitrophenyl-β-galactoside (ONPG) as the substrate. The results showed that the activity of *aku*BGL decreased with increasing the concentration of tannic acid. About 70% activity of *aku*BGL was inhibited with a concentration of 40 μM TNA. When adding EHEP, the more activity of *aku*BGL was recovered with increasing the concentration of EHEP. About 80% activity of *aku*BGL was recovered with a concentration of 3.96 μM EHEP. Therefore, instead of phlorotannins, we used TNA to investigate the protective mechanism of EHEP and the phlorotannin-inhibitory mechanism.

To understand the TNA-binding manner of EHEP, we tried to analyze structures of EHEP in apo and TNA-binding form. Because EHEP is a novel and unique protein, with no structures of homologous proteins were reported, we determined the structure of EHEP by single-wavelength anomalous dispersion using sulfur atoms in the native protein as scattering factors (native-SAD) using a long wavelength. To obtain the structure of EHEP complexed with TNA (EHEP-TNA), we attempted to soak apo crystal into TNA solution for one month. The structure of EHEP consists of three chitin-

binding domains (ChBD1, ChBD2, and ChBD3) linked by two long loops. In the EHEP-TNA structure, TNA is located on the surface of EHEP, and outmost five gallic acids of TNA cannot be visualized due to structural disorder. The TNA-binding pocket was formed by C-terminal α-helix of ChBD1, N-terminal α-helix of ChBD2, and middle part (loop) of ChBD3 with the help of two linker loops. TNA bound to EHEP by hydrogen bond and hydrophobic interactions, and the binding did not induce large conformation changes. Based on structural information, we consider that precipitation of EHEP with TNA should be recovered and successfully dissolve the precipitate of EHEP-TNA using buffer with alkaline pH.

Furthermore, we determined the structure of *aku*BGL. The structure consists of two GH1 domains (GH1 domain1, 29-492 GH1-DI and GH1 domain2, 513-974: GH1-DII) which adopt almost the same structure with a rmsd of 0.59 Å for 371 Cα atoms (40.47 % sequence identity). This is the first time to visualize the structure that contains two GH1 domains of BGLs. As other structures of GH1 family enzymes, each GH1 domain exhibits a classical $(\beta/\alpha)_8$ -barrel fold. D2 domain processes an active site containing two conserved carboxylic acid residues (E675, E885), however, the first catalytic residue in D1 domain was mutated to D192. The two-domain architecture might represent a specific adaption towards brown algae. Moreover, docking was performed to analysis the inhibition mechanism of tannic acid with *aku*BGL.

# Chapter 1. Introduction

## 1.1 Brown algae as promising feedstocks for producing biofuel

With the increasing demand for the fuels due to quickly growth in size of the population and rapidly industrialization development, the world is faced with the decreasing fuel reserves problem. Furthermore, because of high cost and limited sources of fossil fuels, additionally, the greenhouse gasses emission, the focus has turned to develop the unitarization of renewable energy resources ( Fig 1) (Ellabban *et al.*, 2014). Biofuels is considered as the important alternative choice, producing about 10% of world's energy recently (Kopetz, 2013). The first-generation feedstocks for producing biofuel is corn and other oil crops. However, biofuel production from terrestrial biomass meets with many challenges, such as food sacristy and pollution of agricultural land. The second-generation feedstocks to produce biofuels is

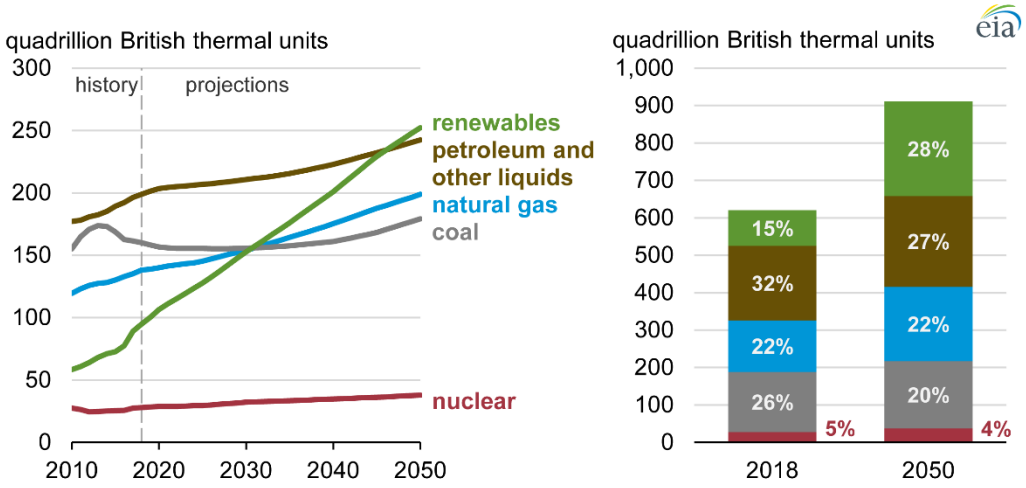



Fig 1. Global primary energy consumption by energy source (Source: U.S. Energy Information Administration, International Energy Outlook 2019)

lignocellulostic feedstock. The biofuel production process mainly consists of the following steps: delignification to liberate cellulose and hemicellulose form their complex with lignin, depolymerization of the carbohydrate to free sugars, and fermentation of sugars to ethanol. Nowadays, the obstacles have not overcome to releasing sugars from recalcitrant lignocellulose (Lin & Tanaka, 2006). Lignin has to be removed before the next step, making this biofuel high cost.

Brown algae is considered as the ideal feedstock for producing biofuel (Enquist-Newman *et al.*, 2014) (Fig 2)because their many advantages. The cultivation does not need arable land, only use fresh water. The area of sea covers about 70% of the Earth. Brown algae utilizes the energy of sunlight and greenhouse gas $CO_2$ and has a rapid growth rate. The algae do not contain lignin, which are essential for structure support and thus can be difficult to be depolymerized (Enquist-Newman *et al.*, 2014) (Wei *et al.*, 2013). Besides, other valuable materials can be separated to support sustainable food production, such as protein and potash fertilizer. Brown algae get their characteristic color from the pigment fucoxanthin and are abundant in nearshore coastal waters with suitable grown in low temperature between 6 and 14 ℃. Brown algae are framed in East-North Asia for commercial use.




Fig 2 The process of microalgae to biofuel. (Download from https://weeklyspy.com/2019/11/25/algae-biofuel-technologies-market-analysis-outlook-2019-2025-industry-growth-top-manufactures-key-segments-and-trends/)

## 1.2 β-glucosidase is responsible for the glucose production from brown algae

Brown algae is considered the promising stocks for producing biofuel and laminarin（Fig 3）is the major storage polysaccharide isolated from brown algae, consisting mostly of linear β-1,3-linked glucose units with small amounts of β-1,6-linkages (Wei *et al.*, 2013). Saccharification of complex polymers to simple sugars is performed by the synergistic action of various different hydrolytic enzymes like cellulases (exoglucanase, endoglucanase, β-glucosidase) and hemicelulases. β-glucosidase breaks cellobiose and cello-oligosaccharides into glucose molecules, playing the final role for accomplishing the degradation of cellulose. Therefore, β-

glucosidase is the final enzyme in lignocellulose degradation process, which determines the hydrolysis rate of the total conversion of lignocellulose into glucose and plays the significant role in the biofuel industry.

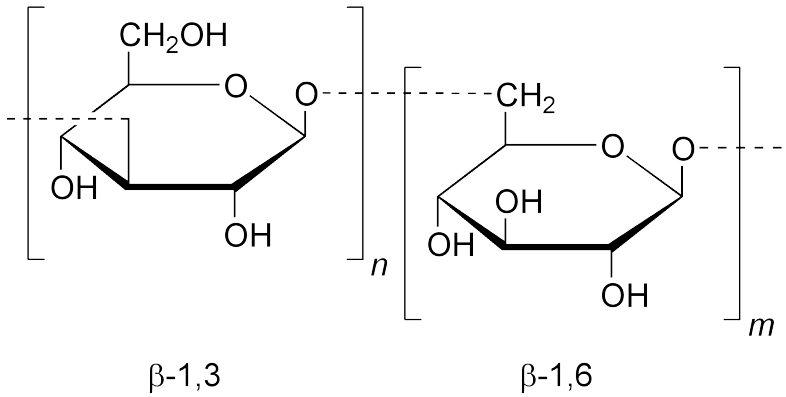



Fig 3. Chemical structure of laminarin (Wei *et al.*, 2013).

*β*-glucosidases (BGL; EC 3.2.1.21) is a group of enzymes that catalyze the hydrolysis of glycosidic bond to release nonreducing terminal glucosyl residues from glycosides and oligosaccharides. It is found among plants, animals, fungi, bacteria and yeasts predominantly. The nonreducing terminal *β*-linked D-glucosyl residues in nature varies, resulting in the assignment of different E.C. numbers for enzymes that hydrolyze glycosidic bond, such as glucosyl ceramidase (3.2.1.45), glucan 1,4- *β* -glucosidases (3.2.1.58), glucan 1,3- *β* -glucosidases (3.2.1.74), steryl- *β*-glucosidase (3.2.1.104), strictosidine *β* -glucosidase (3.2.1.105), amygdalin hydrolase (3.2.1.117), prunasinn hydrolase (3.2.1.118), vicianin *β* -glucosidase (3.2.1.119), raucaffricine *β* -glucosidase (3.2.1.125), and coniferin *β* -glucosidase (3.2.1.126) (Ketudat Cairns & Esen, 2010).

According to the CAZY system (http://www.cazy.org), there are 166 glycoside hydrolase families listed. The *β*-glucosidase that have been found existing in glycoside

hydrolase families GH1, GH3, GH5, GH9, and GH30 (Ketudat Cairns & Esen, 2010). The largest number of characterized $\beta$-glucosidase are belonged to GGH1 family. They utilized the retaining mechanism to perform catalysis (Fig 4) reaction. In the first step, catalytic acid glutamic acid donates a proton to the leaving group, while the catalytic nucleophile attacks from the opposite side. In the next step, the catalytic base glutamic acid extracts a proton from a water molecule, improving its nucleophilic power to attack at the anomeric carbon and displace the enzyme.

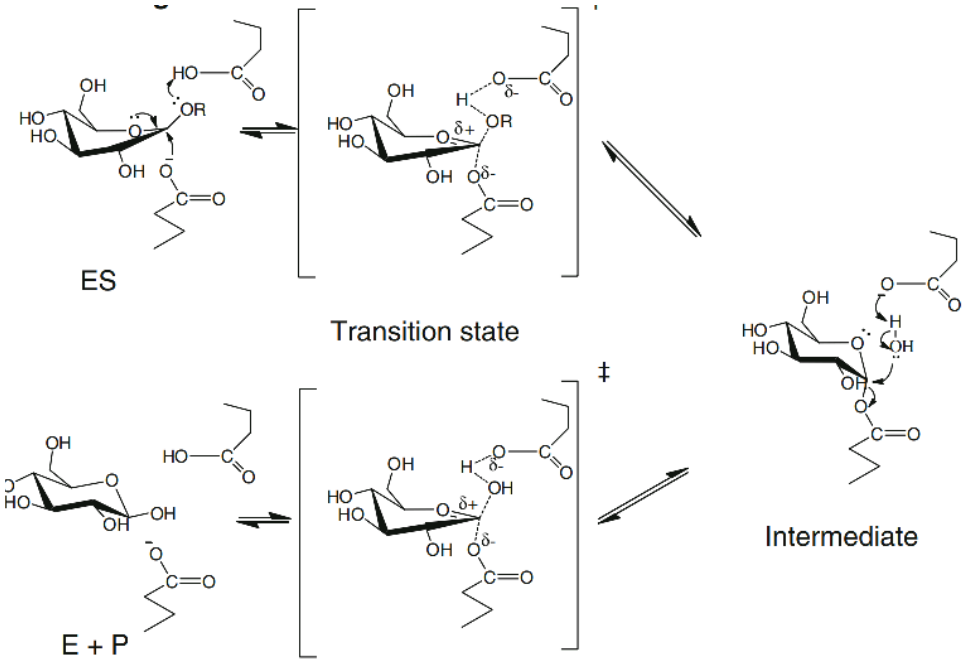



Fig 4 Retaining catalytic mechanism of glucosidase (Ketudat Cairns & Esen, 2010).

The sea hare *Aplysia kurodai* (*A.kurodai*) consumes brown algae *Eisenia bicyclis* as the staple food, liberating large amount of glucose (Tsuji *et al.*, 2017). Therefore, *A.kurodai* is regarded as the excellent model for investigating the glucose production process from brown algae. Firstly, the digestive fluid of *A.kurodai* incubated with brown algae and there is a lot of glucose production. To understand what exactly

enzyme responsible for this reaction, the 110 kDa and 210 kDa *A.kurodai* *β*-glucosidases (BGLs) were identified in the fluid digestive of sea hare *A.kurodai* (Tsuji *et al.*, 2013) that catalyzing the hydrolysis reaction. Both BGLs preferentially hydrolyze the short cellulose fragment produced by endoglucanase digestion. Furthermore, both BGLs also exhibit β-galactosidase and lactase activities. Up to now, only the primary sequence of 110 kDa BGL (*aku*BGL)was published (GeneBank: BAX08664.1). Sequence analysis of *aku*BGL indicated that it contains a signal peptide in the N-terminus and followed by two tandem GH1 domains, which might represent a specific adaption towards brown algae.

Most BGLs belonging to GH1 family contain only one GH1 domain. There are few instances that possess more than one GH1 domain. The mammalian lactase phlorizin hydrolase (LPH) possesses four tandem GH1 domains in the immature form followed by a transmembrane sequence. The first two GH1 domains of LPH protein has shown to act as a chaperone that facilitate the folding of the second half and only two domains (domain 3 and domain 4) are found in the mature protein. Another protein is BGL from shipworm. The shipworm gene encodes an N-terminal signal peptide, followed by six GH1 domains and does not contain transmembrane sequence (Sabbadin *et al.*, 2018). Different from human LPH, all six GH1 domains of shipworm are retained in mature protein. Sequence alignment result indicated that the domains 2, 4, 5 and 6 possess the conserved amino acids for exerting hydrolytic activity, while domains 1 and 3 lack these residues and seems do not have hydrolysis activity and might associates

with protein folding. Another protein identified in the digestive gland of the clam *Corbicula japonica* (*Cj*Cel1A), consisted of two tandem GH1 domains. Based on the amino acid sequence analysis result, it was indicated that the first GH1 domain of the protein does not contain the glycoside hydrolase activity and might instead function as a chaperone, similar to the human LPH. LPH-like sequences are also found in other species, such as insects, reptiles, birds, amphibians, mammals and fish, but not in bacteria, fungi and some animal taxa (Sabbadin *et al.*, 2018) (Fig 5). The multi-domain GH1s from vertebrates typically possess a transmembrane region at C-terminal. Absent of this transmembrane region is a general feature in multi-domain GH1s from invertebrates.

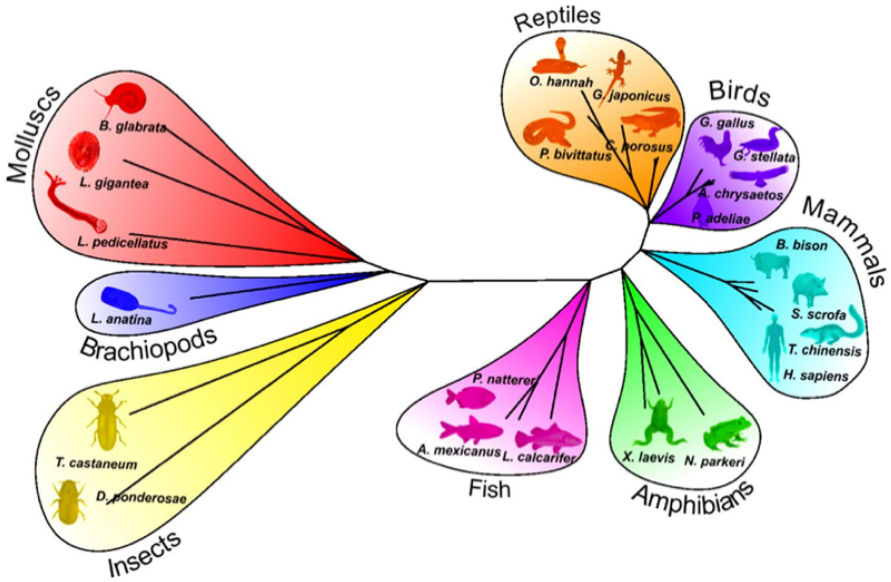



Fig 5. Radial phylogeny of a subset of multi-domain GH1 protein (Sabbadin *et al.*, 2018).

## 1.3 EHEP protects BGL from phlorotannin inhibition

As mentioned above, the sea hare *A.korodai* can digest the brown algae, which

contained high amount of polysaccharide, laminarin to glucose at the function of the *aku*BGLs. However, if the brown alge *Eisenia bicylis* direct incubate with the purified *aku*BGLs, there is no glucose liberated (Tsuji *et al.*, 2017). This was resulted from the phlorotannin, one kind of poly phenol, which are abundant in the brown algae (La Barre *et al.*, 2010, Van Alstyne & Paul, 1990). Phlorotannin binds to *aku*BGLs and thus inhibit its hydrolysis activity. Moreover, phlorotannin also inhibits the activity of several other enzymes, such as hyaluronidase (Shibata *et al.*, 2002), α-glucosidase and α-amylase (Kellogg *et al.*, 2014). How did the sea hare *A.kurodai* overcome the phlorotannin inhibition problem? There should be some special protein in the digestive fluid of *A.kurodai.*

Interestingly, *Eisenia* hydrolysis-enhancing protein (EHEP) was identified in the digestive of *A.kurodai* a few years ago. EHEP is a novel and new protein, because there is no homologous protein present in other organisms. The EHEP contains the motif $CX_{15}CX_5CX_9CX_{12}CX_{5\text{-}9}C$ that is found in peritrophin-A-type binding domains. These regions exhibit sequence homology to a part of the Chitin-binding domains, such as peritrophin-1, peritrophin-95 and tachycitin. However, EHEP exhibited some different characters, it cannot bind to chitin nor have the antimicrobial activity. EHEP was identified to increase the *E.bicyclis* saccharification by *aku*BGL. When *E.bicyclis* was incubated with *aku*BGLs in the presence of EHEP, glucose production was increased prominently, indicating that EHEP protects *aku*BGLs from phlorotannin inhibition. EHEP binds to phlorotannin and forms insoluble complex (Fig 6) (Tsuji *et al.*, 2017).

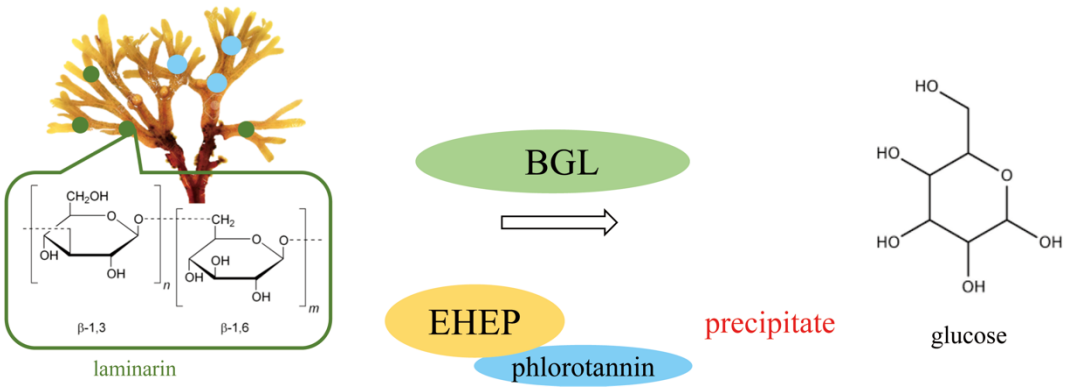



Fig 6 EHEP protects BGL from inhibition.

## 1.4 Phlorotannin and its analogs

Phlorotannins are polyphenol, abundant in brown algae and are formed by the polymerization of phloroglucinol (1,3,5-trihydroxybenzene) monomer units and biosynthesized through the acetate-malonate pathway. Brown algae accumulate a variety of phloroglucinol-based polyphenols. According to the methods of linkage, phlorotannins can be classified into four subgroups: fuhalol and phlorethols (phlorotannin with an ether linkage), fucol (with a phenyl linkage), fucophloroethols (with an ether and phenyl linkage) and eckol (with a dibenzodioxin linkage). Some compounds of phlorotannins that have characterized from *E.bicyclis* are listed in Fig 7 (Jung *et al.*, 2010) (Li *et al.*, 2017).

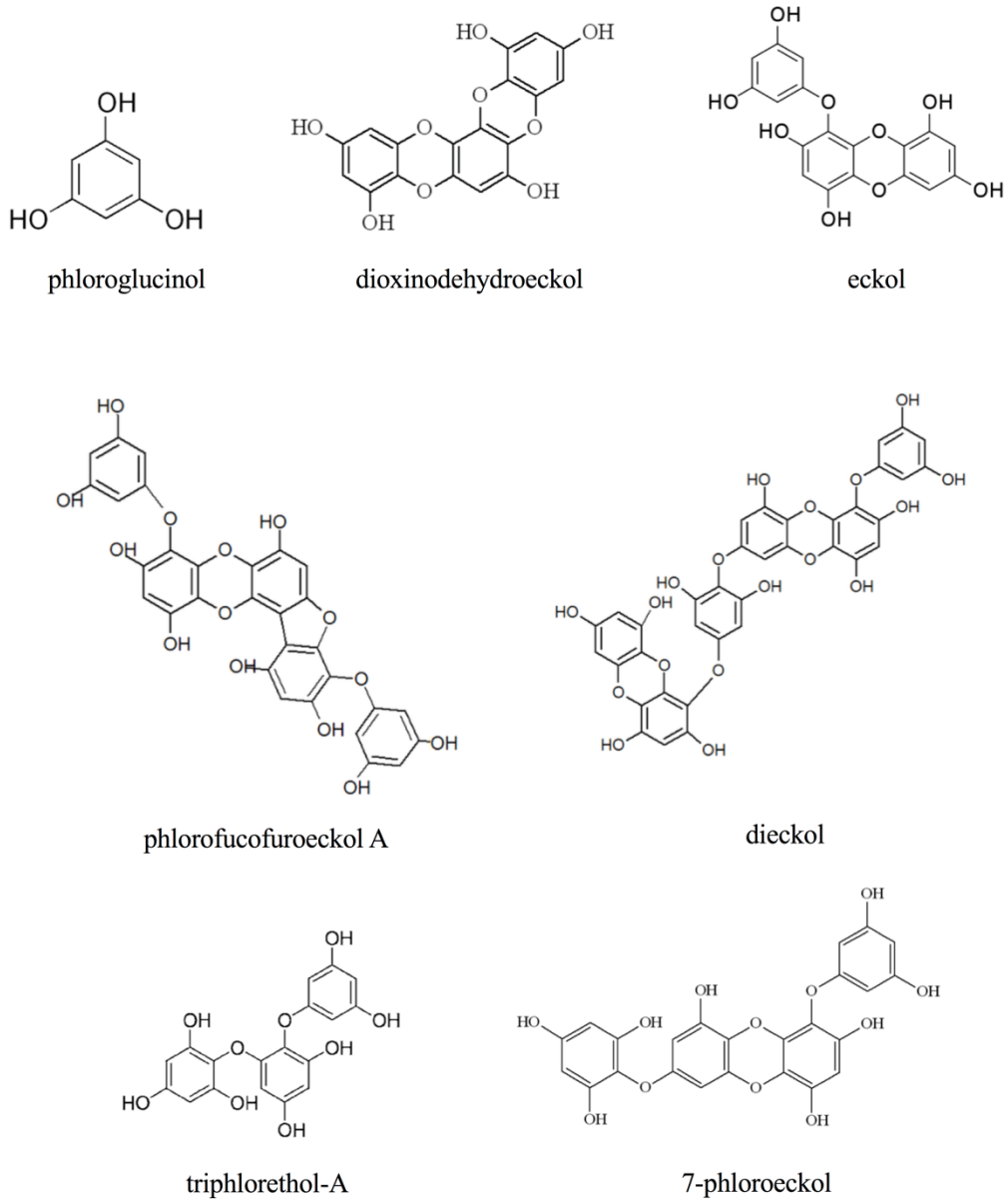



Fig 7. Structures of some phlorotannin derived from marine brown algae *E.bicyclis* (Jung *et al.*, 2010).

Phlorotannin are structurally analogous to terrestrial plant tannins, such as tannic acid. It is a plant polyphenol and has a structure of deca-galloyl glucose consisting of a center glucose molecule esterified at all five hydroxyl moieties with two gallic acid molecules ( Fig 8). Tannic acid also can inhibit the *Kluyveromyces lactis* β-galactosidase activity (Kayukawa *et al.*, 2019).

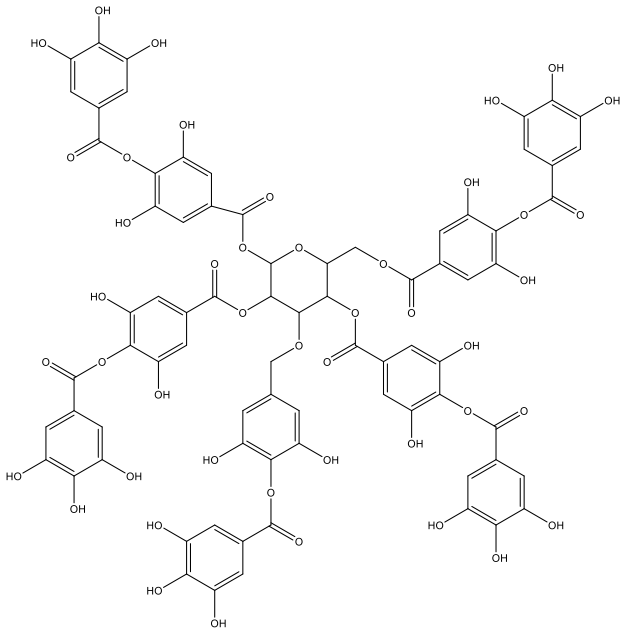


Fig 8. Chemical structure of tannic acid.

## 1.5 Purpose

EHEP is identified in the digestive fluid of *A. kurodai*, is found recently to protect *aku*BGL from this inhibition by binding and then precipitating with phlorotannin, despite not possessing catalytic function. Hence, EHEP and *aku*BGL have high potential for the application in the biofuel industry. However, how does the EHEP bind with phlorotannin for protecting BGL from phlorotannin inhibition and how does the phlorotannin inhibit the BGL activity both remain unknown. An understanding of phlorotannin-binding manner with EHEP and *aku*BGL will help us elucidate protective mechanism of EHEP and the phlorotannin-inhibitory mechanism of *aku*BGL. Consequently, three-dimensional structure could provide information on how to increase the activity of *aku*BGL and how EHEP can be used recycled. The purpose of the research is to reveal the binding manner of EHEP with phlorotannin and also the

inhibition mechanism of BGL with phlorotannin by structure analysis ( Fig 9).

In this research, we explore the effect of tannic acid (analogue of phlorotannin) on the activity of *aku*BGL in chapter 2. In chapter 3, we determined the structures of EHEP and also its complex with tannic acid at a resolution of 1.2 Å and 1.9 Å, respectively. The binding manner was elucidated from the structure, which leads us to re-cycle of the EHEP. In chapter 4, *aku*BGL structure was determined at a resolution of 2.7 Å.

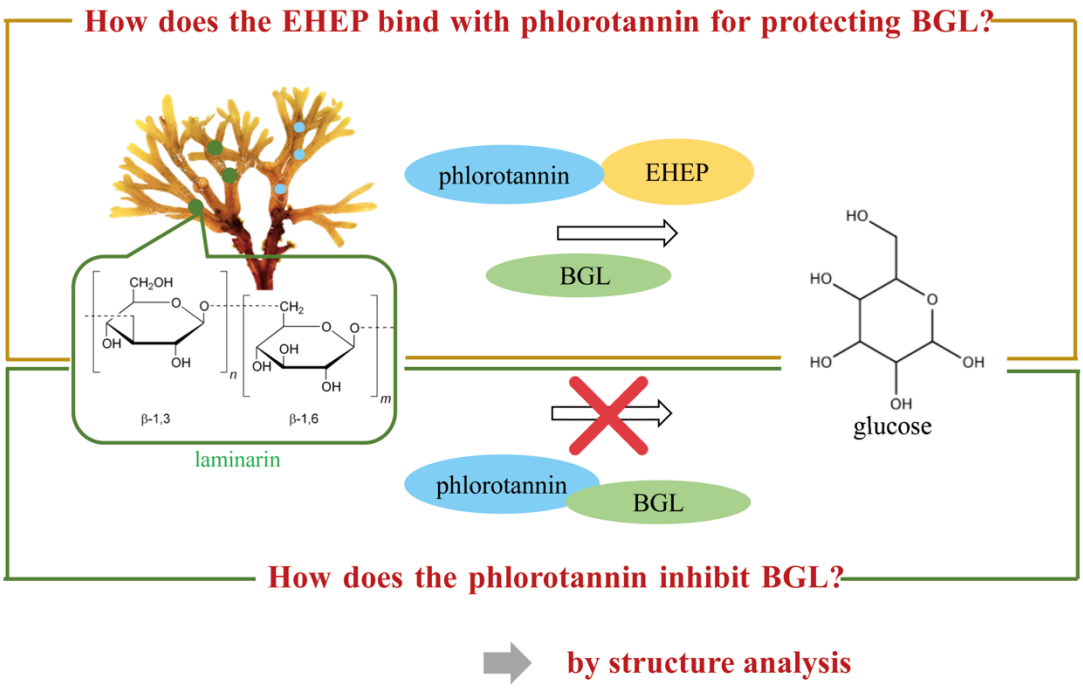



Fig 9 The purpose of this study.

# Chapter 2 The inhibition and protective activity of TNA and EHEP with akuBGL

akuBGL is responsible for catalyzing the saccharification process of *A.kurodai* digestion brown algae process. However, its activity is inhibited by phlorotannin, a polyphenol, which is abundant in brown algae. It is difficult to get from nature and we would like to use the analogue of phlorotannin to study the inhibition mechanism. Tannic acid is a plant polyphenol and consisting of ten gallic acids with a core of glucose and is an analogue of phlorotannin. Therefore, in this part, we investigate the inhibition activity of tannic acid with *aku*BGL and also protective activity of EHEP with *aku*BGL.

## 2.1Material and methods

### 2.1.1 Preparation of ONPG, ONG, tannic acid, EHEP and BGL

Tannic acid (TNA; >98% purity, molar mass of 1701.19 g/mol), o-Nitrophenol (ONP, >98% purity, molar mass of 139.11 g/mol) and ortho-Nitrophenyl-β-galactoside (ONPG; >98% purity, molar mass of 301.25 g/mol) used in this study were purchased from Wako (Japan) company. TNA was dissolved in buffer containing 50 mM CH3COONa, 100mM NaCl, 10 mM CaCl2. For measuring the activity of akuBGL, ONPG was used as substrate. The stock solution of ONPG was prepared as follows: 10

mg ONPG was dissolved into 0.66 mL N, N-dimethylformamide, and stored from light at -30 ℃. The stock solution of ONP was prepared as follows: 6.95 mg ONPG was dissolved into 10 mL MilliQ, and heated until dissolve completely. Native EHEP and akuBGL were provided by Prof. Tsuji (Tokushima University, Japan).

### 2.1.2 The reaction system of TNA inhibits *aku*BGL activity

To explore the inhibition activity of TNA with *aku*BGL, the ONPG was used the substrate to measure the hydrolytic activity of *aku*BGL. A standard ONP curve was constructed. For the enzyme activity assays, the reaction system is shown in the Table 1. The mixtue except *aku*BGL was incubated at room temperature for 30 min, and then *aku*BGL was added. Next, the polypropylene tubes were immediately placed in a thermostated bath (Astec, Japan) at 37 ℃. After 10 min incubation, 100 μL methanol solution was added to each of the samples, to terminate the reaction. Then, the mix was centrifuged at 12000 rpm for 10min at 4 ℃ and the supernatant was used for HPLC analysis.

Table 1 Reaction system of TNA inhibit BGL activity

| Reagent | volume (μL) |
|---|---|
| 50 mM ONPG | 5 |
| BGL 25 nM | 2 |
| TNA 1mM | 0/2/4/6 |
| 50 mM $CH_3COONa$, 100mM NaCl, 10 mM $CaCl_2$ | up to 100 |

### 2.1.3 Reaction system of EHEP protect BGL from inhibition

For the activity of EHEP protecting BGL from inhibition, the reaction system is shown in Table 2. The TNA, buffer, and EHEP was mixed first and incubate at room temperature for 30min and then the other components were added. The subsequent steps were conducted as the same as in the section 1.2.

Table 2 Reaction system of EHEP protect BGL from inhibition activity

| Reagent | volume (μL) |
|---|---|
| 50 mM ONPG | 5 |
| BGL 25 nM | 2 |
| TNA 1mM | 4 |
| EHEP | 0-5.92 μM |
| 50 mM $CH_3COONa$, 100mM NaCl, 10 mM $CaCl_2$ | up to 100 |

### 2.1.4 HPLC analysis

The HPLC system (JASCO, Japan) consisted of PU-4185-binary pump, an AS-4050 autosampler and a spectrum focus UV- Vis detector. Solvents were HPLC grade. Sample separation was performed on a reversed phase (C18, 150 mm ×4.6 mm, 5$\mu$m) column thermostatted at 40 ℃ using mobile phase A (MilliQ, 0.1% HCOOH) and mobile phase B (methanol, 0.1% HCOOH) in a gradient program with a flow of 1 mL/min: 0-30 min: 5% - 100% B; 30-45min, 50% B; 45-60min, 5% B. For quantitative analysis, the wavelength of 290 nm was used. The injection volume was 10 $\mu$L.

## 2.2 Results & discussion

### 2.2.1 TNA inhibits the activity of *aku*BGL

akuBGL activity is inhibited by phlorotannin in nature. However, phlorotannin is difficult to get from brown algae, we would like to use TNA instead of phlorotannin to explore its inhibition mechanism. akuBGL could recognize many substrates, such as 4-MU-β-glucoside, 4-MU-β-galactoside, cellobiose, lichenan, and laminarin (Tsuji et al., 2013). Therefore, we investigate the TNA inhibits the hydrolysis activity of akuBGL using ONPG as the substrate. To quantify the product o-Nitrophenol (ONP), we make a standard curve first.

With the optimized HPLC conditions, the calibration curve was plotted using 0.05, 0.1, 0.5, 1, 2.5 mM concentration of ONP. The calibration curve showed good linearity. The calibration curve for ONP was found to be linear in the concentration range of 0.1 to 2.5 mM with regression value of 0.9994 (Fig 10). The equation derived from three replicates was y=634759x + 15404. The separation of the product ONP and ONPG under optimized conditions is shown in Fig 11 .

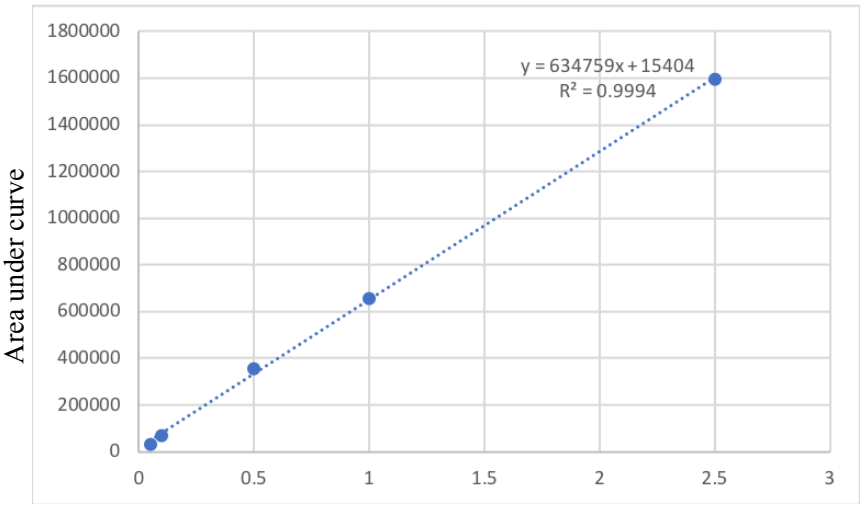



Fig 10. Standard curve for ONP determined by HPLC.

Concentration of ONP (mM)

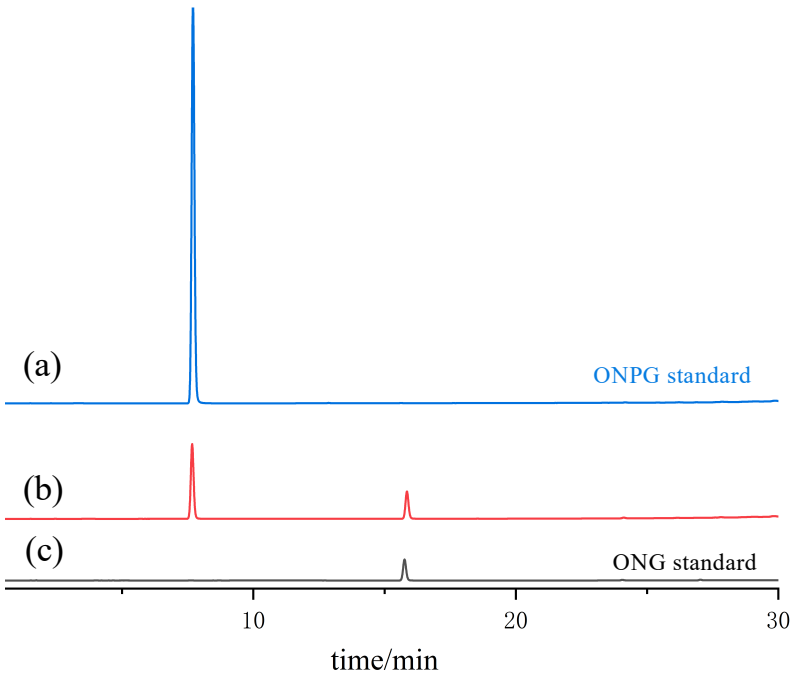



Fig 11. The result of *aku*BGL activity assay by HPLC analysis. (a) HPLC profile of ONPG standard. (b) HPLC profile of *aku*BGL products, (c) HPLC profile of ONG standard.

The interaction of tannic acid with *aku*BGL is presented in Fig 12. As tannic acid concentration increased, the activity of *aku*BGL decreased concomitantly. Thus, tannic acid inhibits the *aku*BGL activity in a dose-dependent manner. The concentration of

TNA that inhibits the 70% activity of the *aku*BGL is 40 μM. Fucofuroekol (FF) and dioxinodehydroeckol (DD) were two kinds of phlorotannin extracted from *E.bicyclis*. FF showed inhibition activity towards of which α-glucosidase and α-amylase, with $IC_{50}$ of 131.34 nmol/L and 42.9 $\mu$mol/L. Similarly, DD also exhibits an inhibition effect on α-glucosidase and α-amylase, with $IC_{50}$ of 93.33 nmol/L and 472.7 $\mu$mol/L. The hydroxyl groups in phlorotannin derivatives may important for having the inhibitory activity(Eom *et al.*, 2012). Except digestion enzymes, tannic acid can also inhibit the activity of acid phosphatase, arylsulfatase and chitinase at high concentrations. Surprisingly, tannic acid did not affect the activity of $\beta$- glucosidase much. Low centration of tannic acid even can speed-up the enzymatic activity, such as acid phosphatase, arylsulfatase, chitinase (Adamczyk *et al.*, 2017). The different influence of tannic acid on the activity of enzymes may be caused by the different binding affinity between them and tannic acid can affect the protein secondary structure.

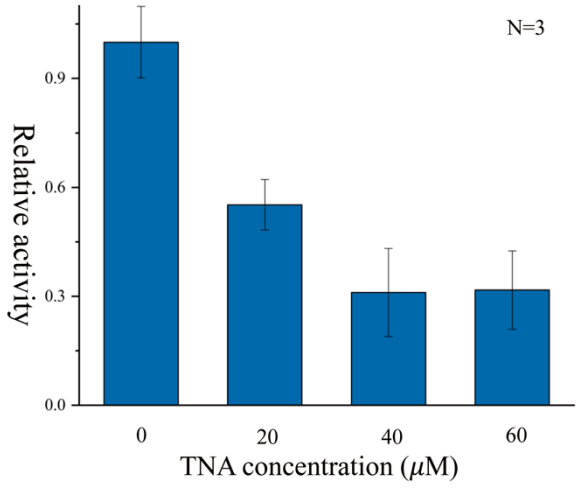



Fig 12 Inhibitory effect of different levels of TNA on *aku*BGL.

### 2.2.2 EHEP protects BGL from TNA inhibition

Next, we investigate the protective activity of EHEP towards the akuBGL inhibition by TNA. As EHEP concentration increased, the activity of akuBGL increased concomitantly( Fig 13). Thus, EHEP protects the akuBGL activity from tannic acid inhibition in a dose-dependent manner. The concentration of EHEP recovery the 80% activity of akuBGL is 2.96 μM. Thus, in the next few parts, we used tannic acid instead of phlorotannin to explore its binding manner with EHEP and BGL.

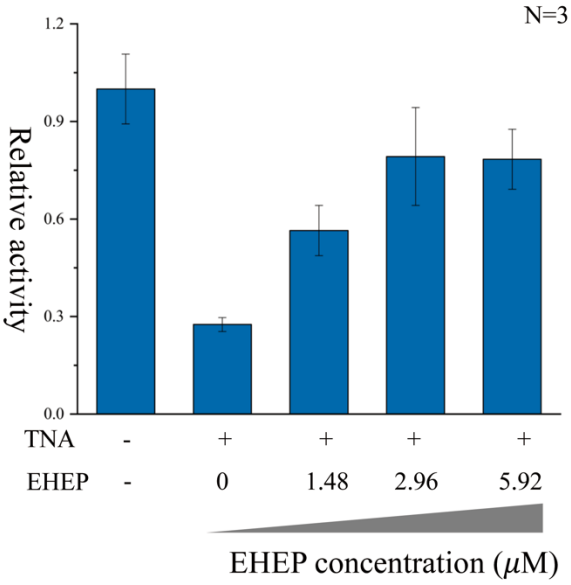



Fig 13. Effect of different levels of EHEP on TNA inhibition *aku*BGL

Even though it has been known for a long time that phenol can inhibit the activity of several different enzymes. Up to now, to our knowledge, EHEP is the first protein that can protect the enzyme from inhibition by binding with phenol. Therefore, EHEP might have a pronounced meaning in industry application. It needs further experiment to investigate the protection effect of EHEP on other enzymes.

# Chapter 3 Binding mechanism of EHEP with TNA

Although we have known that TNA can instead of phlorotannin to inhibit the activity of *aku*BGL and EHEP can protect it from inhibition. It still remains unclear the binding mechanism of EHEP with TNA. In this chapter, we tried to answer this by the structural analysis of the complex structure of EHEP with TNA.

## 3.1 Materials and methods

### 3.1.1 Purification of EHEP

Native EHEP (25 kDa) (GenBank: BAV38197.1) was purified from the natural digestive fluid of *A. kurodai* as described previously (Tsuji *et al.*, 2017). This protein was then further purified using size-exclusion chromatography (SEC) with a Superdex 200 10/300 GL column (GE Healthcare), which was equilibrated with 20 mM sodium acetate buffer containing 100 mM NaCl (pH 6.0). The purified EHEP was examined by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and was then concentrated to 15–25 mg/ml using Vivaspin-4 columns (Sartorius, Göttingen, Germany) with a molecular weight cut-off of 5 kDa.

### 3.1.2 Crystallization and data collection

The initial crystallization screening was performed using the sitting-drop vapor-diffusion method with JCSG Core I-IV and PEGs crystallization kits from Qiagen

(Hilden, Germany) at 20°C. Protein solution (1 μL) was mixed with an equal volume of reservoir solution. Initial crystals were obtained under condition no. 48 [0.2 M lithium sulfate and 20% polyethylene glycol (PEG) 3350] of JCSG Core II, condition no. 7 (0.2 M disodium hydrogen phosphate and 20% PEG 3350) of JCSG Core III, and condition no. 40 [0.1 M HEPES (pH 7.5) and 25% PEG 8000] of PEGs. However, crystals formation could not be replicated under these conditions. Therefore, the crystallization screening was reattempted using the Crystal Screen crystallization kit (Hampton, USA), which resulted in crystals appearing under condition no. 25 [1.0 M sodium acetate and 0.1 M imidazole (pH 6.5)] and condition no. 37 (8% PEG 4000 and 0.1 M sodium acetate). Unfortunately, the crystallization conditions could not be further improved due to the poor reproducibility of crystallization.

To prepare the crystals of EHEP complexed with substrate, we tried to soak crystals in a solution containing 10 mM tannic acid (TNA) at high temperature because the substrate-binding affinity is stronger at high temperature than at room temperature. So that, we crystallized EHEP under condition of 1.0 M sodium acetate and 0.1 M imidazole pH 6.5 with additive material of co-cage-1 and obtain high quality crystals. The EHEP-TNA complex crystals were obtained by soaking crystals in a solution containing 10 mM tannic acid for three months at 310 K (37°) for 2 days and subsequently kept to 293 K (20°) for 2 weeks. For data collection, the crystal was soaked in cryoprotectant solution containing 20% (v/v) glycerol with reservoir solution and placed under a cold nitrogen gas stream at 100 K.

For data collection, the crystals obtained by the initial crystallization screening were soaked in cryoprotectant solution containing reservoir solution supplemented with 20% (v/v) glycerol and placed under a cold nitrogen gas stream at 100 K. Native diffraction data were then collected at a wavelength of 1.0 Å on the beamline BL-5A at Photon Factory (Tsukuba, Japan). For determining EHFP structure, native-SAD data sets were collected using a long wavelength of 2.1 Å on the beamline BL-17A at Photon Factory. To reduce solution absorption at a long wavelength, the solution-less crystal mount method was applied for native SAD data collection (Kitago *et al.*, 2005) using solution-less crystal mount tools (solution-less loop and extraction-freezing robot AFERO (SYSCON Inc. Japan) developed in our laboratory. To obtain a strong anomalous signal with high redundancy, eight native-SAD data sets (720 degrees per data set) were collected through the exposure of eight positions on a single crystal.

For EHEP-TNA complex crystal, diffraction data were collected on beamline BL-17A at Photon Factory at a wavelength of 1.0 Å. The statistics of data collection and process were summarized in Table 3.

### 3.1.3 Structure solution and refinement

All of the data sets were then indexed, integrated, scaled, and merged using the *XDS/XSCALE* program (Kabsch, 2010) (Sheldrick, 2010). This initial model was used as a search model for molecular replacement using *Phaser* for determining a native structure. Full structure of EHEP was automatically built using *Phenix.autobuil* in the

*Phenix* program suite and several rounds of refinement were performed using *Phenix.refine* in the *Phenix* program suite alternating with manual fitting and rebuilding using *COOT* at 1.15 Å. The final refinement statistics and geometry are shown in Table 3.

The structure of EHEP-TNA complex was determined by molecular replacement method using the structure of EHEP as the search model using *Phaser*. Several rounds of refinement were performed using *Phenix.refine* alternating with manual fitting and rebuilding using *COOT* (Adams *et al.*, 2010) (Emsley & Cowtan, 2004), electron density block of TNA were showed clearly in both $2F_o-F_c$ and $F_o-F_c$ maps. The structure of TNA was built manual. The final refinement statistics and geometry are shown in Table 3.

### 3.1.4 MALDI-TOF MASS spectrum analysis

Previous study has predicted that native EHEP possesses a signal peptide in the N-terminus(Tsuji *et al.*, 2017). To confirm the molecular weight of mature native EHEP, MALDI-TOF mass spectrometry analysis of purified EHEP was carried out using an Autoflex III Smartbeam mass spectrometer (Bruker). After SEC purification, the EHEP was demineralized using ZipTip (Merk) pipette tips packed with C4 resin. The sample for MALDI-TOF mass spectrometry was then prepared by mixing 1 $\mu$L EHEP with 1 $\mu$L matrix (sinapinic acid). The sample was applied to the MALDI target plates and dried with a drier for several minutes. The spectra were then automatically acquired

and analyzed using the Flex Analysis software.

### 3.1.5 Expression and purification of recombinant EHEP

The cDNA of EHEP was amplified using the high-fidelity DNA polymerase KOD according to the protocol. The PCR product was analyzed by agarose gel electrophoresis and purified by Qiagen kit. The purified fragment was digested by restriction enzyme NdeI and XhoI and then purified. The fragment was ligated to expression vector derived from pET28a (Novagene, Darmstadt, Germany) or pMAL-p2E expression vector respectively using T4 ligase. The ligation product was transformed into *E.Coli* DH5α and spread into LB plate containing Km or Amp antibiotic, incubating at 37 ℃ overnight. Then colonies were used as template to perform the colony PCR. The PCR products were checked by agarose gel electrophoresis. The colony corresponding to the right band was sequenced. The recombinant plasmid was used for expression.

*E.Coli* B834DE3pRAE2 cells were transformed with each expression vector, and cultured in LB medium containing 100mg/L Km or Amp for the pMAL-p2E expression vector or 50 mg/L kannmycin for the pET28a derived expression vector at 37 ℃. When the optical density of the culture at 600 nm reached about 0.6, isopropyl-β-thioga-lactopyranodise (IPTG) was added to a final concentration of 0.5mM to induce protein expression and the culture was further continued for 20 h at16 ℃. The cells transformed with pET-28 derived vector was disrupted by sonication in buffer A, containing 50 mM

Tris-HCl (7.4), 300mM NaCl, and purified by a Histrap HP 5mL column (GE Healthcare). Proteins were eluted with buffer B containing 50 mM Tris-HCl (7.4), 300mM NaCl, 500 mM imidazole. The fractions corresponding to peaks of absorbance at 280 nm were checked by SDS-PAGE. The fractions containing the desired protein was dialyzed to buffer C (50 mM Tris-HCl (7.4), 5mM NaCl) and loaded onto Hitrap Q column (GE Healthcare). EHEP was eluted by increasing the concentration of buffer D (50 mM Tris-HCl (7.4), 1M NaCl). The peaks were checked by SDS-PAGE. The proein was further purified by a Hiload 16/60 superdex 75 column (GE Healthcare) in 20 mM sodium acetate buffer containing 100 mM NaCl (pH 6.0). The cells transformed with the pMAL-p2E expression vector was disrupted by sonication in 20 mM Tris-HCl buffer (pH 7.4) containing 200 mM NaCl, 1mM EDTA and purified by a MBP trap column 1mL (GE Healthcare) by eluting by the buffer B (20 mM Tris-HCl buffer (pH 7.4) containing 200 mM NaCl, 1mM EDTA, 10mM maltose). Then, the MBP tag of MBP-EHEP was cleaved by TEV protease, and the sample was further purified by a Hiload 16/60 superdex 75 column (GE Healthcare) in 20 mM sodium acetate buffer containing 100 mM NaCl (pH 6.0). Finally, the purified recombinant EHEP was stored in 4 ℃.

### 3.1.6 Activity of recombinant EHEP

The activity of recombinant EHEP was measured using the method published previous (Tsuji *et al.*, 2017). Briefly, the recombinant EHEP was incubated with tannic

acid in buffer 50 mM acetate sodium (pH 5.5) containing 0.1 M NaCl and 10 mM $CaCl_2$ at 25 ℃ for 90 min. Then the mixture was centrifugated at 12000g for 10min. The resulting precipitate was washed twice with the reaction buffer and then resuspended in the SDS-PAGE loading buffer, treated at 95 ℃ for 10 min and then applied 15% SDS-PAGE for analysis.

### 3.1.7 Isothermal titration calorimetry (ITC)

Experiments were carried out at 25 ℃ in 20 mM $CH_3COONa$, 100 mM NaCl, 10 mM $CaCl_2$ using Nano ITC instrument. The EHEP 0.1 mM filled in the cell was titrated with 5 mM tannic acid. All the solutions were filtered and degassed before use. The reference experiment was performed by titrating 5mM tannic acid into buffer without EHEP and was subtracted from titration experiments (Deaville *et al.*, 2007). 2$\mu$L tannic acid was delivered into the cell with an interval of 3min between injections. The data was analyzed using the software NanoAnalyze.

### 3.1.8 Molecular docking

The Schrondinger-Maestro (Sastry *et al.*, 2013) program was used for the docking study. The ligand used for docking were prepared using LigPrep module in Schrodinger Maestro. Glide (Grid-based ligand docking with energetics) was used to docking calculations. The glide score was used to choose the best docked structure from the output. The docked structures were analyzed using PyMol.

### 3.1.9 Dissolve of precipitate of EHEP-TNA

The precipitate was produced as in Section 1.5. Then the precipitate was dissolved in buffer 50 mM Tris-HCl (pH 8.0) containing 0.1 M NaCl and 10 mM $CaCl_2$. Several hours later, the mix was become clear and then was centrifugated at 12000g for 10min. The resulting supernatant was analysis by the HPLC. Method is the same as in chapter 1.

Table 3 Data collection and refinement statistics

[a]Values in parentheses are for the highest resolution shell.

| | Native-SAD of EHEP | Native of EHEP | EHEP-TNA |
|---|---|---|---|
| Data Collection | | | |
| Beamline | PF BL17A | PF BL17A | PF BL17A |
| Wavelength (Å) | 2.1000 | 0.9800 | 0.9800 |
| Resolution range (Å) | 50-2.34 (2.48-2.34)[a] | 50–1.15 (1.22–1.15) | 50–1.9 (2.01–1.9) |
| Space group | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ |
| Unit-cell parameters (Å) a, b, c | 42.18, 65.32, 66.51 | 42.13, 65.2，66.49 | 42.46，65.35，67.17 |
| Completeness (%) | 92.0 (62.3) | 93.4（82.2） | 99.7（98.8） |
| Redundancy | 12.6 | 6.64 | 5.69 |
| Average I/σ(I) | 26.04 (6.56) | 19.28 (2.28) | 14.58（1.94） |
| $R_{merge}$[b] | 8.2% (26.2%) | 7.3% (83%) | 8.9% (81.2%) |
| $CC^{1/2}$ | 99.8 (96.4) | 99.9（75.2） | 99.8（74.2） |
| Ano/Sig | 1.3 (1.3) | 0.753(0.722) | 0.798(0.745) |
| Molecules/ asymmetric unit | 1 | 1 | 1 |
| Refinement | | | |
| $R_{work}/R_{free}$ (%)[c] | | 16.45/18.11 | 20.05/22.58 |
| Atoms | | 3003 | 1554 |
| Amino acid residues | | 207 | 205 |
| Water molecules | | 358 | 73 |
| Ligands | | 0 | 1 |
| RMSD from ideality | | | |
| Bond length (Å) | | 0.013 | 0.004 |
| Torsion angle (°) | | 1.469 | 1.214 |
| Ramachandran plot (%) | | | |
| Favoured | | 98.26 | 98.03 |
| Allowed | | 1.74 | 1.97 |
| Outliers | | 0.00 | 0.29 |

[b]$R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I_i(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $i$ is the number of observations of a given reflection and $I(hkl)$ is the average intensity of the $i$ observations. $R_{free}$ was calculated with a 5% fraction of randomly selected reflections evaluated from refinement. The highest resolution shell is shown in parentheses.

[c]$R_{work} = \sum_{hkl}||F_{obs}| - |F_{calc}|| / \sum_{hkl}|F_{obs}|$, $R_{free}$ was calculated for 5% randomly selected test sets that were not used in the refinement.

## 3.2 Results & discussion

### 3.2.1 Overall structure of EHEP

In this study, native EHEP was successfully obtained from *A. kurodai* at a high purity (Fig 14a). Initial crystallization screening produced high-quality crystals for the diffraction experiment. However, it proved difficult to reproduce these crystals even when the same crystallization kits were used or the conditions were expanded around the growth conditions outlined in section 2.2 by adjusting the pH and the concentrations of buffer and precipitants (salt, organic solvent). Therefore, after obtaining a native data set at a resolution of 1.2 Å using the initial screening crystals produced under condition no. 48 (JCSG Core II) (Fig 16a), we attempted to solve the structure of EHEP using the molecular replacement method. Since no homologous protein or domain structure was available [identities <28% for a fragment with about 1/2 of full-length EHEP], we undertook the modeling using *Phyre2* (http://www.sbg.bio.ic.ac.uk/servers/phyre2) as a search model. However, this failed to provide any answers. EHEP is a cysteine-rich protein, and the diffraction ratios ($<\Delta F>/<F>$) of the anomalous scattering of sulfur atoms at wavelengths of 1.9 Å and 2.1 Å were calculated to be 1.91% and 2.28%, respectively (Hendrickson & Teeter, 1981), which is considered sufficient for native-SAD phasing. Therefore, we used the crystal obtained under condition no. 25 (Crystal Screen TM) to collect native-SAD data with a wavelength of 2.1 Å, applying the solution-less crystal mount method to reduce the effect of solution absorption at a long

wavelength and thus increase the signal to noise (S/N) ratio of the diffractions (Fig 15b). A total of 11520 (= 8 × 720/0.5) frames were collected using an oscillation angle of 0.5 degrees, all of which were used for data processing and phasing.

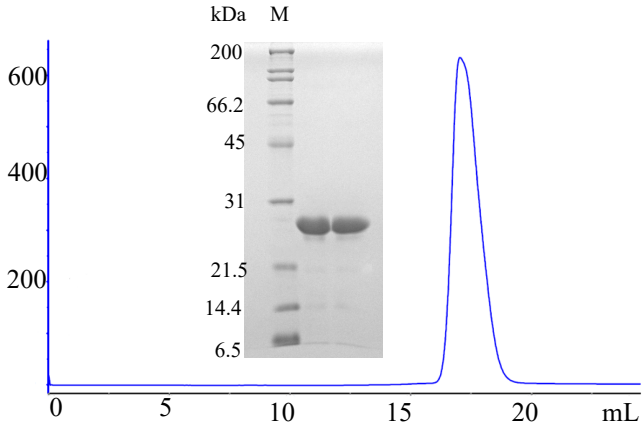



Fig 14. Size-exclusion chromatography of EHEP. Purified EHEP was checked by SDS-PAGE. Lane M: molecular-weight protein markers.

The preliminary X-ray diffraction pattern showed that the crystals belong to space group $P2_12_12_1$. Assuming one molecule per asymmetric unit, the Matthews coefficient (Matthews, 1968) was estimated to be 1.83 $Å^3$/Da, which corresponds to a solvent content of 32.7%. All sites of the 20 sulfur atoms were located using the *SHELXCD* program (Sheldrick, 2010), and model-building was performed using *Phenix.autosol* in the Phenix software suite (Adams *et al.*, 2010) with rephasing and phase-improvement. This resulted in 58% of the structure of EHEP being automatically built. This initial model was used as a search model for molecular replacement using *Phaser* (McCoy *et al.*, 2007) for determining a native structure.

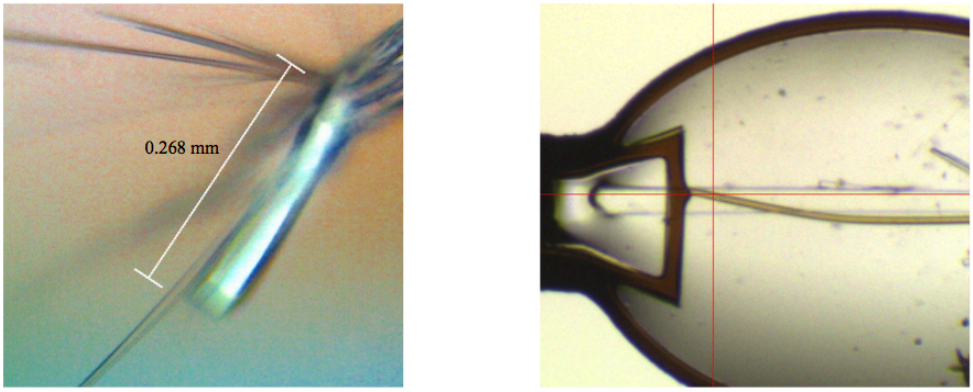



Fig 16. Photo of EHEP crystals. (a) EHEP crystal for collecting native data. (b) EHEP crystal for collecting native-SAD data frozen using solution-less loop.

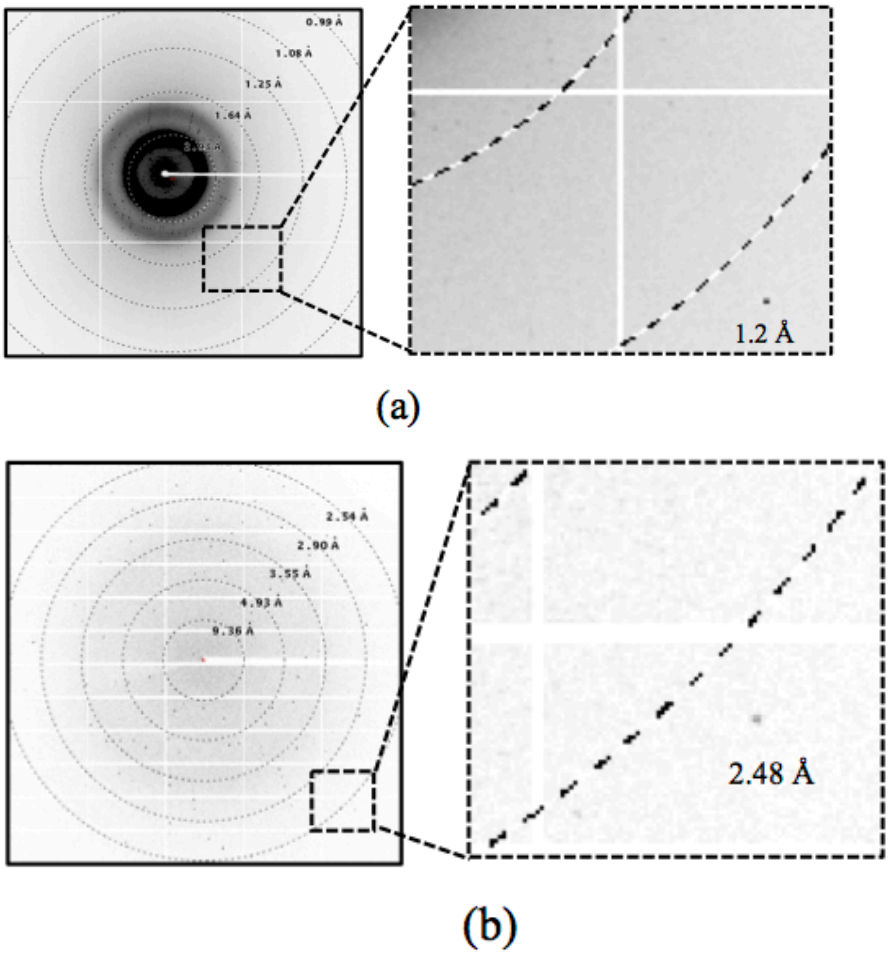



Fig 15. X-ray diffraction pattern of EHEP crystals. (a) Native data. (b) native-SAD data.

The structure of natural EHEP was determined at resolution of 1.8 Å with a $R_{work}$ and $R_{free}$ of 0.16 and 0.18, respectively (Table 3). The amino acids were well modeled except the first 20 amino acids of the protein (1-20) as a result of lacking electron density. EHEP structure consists of three tandem chitin-binding domains (ChBD); ChBD1 (residues 23-79), ChBD2 (residues 92-149) and ChBD3 (residues 167-227). These three domains are linked by two long loops (Fig 17).

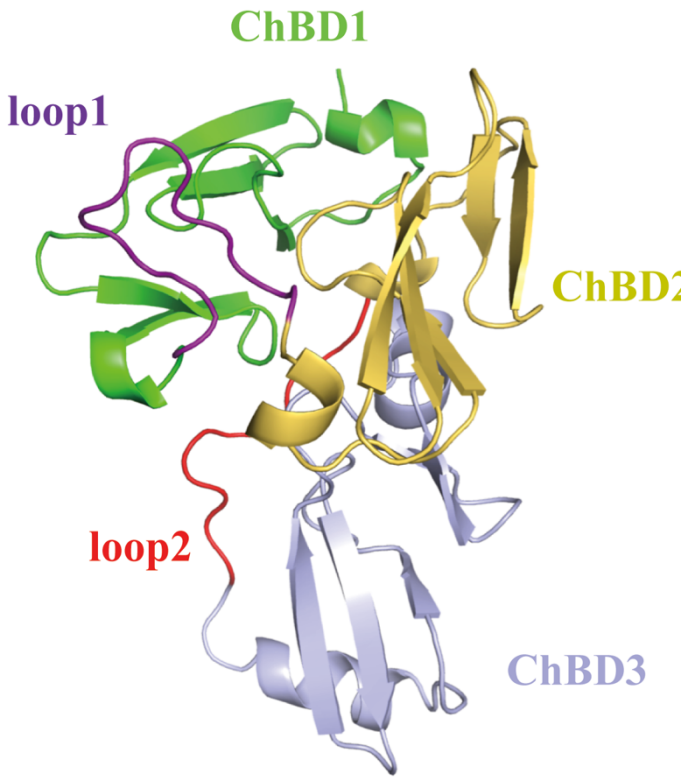



Fig 17. Overall structure of EHEP. Three domains and two linkers are colored by orange, green, light-blue, blue respectively.

Amino acid sequence analysis indicates that all the three ChBD domains belong to Carbohydrate-binding module family 14 (CBM14 family). A search in the carbohydrate-Active enZYmes (CAZy) database reveals that there are seven CBM14 proteins structures are available up to now, two of which are chitinase, while other

proteins are a single domain difference to chitinase. It is of interest to note that the different character for EHEP was that it is consisted of three tandem ChBDs. The three ChBDs almost adopt the similar structure with a root-mean-squared deviation (RMSD) of 0.948 Å over 55 Cα atoms (the core structure), although sequence identify is 21.8% (Fig 18). The ChBD has a typical CBM fold consisting of a $\beta$-sandwich of two $\beta$-sheets containing three N-terminal anti-parallel $\beta$-strands and two C-terminal anti-parallel $\beta$-strands.




Fig 18. Sequence alignment of three ChBD domains of EHEP.

A search of the protein database Dali (Holm & Rosenstrom, 2010) for structural similarity revealed two most similar structures of mite allergen Der p 23 (PDBID 4ZCE)17 and Avr 4 (PDBID 6BN0)18 with RMSD of 1.0 Å, 0.9 Å, respectively. Structure alignment shows the two structures both align with the ChBD1 domain well. Der p 23 shares a similar conformation with ChBD1 domain, except that it lacks the two α helixes at the N and C terminals (Fig 19a). Avr 4 is a chitin-binding protein. The key residues of Avr4 involved in chitin binding are W100 and D102, mutating one of the two residues can abolish the chitin binding activity. W100 provided the main facial interaction between Avr4 and chitin through CH-π bond. and the side chain of D102 forms hydrogen bonds with the chitin. As shown in Fig 19b. D102 and W102, the corresponding residue in EHEP was N69 and D71, respectively. This may partly explain why the EHEP cannot bind to chitin although EHEP shares the similar structure

of ChBD with chitin-binding proteins.

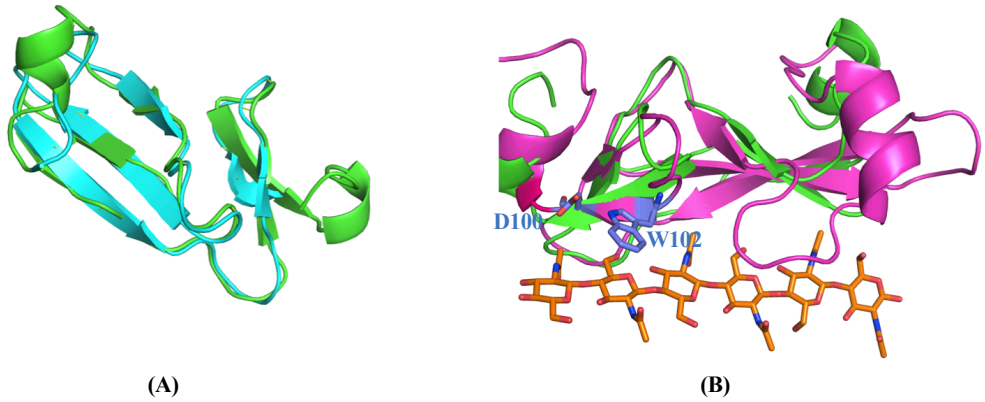



Fig 19. Structure alignment of ChBD1 domain with Dfer P (A) and Avr4 (B). The ChBD1 domain was colored in green, Dfer P was colored in cyan, and Avr4 was colored in magenta. The $(GlcNAc)_6$ bound with Avr4 was drawn in orange representation with stick. The key residues interacting with $(GlcNAc)_6$ were shown in sticks with skyblue.

EHEP is a cystines-rich protein. All three ChBD domains exhibit a motif of $C^{No1}X_{15}C^{No2}X_5C^{No3}X_9C^{No4}X_{12}C^{No5}X_{5\text{-}9}C^{No6}$ (superscript number show the number of cystines and subscript shows residues number between the adjacent cysteines), and these six cystines form disulfide bonds by disulfide pairs: $C^{No1}$-$C^{No3}$, $C^{No2}$-$C^{No6}$, $C^{No4}$-$C^{No5}$. Such rich disulfide bonds can be considered to play the important role for structure folding, even if more than 70% of the EHEP structure is loops. Such motif with disulfide bonds was also found in peritrophin A-type chitin-binding domains existed in tachycitin19, Avr420 and chitinase21. However, Der p 23 contains only 4 cystines formed two disulfide bonds, although Der p 23 shares a very similar structure with ChBD. Another protein allergen Der f 23 (PDB ID: 5ZJL) which also belongs to the CBM14 family but did not have a typical CBM structure, has four cystines in the

primary sequence, however, only one disulfide bond was formed. Taken all sequence and structural characteristics together, it is suggested that the two disulfide bonds ($C^{No1}$-$C^{No3}$, $C^{No4}$-$C^{No5}$) might be conservation structural feature for the traditional CBM folding.

EHEP belongs to CBM14 family that members share CBM folding although have different biochemical functions. For example, tachycitin has both the chitin binding and antibacterial activities(Suetake *et al.*, 2000), allergen Der p 23 neither bind to chitin nor has antibacterial activity(Mueller *et al.*, 2016). Avr4 and human chitinase CHIT1 maintain the chitin-binding activity, but they utilize a very different binding mechanism(Hurlburt *et al.*, 2018, Fadel *et al.*, 2016). CBM classifications has been proposed in which these proteins are grouped into three types by substrate binding mode: surface-binding CBMs (Type A), glycan-chain-binding CBMs (Type B), small sugar binding CBMs (Type C). In all the three mechanisms, the aromatic amino acid, piratically the tryptophan play an important role in the substrate recognition (Guillen *et al.*, 2010), which form stacking interactions with the sugar rings and the conserved residues. EHEP is the first crystal structure that has a three tandem (typical CBM fold) ChBDs organization in CBM14 family.

### 3.2.2 Characterization of EHEP

In the natural EHEP crystal structure, the first 20 amino acid cannot be building because no electron density map was obtained (Fig 21a). To address whether it was

result from this part of EHEP structure flexible or other reasons, we performed N-terminal amino acid sequencing of the natural EHEP by Edman degradation. However, the result showed that the N-terminal was blocked, suggesting that N-terminal has been modified, which is consistent with previous study(Tsuji et al., 2017). Then, molecular weight analysis was carried out by using MALDI-TOF MASS. The result showed that natural EHEP has a molecular weight of 22.5 kDa (Fig 20), which is consistent with the estimated molecular weight of EHEP without first 20 amino acids. This result suggest that the first 20 amino acids is cleaved off during maturation process as signal peptide. The cleft site (between 20th and 21st residue) is slightly different from the predication (between 17th and 18th residue) on website signal peptide service of SignalP-5.0 (http://www.cbs.dtu.dk/services/SignalP/). What exact the modification of its N-terminal residue Ala21 have? According to the electron density map, the acetylation modification on the N-terminus of 21st residue was matched well (Fig 21b), which is good agreement with the results of N-terminal sequencing and MALDI-TOF MASS.

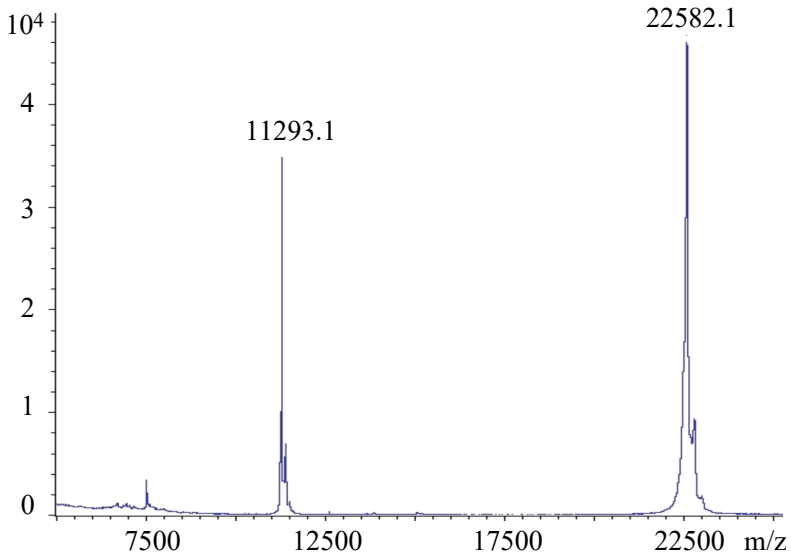
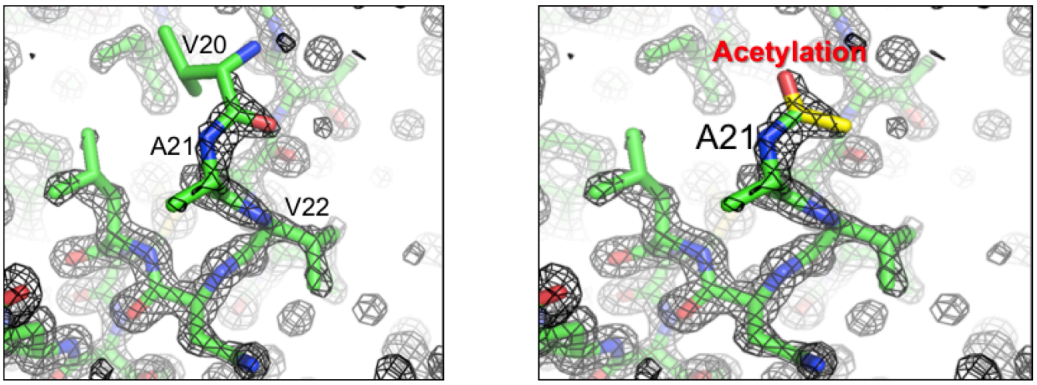



Fig 20. MALDI-TOF MASS result of EHEP.




(a) (b)

Fig 21. Structure of the EHEP. The backbone is shown in a stick representation, with a $2F_o$-$F_c$ electron density. (a) Electron density map pf EHEP, (b) acetylation modification of Ala21 residue.

The acetylation modification of N-terminal is a frequent phenomenon for majority of eukaryotic proteins. The modification might have important biological functions, such as protein half-life, protein secretion, protein-protein interaction, protein-lipid interaction(Silva & Martinho, 2015), metabolism and apoptosis(Hollebeke *et al.*, 2012). Besides, the N-terminal acetylation is expected to play an important role in determining protein stability (Lange & Overall, 2011). It need further verify that the exact function

of EHEP acetylation modification in sea hare *A.kurodai*.

### 3.2.3 The effect of acetylation modification on the activity of EHEP

To explore the effect of modification on the EHEP activity, we prepared recombinant EHEP using *E.Coli* expression system since *E.Coli* expression system lacks the capacity to modify its expressed proteins(Rosano & Ceccarelli, 2014). Because the first 20 amino acids are cleaved off during maturation process, the nucleotides of full-length EHEP except the nucleotides coding the first 20 amino acids were ligated to expression vector to express EHEP. Two vectors were tried to express EHEP.

For the EHEP (-20)-28M vector, the presence of $His_6$-Tag at the N-terminal greatly facilitates the purification procedure. EHEP (-20)-28M was expressed in small amount with majority of the protein insoluble in pellet. The chromatogram from Ni affinity column is shown in Fig 22a. EHEP was purified and eluted with a 4% buffer B first, and then eluted with a linear gradient of 4%-100% buffer B over 10 CV. The two peaks were checked by SDS-PAGE. The fractions containing the protein was purified by Q column. Several peaks appear and checked by SDS-PAGE. SDS-PAGE shows that the yellow peak contains the desired protein. Finally, EHEP was purified by SEC column( Fig 22c). It contains two peaks at 280nm. Fraction corresponding to these peaks were collected and analyzed with SDS-PAGE. According to the SDS-PAGE, the peak 2 contains the desired protein. After three step purification process, His-EHEP (-

20) was obtained with extremely low yield.

Because majority of EHEP was in the precipitate, we fused EHEP with the MBP tag to increase its solubility. The MBP-EHEP was expressed in large amount with majority remain in the supernatant. MBP-EHEP was purified by MBP column (Fig 23a), and the chromatogram is shown. MBP-EHEP is eluted by 100% elution buffer. The elution peak was tested by SDS-PAGE, showing that all the fractions is presented MBP-EHEP. Then, MBP-EHEP is treated with TEV protease and was purified by MBP column second time (Fig 23b). Finally, EHEP was purified by SEC column ( Fig 23c). The chromatogram from gel filtration of fractions contains three peaks at 280nm. Fractions corresponding to these peaks were collected and checked with SDS-PAGE. According to SDS-PAGE, most EHEP elutes in peak 1. We can only get a little protein after three step purification process. It may result from the EHEP has many disulfide bonds as we discussed above, and the *E.Coli* cannot provide a propriate environment to fold.

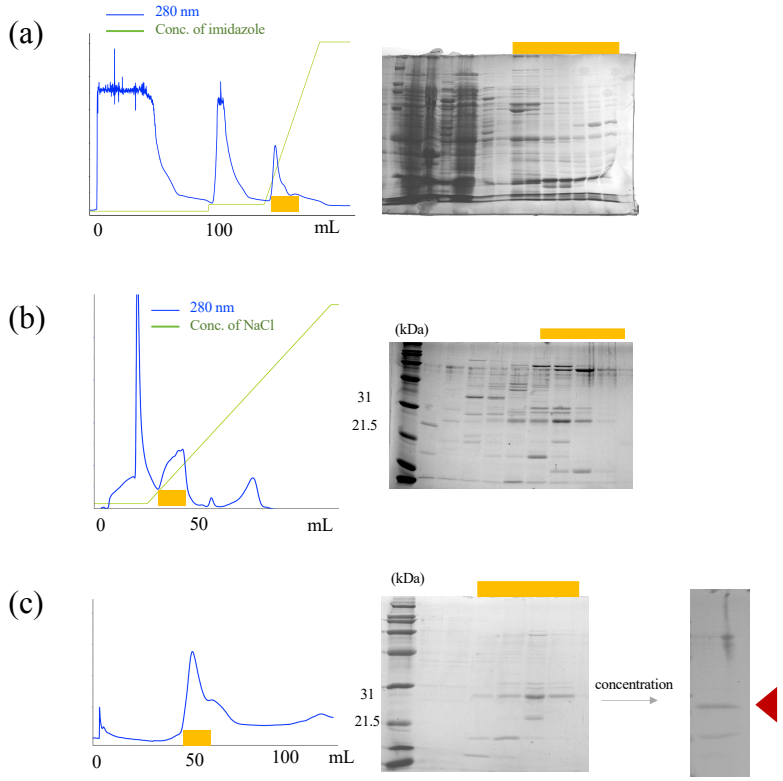



Fig 22. Purification and related SDS-PAGE results of EHEP-20-28M. (a) Purified by Ni column.

(b) Purified by Q column. (c) Purified by SEC column.

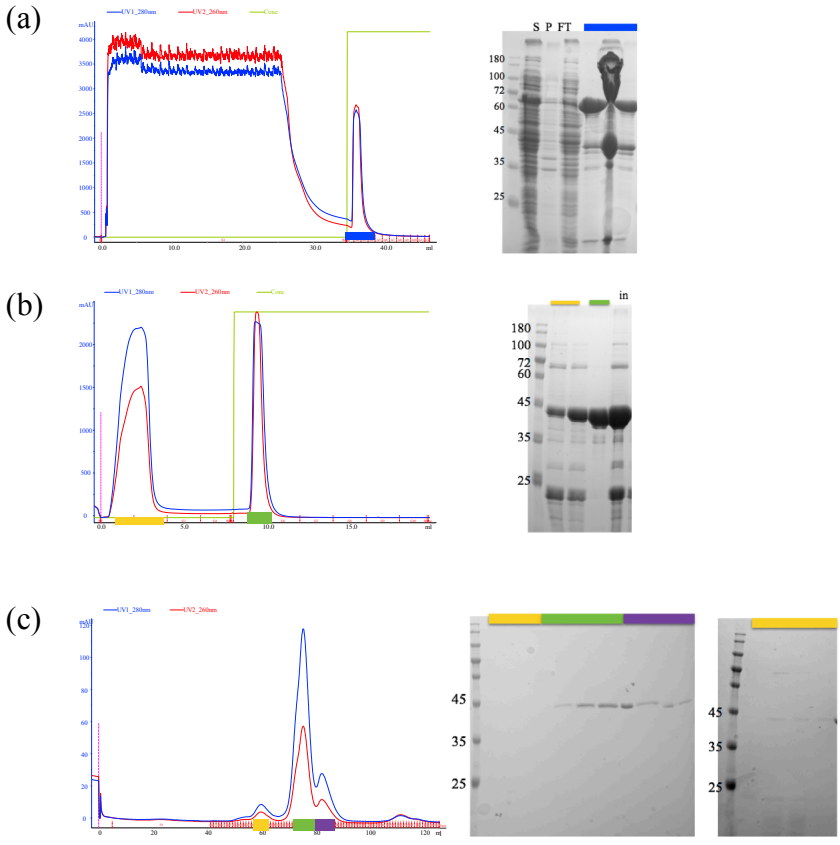



Fig 23. Purification and related SDS-PAGE of EHEP-20-MBP. (a) Purified by MBP column. (b) The sample after TEV treatment was purified by MBP column. (c) Purified by SEC column.

Recombinant His-EHEP (-20) was used for checking its activity using tannic acid as the substrate. The result showed that the recombinant EHEP was precipitated after incubating with tannic acid (Fig 24), indicating that the acetylation modification is not indispensable for the EHEP activity. During the EHEP purification process, we can see some impurity band which might be the degradation part of EHEP, indicating that the acetylation modification of EHEP might has an effect on the EHEP stability.

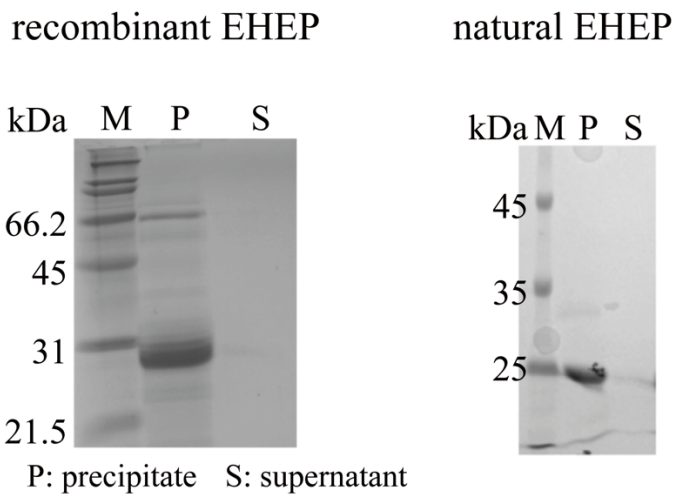



Fig 24. Activity of recombinant EHEP. Lane M: molecular-weight protein markers. Lane P: Precipitate resolving in SDS-PAGE loading buffer.

### 3.2.4 Complex structure of EHEP with TNA

In sea hare, EHEP binds to phlorotannin, which can in turn protect *aku*BGL from phlorotannin-inhibition(Tsuji et al., 2017). In vitro binding assay, it was clear that tannic acid was also bound to EHEP as a homology of phlorotannin. In order to gain detailed insights into the binding manner, we tied to obtain a structure of EHEP complexed with tannic acid (EHEP-TNA) by soaking method. Because the EHEP crystal is difficult to repeat, we used the material developed by our lab to get the crystal and then for soaking at high temperature with long time (Fig 25).

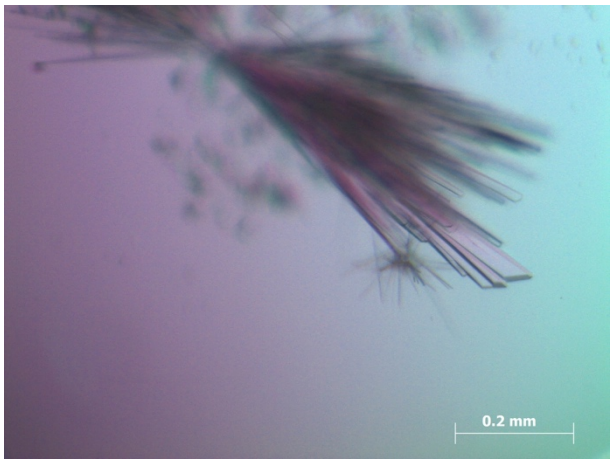



Fig 25. EHEP crystal grown on the material and for soaking tannic acid.

The structure of EHEP complexed with 1,2,3,4,6-pentagalloylglucose (EHEP-PGG) was determined by MR method using native EHEP as a search model at 1.9 Å resolution (Table 3, Fig 26). After rigid-body refinement of MR calculation, the both 2Fo-Fc and Fo-Fc maps (Fig 27B) showed a well-defined electron density blob for a pentagalloylglucose, not tannic acid. The pentagalloylglucose is a major part of tannic acid which absents the outmost five gallic acids of tannic acid (Fig 27A). We do not know whether the outmost five gallic acids of tannic acid were disordered due to its structure flexible or hydrolyzed during soaking. According to the previous studies, acid-catalytic hydrolysis of tannic acid was achieved at high temperature of 130 ℃ (Jie Fu *et al.*, 2015). Even with a polystyrene-hollow sphere catalyst, the temperature of 80 ℃ was needed (Luo et al., 2018). Therefore, we suggest that the outmost five gallic acids cannot be seen due to its structure flexible. EHEP bound in a 1:1 stoichiometric ratio with pentagalloylglucose. There are not many conformation changes between the apo form and the pentagalloylglucose complex structures with a RMSD of 0.283 Å.

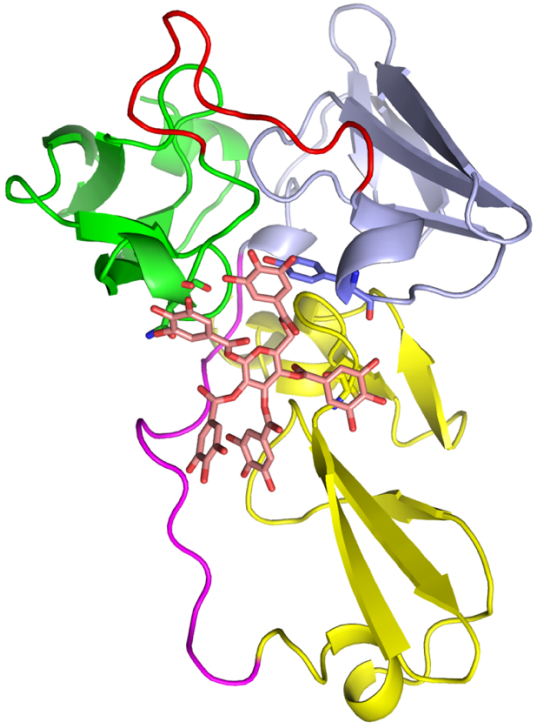
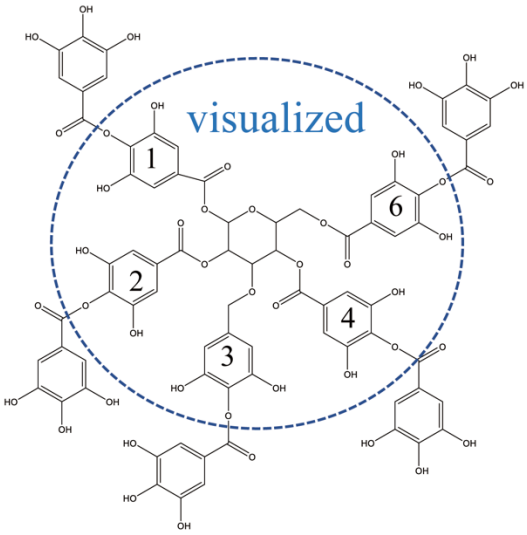
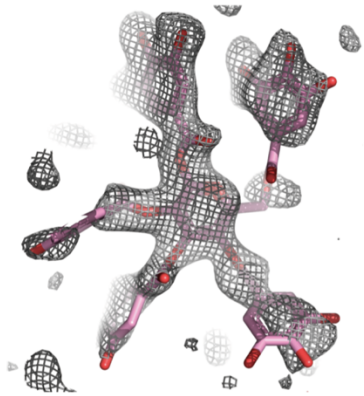


Fig 26 Complex structure of TNA-EHEP. Three domains and two linkers are colored by green, light-blue, pale-yellow, magenta, and red respectively. TNA is colored in pink, shown as stick representation.




(A)



Fo-Fc map of tannic acid, 2σ level

(B)

Fig 27 (A) Chemical structure of tannic acid. (B)Fo-Fc map of tannic acid. The map is colored in grey and countered at 2σ level.

Although the overall tertiary structure of EHEP did not show much changes after tannic acid binding, superstition of the two structures shows some rearrangement (Fig 28). Tannic acid binding to EHEP caused an increase both in α-helix and $\beta$-sheet contents, especially for ChBD2 and ChBD3 domains. For ChBD2 domain of apo form, it consists of a $\beta$-sheet with three antiparallel $\beta$-strand. In the complex, ChBD2 domain consists of two $\beta$-sheets and two α-helix. The different is caused by the rearrangement of the N-ter region (residues 93-96) and C-ter portion (residues 129-149). ChBD3 domain of apo form consists of two $\beta$-sheet with two antiparallel $\beta$-strand and a C-ter helix. After tannic acid binding, the N-ter region (residues 168-179, 197-199) rearrangement was observed, resulting in the incorporation of N-ter α-helix and one more $\beta$-strand to the first $\beta$-sheet. For ChBD1 domain, there is only small change resulted from the ligand interaction, $\beta$-strand3 and C-ter helix become longer.

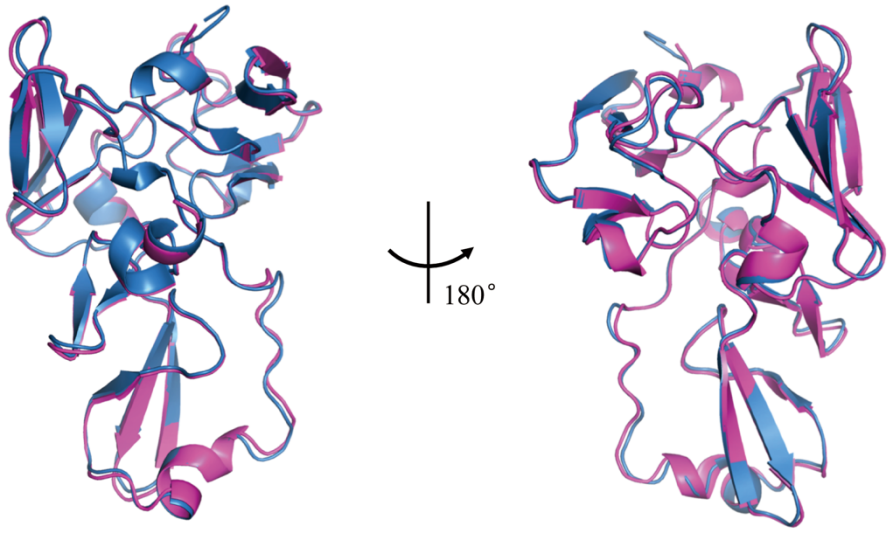



Fig 28. Alignment of EHEP form with TNA complex form. apo form is colored in pink and complex form colored in blue (TNA did not display).

The complex structure showed that tannic acid is bound to the center of the three ChBDs (Fig 26) without any symmetric binding manner. The binding pocket on the surface of EHEP, was formed by C-terminal α-helix of ChBD1, N-terminal α-helix of ChBD2, and middle part (loop) of ChBD3 with the help of two linker loop. The tannic acid bound to EHEP mainly by hydrogen bond (Fig 29) and hydrophobic interactions (Fig 30). Three of the gallic acids, gallic acid1, gallic acid 4 and gallic acid 6 interacted to EHEP by hydrogen bonds and hydrophobic interaction. Gallic acid 3 also hydrophobically contacted with EHEP, and form hydrogen bond with the water molecules. Additionally, gallic acid 1 and gallic acid 4 also share hydrogen bond with water molecule. Hydroxyls group of gallic acid1 forms hydrogen bind with N75 of ChBD1 with a hydrogen bond length of 2.7 Å.

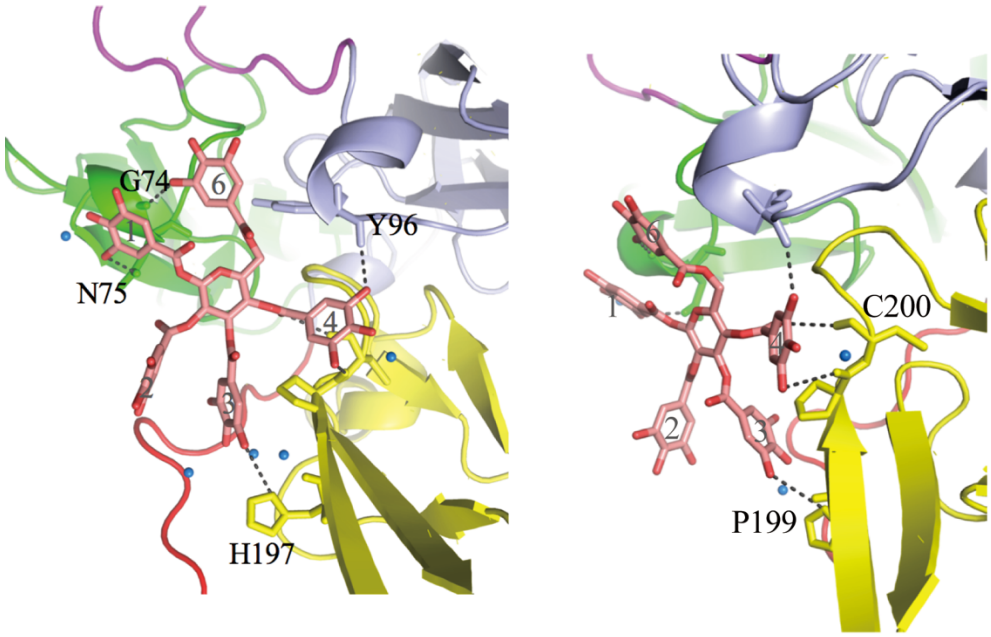



Fig 29. The domains of EHEP are colored using the same color scheme as in Fig. 21. Hydrogen-bonding interactions between the tannic acid and the protein are shown as dashed lines. Water

molecules within hydrogen bonding distance of the tannic acid are indicated by small cyan spheres, but the hydrogen bonds are shown in grey dashed lines.

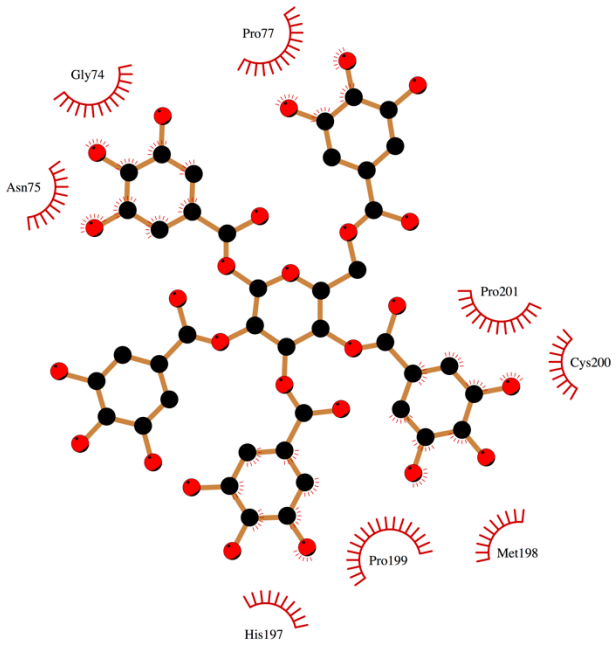



Fig 30. Hydrophobic interaction diagram of tannic acid with EHEP.

The gallic acid4 hydrophobically interacted with M188, P199, C200, and P201 of ChBD2. Gallic acid 6 formed hydrogen bond with main side of G74 and hydrophobic contacted also to P77 of ChBD1. Of particular interest was the CH-**π** interaction. Additionally, there are also some hydrogen bonds existing between tannic acid with water molecule (Fig 29). Comparing EHEP-TAN with apo form, we can clearly see that most of these residues are located in the motif newly formed upon the ligand binding.

The interaction of protein with phenols is widely studied on proline-rich proteins and histatins. Phenol binding to protein has may sides of effect on proteins, for instance, the protein structure, protein functional properties and nutritional value (Ozdal *et al.*, 2013). Understanding the interactions mechanism of tannin-protein has been interests

of the ecological and botanical communities and the human healthy and food science communities. However, due to lack of crystals, the interaction cannot be understood at the molecular level. Because there is almost no three-dimensional structure of protein-phenol available, various indirect methods were used for detect changes in the protein secondary structures, such as circular dichroism (CD) and Fourier transform infrared spectroscopy (FTIR). Papadopoulou et al reported that the association between flavonoids and BSA did not change molecular conformation of BSA by the fluorescence quenching method (Papadopoulou *et al.*, 2005). However, Wu et al found that there is a slight alteration on the secondary structure of $\beta$-lactoglobulin upon binding of EGCG, α-helical content of the protein increased from 21% to 25% (Wu *et al.*, 2013). Similar results were obtained by Roy et al (Roy *et al.*, 2012), studying the interaction of two stereoisomeric antioxidant flavonoids, catechin (C) and epicatechin (EC) with bovine serum albumin (BSA) and human serum albumin (HSA), showing that C and EC stabilize the α-helix at the cost of a corresponding loss in the $\beta$-sheet structure. Kanakis investigated the interaction of milk $\beta$-lactogolobulin with tea polyphenols and showed that there is an increase in $\beta$-sheet and α-helix after binding (Kanakis *et al.*, 2011).

In this research, we utilize the crystallography to evaluate the interaction of phenol with TNA. Previous researches studies about the structure alterations mainly by indirect method. Here, the complex crystal structure shows that tannic acid binding induces an increase of $\beta$-sheet and α-helix content, similar with milk $\beta$-lactogolobulin with tea polyphenols.

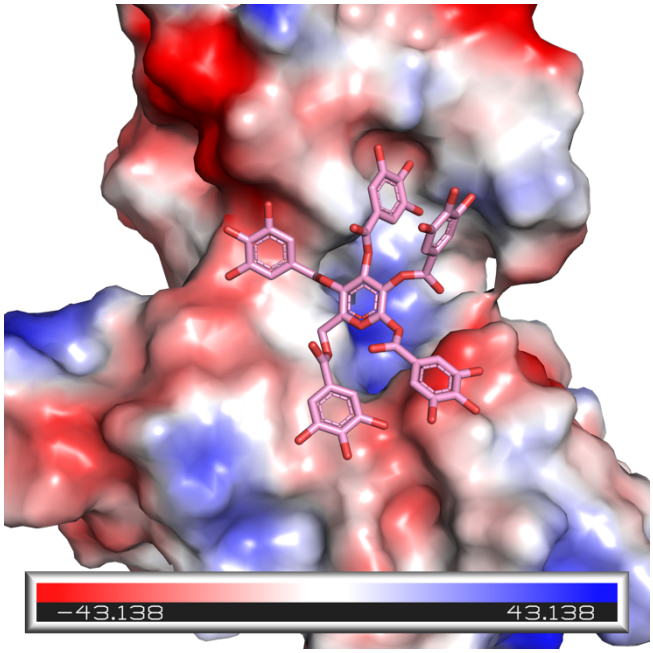



Fig 31. Representation of the electrostatic potential at the binding pocket of tannic acid. Color-codes depends on the electrostatic potential (red: negative charge; blue: positive charge; and white: neutral charge). tannic acid is colored in pink representation by sticks.

To get more insights into the tannic acid binding mode of EHEP, we have also investigated its electrostatic surface properties of the binding pocket. Tannic acid in bound at an almost neutral charge pocket (Fig 31). This issue will be further leading us to resolve the precipitate of EHEP with tannic acid.

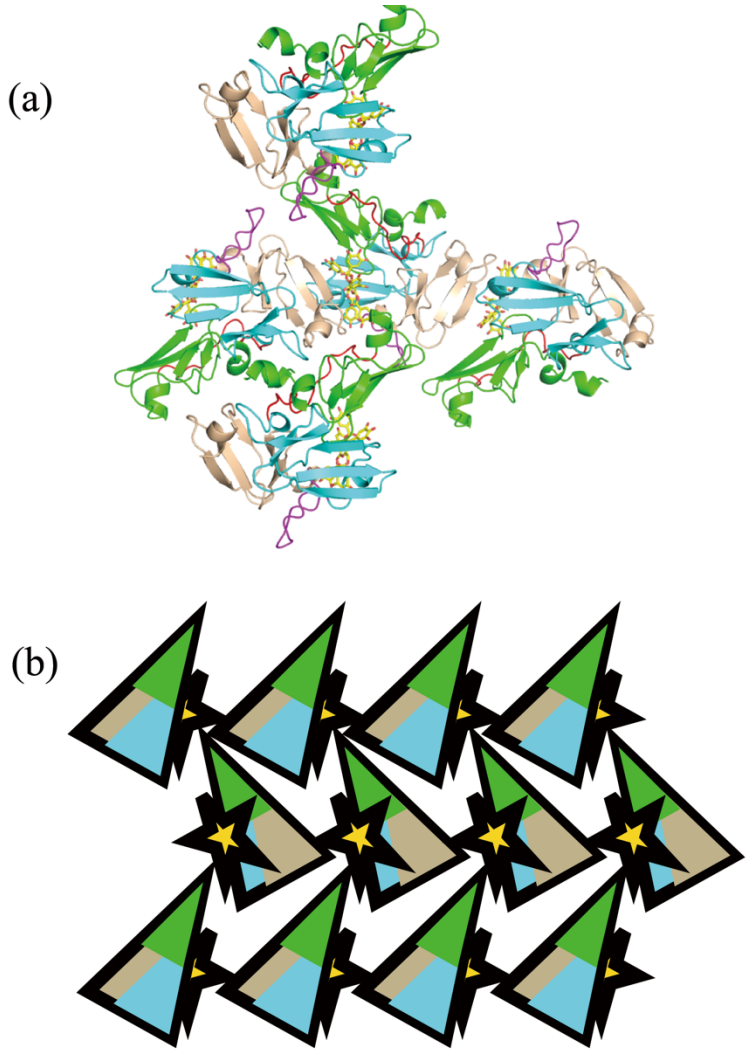



Fig 32. Crystal packing analysis of EHEP-TNA complex. (a) EHEP-TNA molecule in the crystal. (b) Diagram of EHEP-TNA molecule packing. One triangle represents a EHEP-TNA molecule, different color (blue, green, orange) represents different ChBD domains. Yellow pentagram represents the TNA.

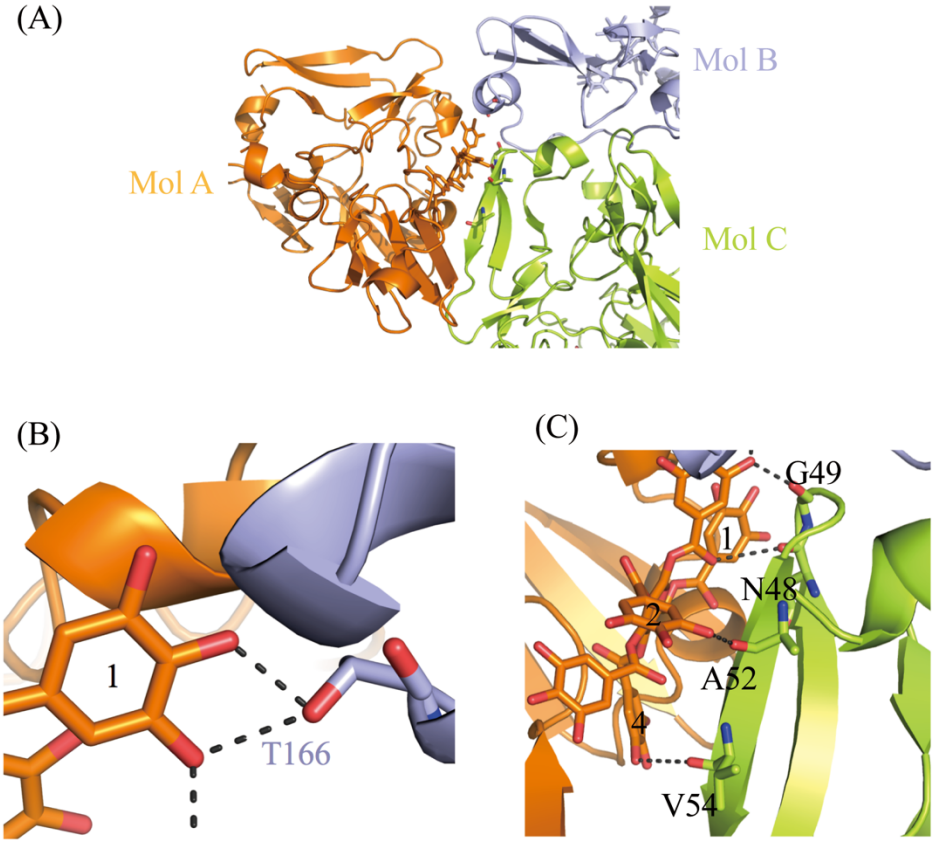



Fig 33. Interaction analysis between different molecules. (A) Overall view of the three molecules interactions. Three different molecules are colored in orange, green, light blue, respectively. (B) Interactions between TNA of molecule A with molecule B. The residue participate in interaction is shown as sticks. (C) Interactions between TNA of molecule A with molecule C. The residue participate in interaction is shown as sticks.

The EHEP crystal form contains a single EHEP in the asymmetric unit. Fig 32 shows the EHEP crystal packing. All the molecules in this crystal are related by crystallographic symmetry operations and located in equivalent packing environments. In addition to the intermolecular interaction formed by crystallographic symmetry, the tannic acid also associates with the other two adjacent EHEP molecules.

Fig 33 shows the detail information about how the tannic acid interacts with two

other molecules. The gallic acid 1 of tannic acid from molecule A forms hydrogen bonds with T166 on the loop connecting of ChBD2 and CHBD3 domain of molecule B and N48, G49 of molecule C. Gallic acid 2 forms hydrogen bond with A52 of molecule C. Gallic acid 4 forms hydrogen bond with V54 of molecule C. The $\beta$-strand3 of ChBD1 domain provide several hydrogen bonds to stabilize the tannic acid. Because tannic acid can associate with three EHEP molecules at the same time. EHEP might crosslink with each other at the help of tannic acid, leading to the assembly of huge complex.

### 3.2.5 ITC result

The result of ITC shown that the titration of tannic acid to EHEP produced a positive heat deflection with two discrete stages (Fig 34). During the first stage (1-3 injections), the exothermic increase for each successive injection, implying that tannic acid already titrated into the cell and bound to EHEP, and have a favorable effect on the new binding sites(Poncet-Legrand *et al.*, 2007), (Frazier *et al.*, 2003). After entering the second stage, the exothermic decrease for each successive injection. BSA and gelatin are usually used for investigation of binding affinities of tannins. BSA is a globular protein that is low in proline. Gelatin is a proline-rich protein and has an open random coil conformation (Deaville *et al.*, 2007). Previous studies suggest that protein structure influences the interaction mechanism, because BSA has a single-stage interaction mechanism and gelatin shares a two- discrete stage (Frazier *et al.*, 2003). The binding mechanism

between EHEP and tannic acid is similar with the gelatin.

The model we showed here was obtained by omitting the first three peaks, due to the first part of titration showed some cooperative effect. The value of ΔH may be not accurate, because of excluding the first three peaks. The process of EHEP binding to tannic acid is a spontaneous process, because of with a negative Gibb`s free energy. The equilibrium constant $K_d$ of the tannin acid and EHEP system is 0.1 mM, indicating the interactions is weak, and is comparable with the interaction between salivary proline-rich proteins IB5 and tannin(Canon *et al.*, 2013). The positive entropy change and negative enthalpy change suggest that the association is driven by enthalpy and entropy synergistically(Li & Ni, 2016) and that hydrophobic interactions were favored as main binding mechanism(Watrelot *et al.*, 2015).

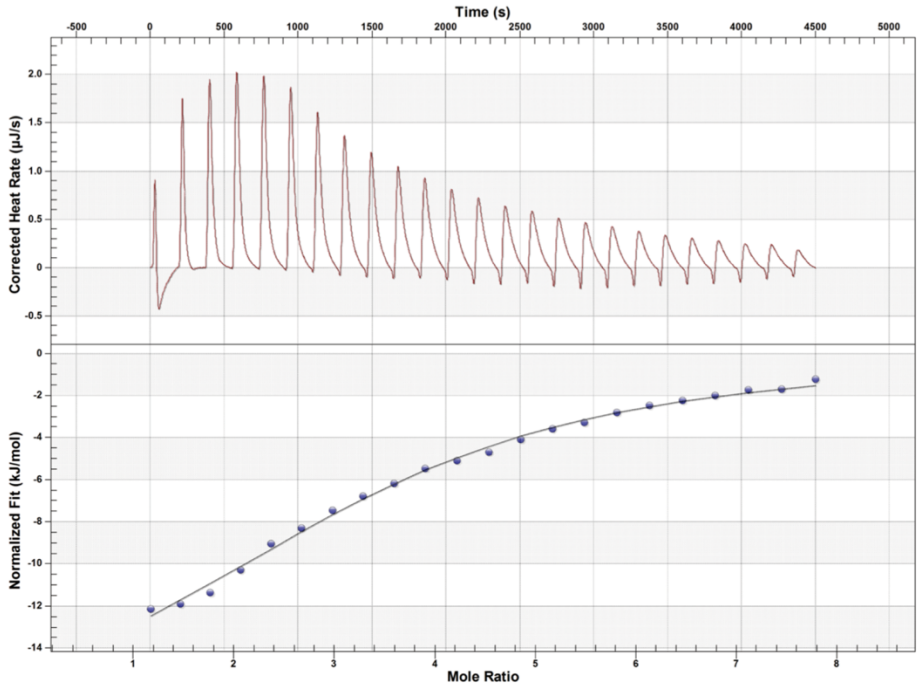



Fig 34 ITC result of tannic acid into EHEP. The upper panel shows the raw heat changes and the lower panel the integrated heats corrected for heat of dilution.

### 3.2.6 Molecular docking result of EHEP with phlorotannins

Because EHEP binds with phlorotannins in nature, it is difficult to isolate one compound. Therefore, molecular docking is performed to characterize the binding manner of EHEP with phlorotannins. Three compounds of phlorotannins: eckol, phloroglucinol and phlorofucofuroeckol A were used for the molecular docking studies. All these three compounds were docked into the binding surface of EHEP (Fig 35).

The glide score of EHEP with eckol is -7.043 kcal/mol. Fig 35 A shows the binding manner of EHEP with eckol. All the three benzene rings form hydrogen bonds with EHEP. The hydroxyl groups form hydrogen bonds with Aan64, Asn75, Val 58, Cys200, Ala 202, Cys204.Additionally, the benzene rings form **π-π** stacking with the residues Tyr819, Trp857, Trp935, Phe689, Phe677, Trp681.

The glide score of EHEP with phloroglucinol is -6.004 kcal/mol, which is the lowest. The binding manner of EHEP with phloroglucinol is shown in Fig 35 B. Two hydrolysis groups form three hydrogen bonds with residues Ser205, Gly203, Cys200. Also, the complex is stabilized by hydrophobic interactions with Val 678, Ile736, His 818, Tyr 819, Trp 857.

The glide score of EHEP with phlorofucofuroeckol A is lowest, with –7.768 kcal/mol, indicating that it has the highest binding affinity. Fig 35 C shows the binding manner of EHEP with phlorofucofuroeckol A. Similarly, this complex is stabilized by hydrogen bond and hydrophobic interactions. The residues that form hydrogen bond with benzene

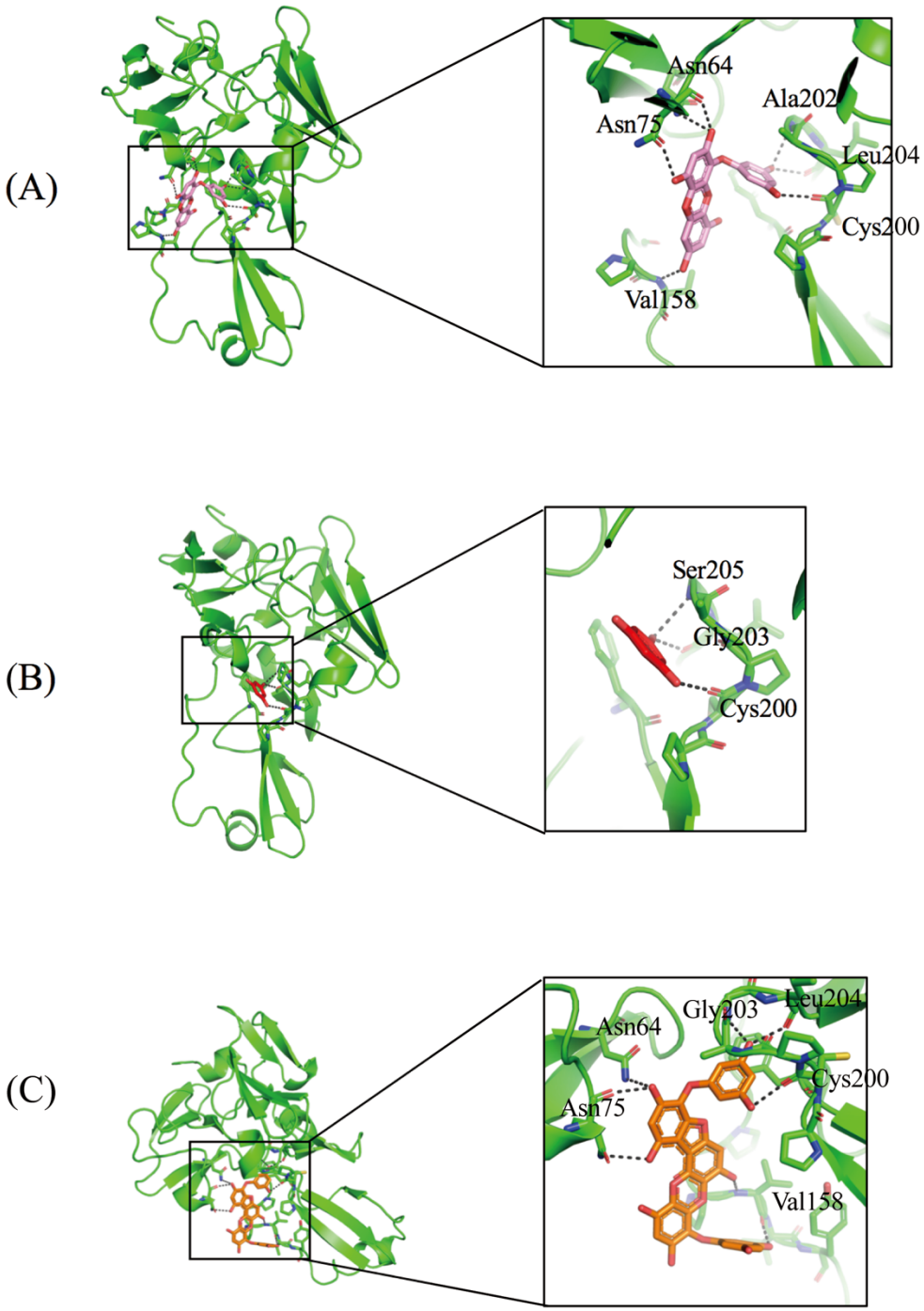



Fig 35 Docking result of EHEP with different phlorotannins. EHEP is colored in green, represented in cartoon. Hydrogen bond is shown in gray dash. (A) Dock of eckol into EHEP. Eckol is colored in pink. (B) Dock of phloroglucinol into EHEP. Phloroglucinol is colored in red. (C) Dock of phlorofucofuroeckol A into EHEP. Phlorofucofuroeckol A is colored in orange.

rings are same with eckol complex. However, the residues that take part in hydrophobic interaction is different, including Val678, Leu 682, Ile 686, Ile 688, Phe 689, Ile 842, Ala 855, Trp 857, Phe 943. It seems that with the number of hydroxyl increase, the higher binding energy.

### 3.2.7 Dissolution of the precipitate of EHEP-TNA

The binding mechanism of phenol and protein has long been investigated. The mechanism involves covalent or/and non-covalent interactions, resulting in reversible or irreversible binding. Non-covalent interactions mainly include hydrogen bond, hydrophobic interaction and van der Waals.

The complex structure of EHEP-TAN showed that the tyrosine and proline are most important residues for binding assay and mainly by hydrogen bond and hydrophobic interactions, indicating that the precipitate might be reversible binding. EHEP precipitated after binding with phlorotannin, which limit its industry application. How we could dissolve the precipitate? if we can use buffer containing high concentration of salt or using buffer containing SDS? A low pH is benefit for the TNA binding, and alkaline pH is disadvantage for the EHEP binding with TNA, leading us to wonder if we could dissolve the precipitate in alkaline pH.

We tried to dissolve the precipitate in the buffer pH 8.0. After several hours, we can clearly see that the precipitate becomes clear. The character of solubilized EHEP was checked by SDS-PAGE and HPLC. The result showed that the EHEP is in the

supernatant, furthermore, the retention volume of solubilized EHEP is the same as native EHEP (Fig 36), indicating that the solubilized EHEP has the same character with native EHEP.

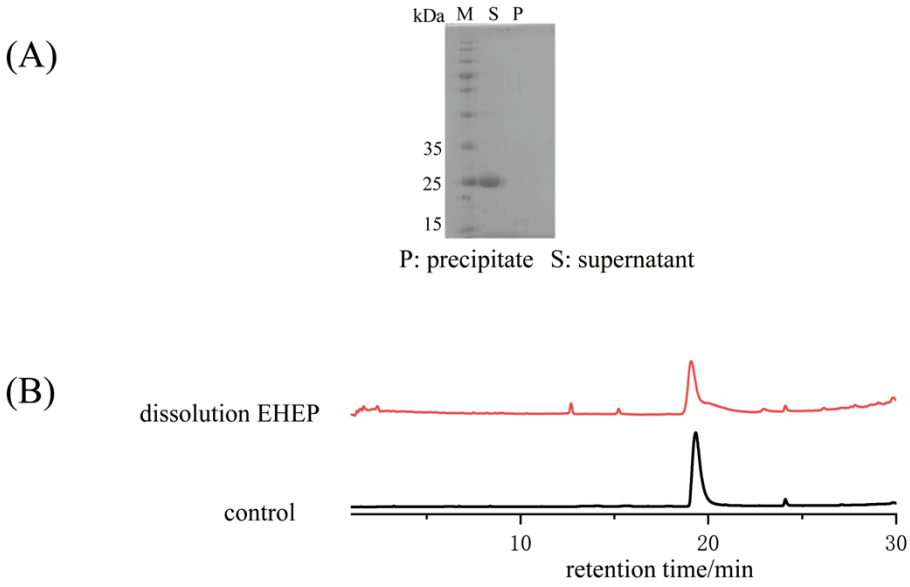



Fig 36 Solubilization of the precipitate of EHEP-TNA. (A) Checked by SDS-PAGE. (B) HPLC profile of solubilized of EHEP-complex. Black line: control EHEP. Red line: solubilized EHEP.

# Chapter 4 Structure of *aku*BGL

*aku*BGL is responsible for catalyzing the glucose releasing during the *A.kurodai* digestion brown algae process. The sequence analysis show that it possesses two tandem GH1 domains linked by a long linker. However, in PDB database, there is no β-glucosidases sharing two tandem GH1 domains available. The two-domain architecture might represent a specific adaption towards brown algae. Therefore, in this chapter, we tried to solve the structure of *aku*BGL.

## 4.1 Materials and methods

### 4.1.1 Sample preparation

BGL was purified from digestive fluid of *A.kurodai* and was kindly provided by the Professor Tsuji of Tokushima university.

### 4.1.2 Crystallization and data collection

The initial crystallization screening was performed using the sitting-drop vapor-diffusion method with Screen classics, Classics II crystallization kits from Qiagen (Hilden, Germany) PACT crystallization kit from Molecular Dimensions (Anatrace, Inc.) at 20°C. Protein solution (0.5 μL) was mixed with an equal volume of reservoir solution. Initial crystals were obtained under condition no. 41 [0.1 M sodium acetate

(pH 4.5) and 25% polyethylene glycol (PEG) 3350] of Classics II, condition no. 13 [0.1 M MIB buffer (25 mM sodium malonate dibasic monohydrate, 37.5 mM imidazole, 37.5 mM boric acid), 4.0 and 25% PEG 1500], no.37 [0.1 M MMT buffer (20 mM DL-malic acid, 40 mM MES monohydrate, 40 mM Tris), 4.0 and 25% PEG 1500] of PACT. After optimization (varying the buffer pH, precipitant concentration, using material Co-cage1), the best crystals obtained were harvested into loops and cryoprotected by the addition of glycerol to a 20% final concentration. Diffraction data were collected at 100 K on the BL-1A beamline at the photon Factory (Tsukuba, Japan) using a Dectris EIGER X4M detector. 360° of data were taken with an oscillation angle of 0.5° to obtain full data sets. Data were indexed using XDS (Kabsch, 2010)and scaled with aimless (Evans & Murshudov, 2013).

### 4.1.3 Structure solution and refinement

The structure of *aku*BGL was determined by molecular replacement with *Phaser* in *PHENIX* (Adams *et al.*, 2010). The structure of β-klotho (PDB entry: 5VAN) (Lee *et al.*, 2018), which shares 30% amino-acid sequence identity with *aku*BGL as the search model. Early model was built using the *AutoBuild* wizard in *PHENIX* (Adams *et al.*, 2010). Cycles of refinement and model building were performed using *PHENIX* and *Coot* (Emsley & Cowtan, 2004*)*. Data collection and processing statics are summarized in Table 4. Despite being added to the solution, TNA was not observed in the crystal structure. All structural figures were prepared using the PyMOL molecular-graphics

system.

### 4.1.4 Molecular docking study

Prior to docking, the crystal structure of BGL was prepared. The process of docking is similar with descriptions in 1.6 of chapter 3.

Table 4. Data collection and refinement statistics

| | *aku*BGL |
|---|---|
| Data Collection | |
| Beamline | PF BL1A |
| Wavelength (Å) | 1.0 |
| Resolution range (Å) | 50-2.7 (2.86-2.7) [a] |
| Space group | P62 |
| Unit-cell parameters (Å) a, b, c | 191.7, 191,7, 112.6 |
| Completeness (%) | 99.9 (99.4) |
| Redundancy | 10.74 (10.98) |
| Average I/σ(I) | 10.69 (2.76) |
| $R_{merge}$ [b] | 18.4% (73.8%) |
| $CC^{1/2}$ | 99.5 (86.4) |
| Molecules/ asymmetric unit | 2 |
| Refinement | |
| $R_{work}/R_{free}$ (%)[c] | 0.20/0.27 |
| Atoms | 14596 |
| Water molecules | 593 |
| Ligands | 0 |
| RMSD from ideality | |
| Bond length (Å) | 0.010 |
| Torsion angle (°) | 1.145 |
| Ramachandran plot (%) | |
| Favored | 92.39 |
| Allowed | 6.44 |
| Outliers | 1.18 |

[a]Values in parentheses are for the highest resolution shell.

[b]$R_{merge} = \sum_{hkl}\sum_i|I_i(hkl)-<I_i(hkl)>|/\sum_{hkl}\sum_i I_i(hkl)$, where $i$ is the number of observations of a given reflection and $I(hkl)$ is the average intensity of the $i$ observations. $R_{free}$ was calculated with a 5% fraction of randomly selected reflections evaluated from refinement. The highest resolution shell is shown in parentheses.

[c]$R_{work} = \sum_{hkl}||F_{obs}|-|F_{calc}||/\sum_{hkl}|F_{obs}|$, $R_{free}$ was calculated for 5% randomly selected test sets that were not used in the refinement.

## 4.1 Results & discussion

### 4.2.1 Crystallization of *aku*BGL

The quality of *aku*BGL was checked by SDS-PAGE (Fig 37). Then *aku*BGL was exchanged to buffer 20 mM Bis-tris (pH 6.0) using Vivapsin-4 columns with a molecular-weight of 50 kDa.

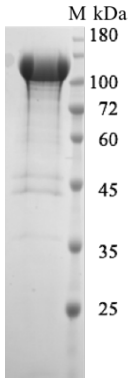



Fig 37. *aku*BGL was checked by SDS-PAGE. Lane M: molecular-weight protein markers.

After *aku*BGL was concentrated to 9.6 mg/mL, initial screening was performed with the kit Classics II, Classic, and PACT kits. During initial crystallization screening, several crystals appeared under some conditions (Fig 38). However, only the crystal condition obtained under Classic II 5D can be repeated and thus optimized to improve the crystal quality.

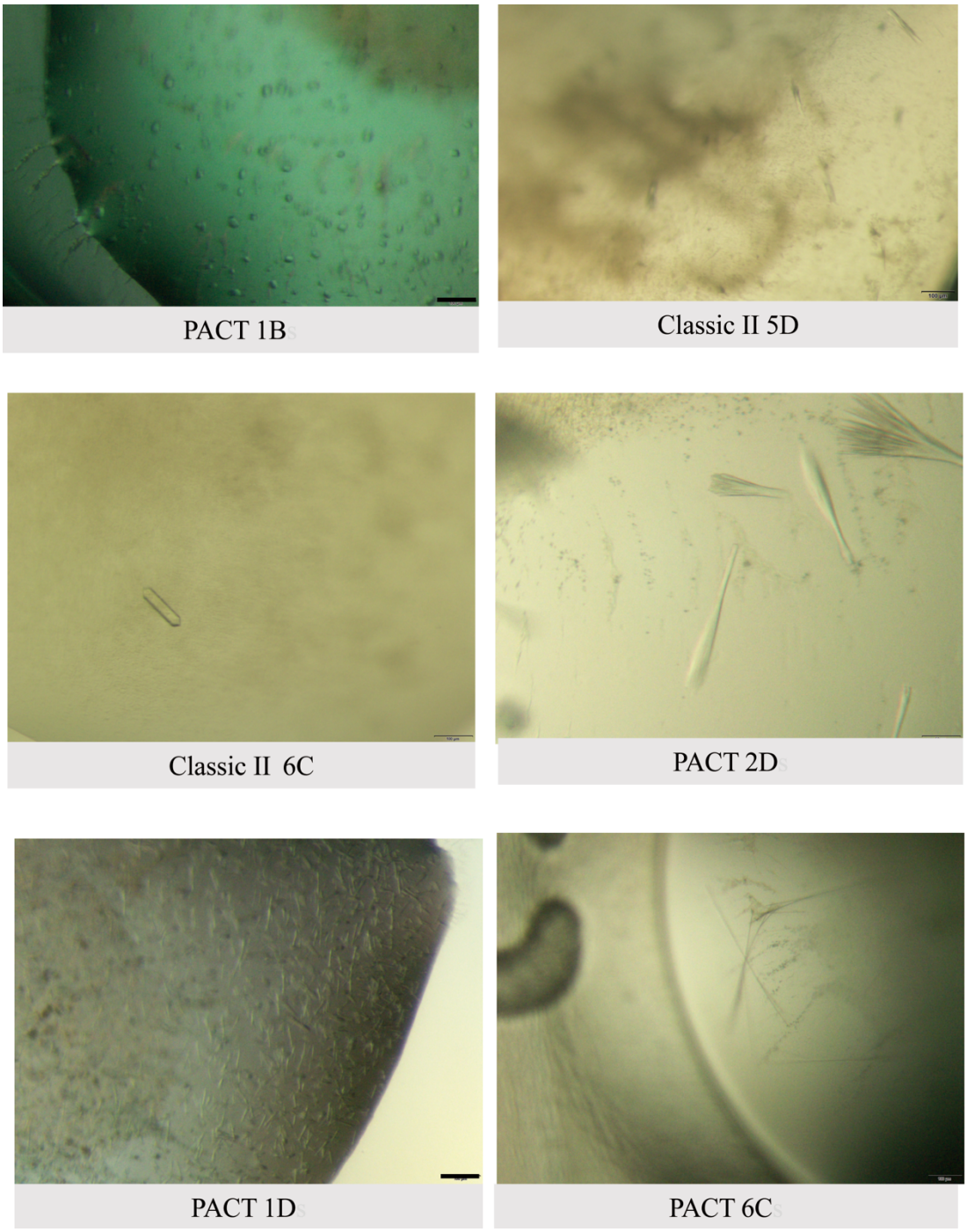



Fig 38. Photos of the crystals in initial screening plates.

To optimize the crystallization of *aku*BGL, the conditions were expanded around the growth condition outlined in section 1.1 (protein concentration, precipitant concentration, buffer pH). The best crystal was obtained in 0.1 M sodium acetate (pH 4.5) and 20% polyethylene glycol (PEG) 3350 adding the material developed by our

laboratory using hanging- drop method (Fig 39). To get the complex structure of *akuBGL*, the crystal was soaked with 5mM tannic acid at 20 ℃ for 4h. The crystal diffracted to a resolution of 2.7 Å (Fig 40).

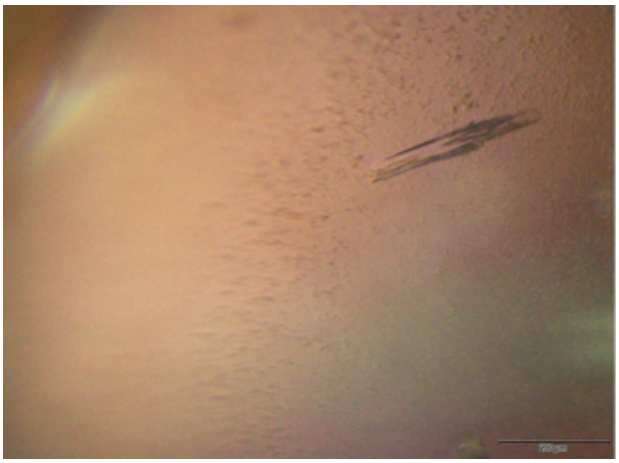


Fig 39. The best crystal for collecting data.

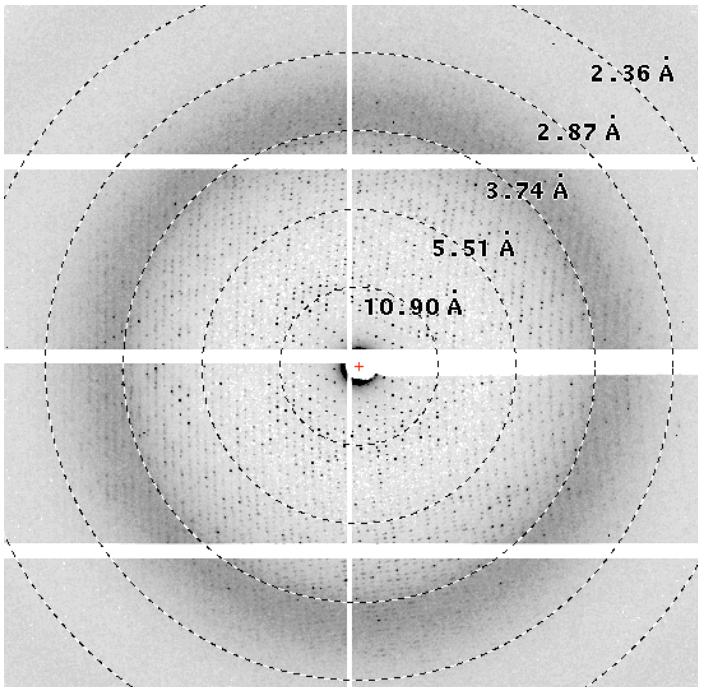



Fig 40. X-ray diffraction pattern of *aku*BGL.

### 4.2.2 Overall structure of *aku*BGL

*aku*BGL crystallized in the hexagonal space group *P*62 with two molecules per asymmetric unit. Assuming there are two molecules, A and B, in the asymmetry unit, the Matthews coefficient (Matthews, 1968) was estimated to be 2.71 $Å^3$/Da, which corresponds to a solvent content of 54.69%. Residues 34-976 of molecules A and B could be built into the electron density. The two molecules adopt the same conformation; therefore, chain A is used for discussion in the next part. The N-terminal 23 residues is signal peptide according to the SignalP-5.0 predication result (Fig 41).

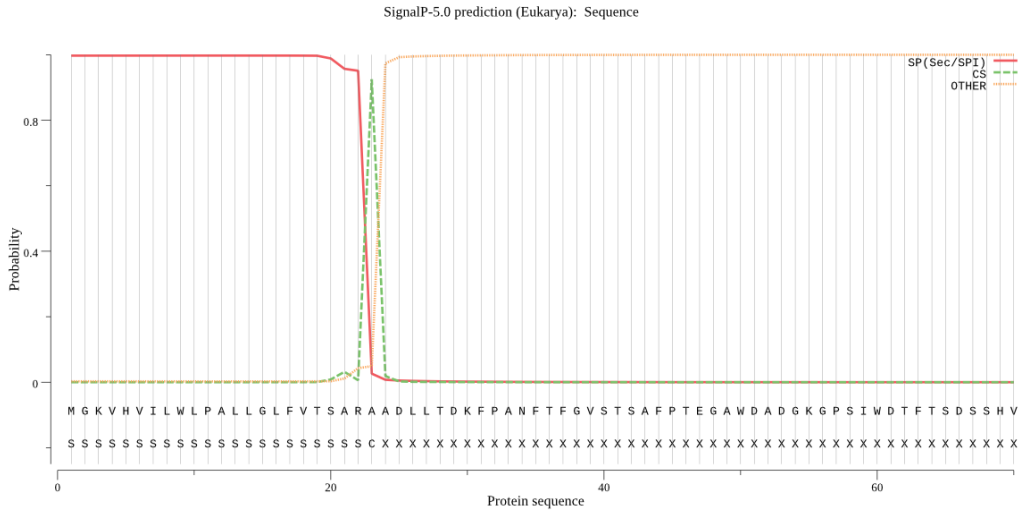
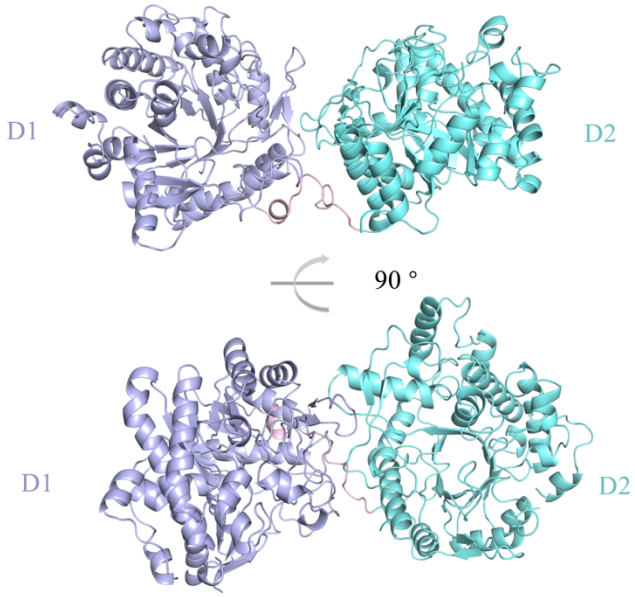


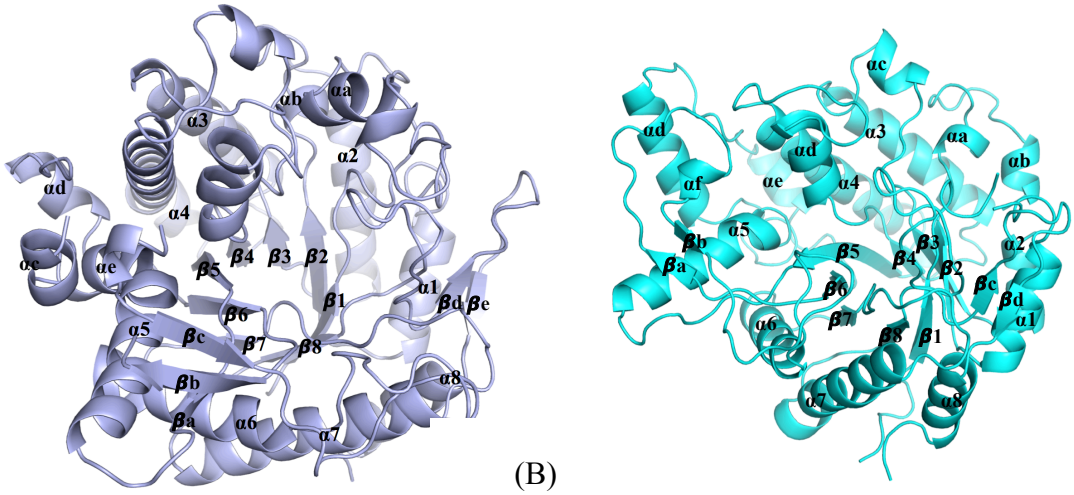

Fig 41 Signal peptide predication result of *aku*BGL.

The overall structure features two tandem glycoside hydrolase domains, D1 (residues:24-492) and D2 (residues:513-974), which are connected by a flexible linker (Fig 42a). Even though we soaked the tannic acid into crystal for several hours, there is no tannic acid in the density map. The D1 and D2 domains share the canonical $(\beta/\alpha)_8$ -barrel of the GH1 family with some extra secondary-structure elements: a three-stranded antiparallel $\beta$-sheet, a two-stranded antiparallel $\beta$-sheet, five α-helix in D1

domain; two two-stranded antiparallel $\beta$-sheets, six α-helix in D2 domain (Fig 42b).




(A)




(B)

Fig 42. (A)The overall structure of *aku*BGL. D1 domain was colored in light blue, D2 was colored in cyan, and the linker was colored in light pink. (B) The helices and strands which constitute the $(\alpha/\beta)_8$ TIM-barrel fold is labelled α1-α8 and $\beta$1- $\beta$8 respectively. The extra secondary structure is labelled αa-αf and $\beta$a-$\beta$e respectively.

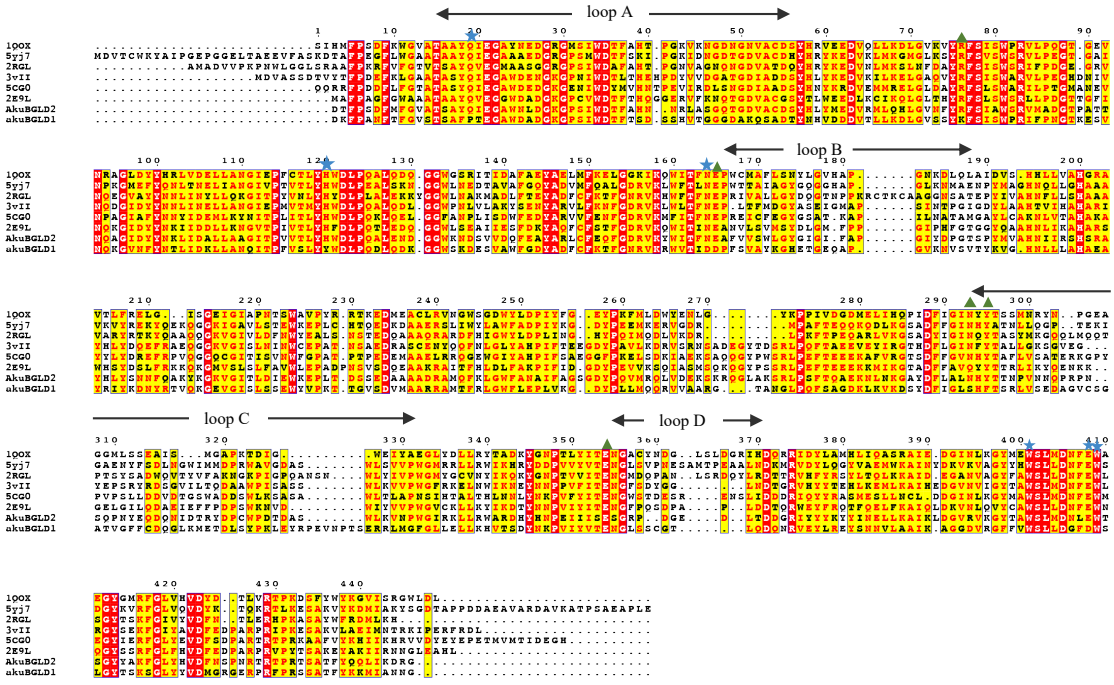
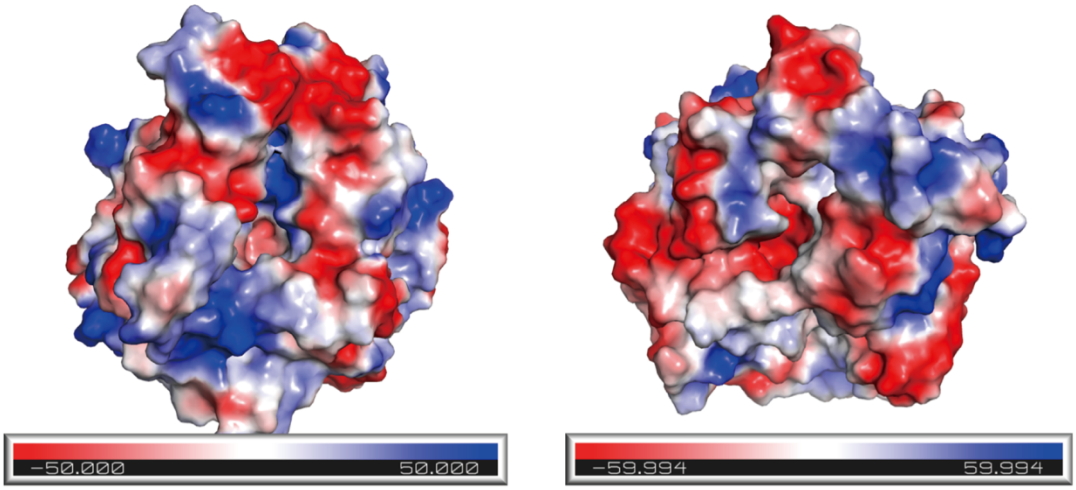



Fig 43. Sequence alignment of D1 domain, D2 domain of *aku*BGL with $\beta$-glucosidase from termite Neotermes (PDB entry: 3VII), $\beta$-glucosidase from spodopera frugiperda (PDB entry: 5CG0), $\beta$-glucosidase from Bacillus circulans sp. alkalophilus (PDB entry: 1QOX), Cytosolic Neutral beta-Glycosylceramidase from human (PDB entry: 2E9I), $\beta$-glucosidase from oleaginous microalgae Nannochloropsis(PDB entry: 5YJ7). Rice BGlu1 (PDB entry: 2RGL). Strictly conserved and homologous residues are shown in red and yellow boxes, respectively. Conserved residues of GBS site is marked with blue pentagram, conserved residues of CR site are marked with green triangle.

There are mainly four loops surrounding the active site in the GH1 family enzymes, which are different among enzymes. The four loops is shown in Fig 43, including residues T41-D81 in D1 domain, T525-D565 in D2 domain of loop A, residues P193-V195 in D1 domain, A676-S699 in D2 domain of loop B, residues H328-G381 in D1 domain, H818-P862 in D2 domain of loop C, residues N405-D415 in D1 domain, S886-D897 in D2 domain of loop D. Plant GH1 enzymes contains a disulfide bond in loop

B, which is absent in *aku*BGL. Rice BGL contains a longer loop B compared with others. D1 domain of *aku*BGL has a longer loop C compared with others. While *β*-glucosidase from Bacillus circulans sp. alkalophilus has a longer loop D. Both D1 domain and D2 domain of *aku*BGL have a negatively charged pocket (Fig 44), while D2 domain is more negatively charged compared with D1 domain. At the top of the pocket, both domains have positively charged residues along one side, which is similar with the rice BGlu1 (Chuenchor *et al.*, 2008). These characters are suitable for binding oligosaccharide chains. In addition, we can see that at the surface potential of two domains is quite different. The surface of GH1 domain is a little positively charged, while the surface of GH2 domain is negatively charged.




(a) (b)

Fig 44. Electrostatic surface of the *aku*BGL. Positively charged, negatively charged, and neutral regions colored in blue, red and white, respectively. (a) D1 domain of *aku*BGL. (b) D2 domain of *aku*BGL,

A structural similarity search was performed with the atomic coordinates of *aku*BGL using the *DALI* server. The server identified several strong matches (Z-

score>55), among the 5 structures with highest scores, four of which belong to the *β*-glucosidase, the other one is cytosolic neutral *β*-glycosylceramidase. NkBgl(Jeng *et al.*, 2012) gave the highest score, followed by Sfβgly(Tamaki *et al.*, 2016), *β*-glucosidase from Bacillus circulans sp. alkalophilus (Hakulinen *et al.*, 2000), KLrP from human (Hayashi *et al.*, 2007) and BGLN1.Notably, all the five structures contain one GH1 domain and an alignment of their sequence is shown in Fig 43. The alignment of the *aku*BGL with these structures with root mean square difference (RMSD) values of 1.0Å for 455 Cα-atoms of *Neotermes koshunesis* glucosidase( 57.3% sequence identity), 1.1Å for 450 Cα-atoms of *spodoptera frugipersa* glucosidase (44% sequence identity), 1.0 Å for 453 Cα-atoms of KLrP from human (48% sequence identity), 1.4 Å for 431 Cα-atoms of *Bacillus* glucosidase (41% sequence identity) and 1.4Å for 451 Cα-atoms of BGLN1 (44% sequence identity).

The GH1 family glucosidases mostly utilizes the retaining mechanism during which two glutamic acids responsible for catalyzing the reaction. Sequence alignment shows that in the GH11 domain, the first glutamate is replaced by Asp192, and the second glutamate Glu 404 remained. The GH2 domain contains the conserved two glutamic acids, Glu 675 and Glu885. The two pairs of catalytic residues are located at the at the carboxyl termini of *β*-strands 4 and 7. The oxygen atoms of the carboxylate groups in the side chains of Asp 192 and Glu404 is 7 Å apart from each other (Fig 45a)and the distance of between Glu 675 and Glu885 is 5 Å (Fig 45b). The distance between two catalytic residues in the GH2 domain is similar with Sf*β*gly (3.9 Å to 4.8 Å) (Tamaki

*et al.*, 2016), and also with $\beta$-gycosylceramidase (5.3 Å) (Hayashi *et al.*, 2007). For the retaining mechanism, the distance of the two catalytic is about 5 Å and 10 Å in the inverting mechanism. In *aku*BGL, the distance in GH1 domain belongs to neither of them. More importantly, the first catalytic residue was replaced to Asp, although Asp and Glu has the similar characters. For $\beta$-gycosylceramidase, mutation the first catalytic residue Glu to Asp, its activity reduced to almost half compared to wild type (Hayashi *et al.*, 2007). Mutation of first catalytic residue Glu to Asp in Sf$\beta$gly resulted in that large activity of Sf$\beta$gly decrease (Marana *et al.*, 2003). Mutation of first catalytic residue Glu 193to Asp in *NkBgl* increases the distance between amomeric carbon group of aspartates relative to that of glutamate. The trans glycosylation reaction is more favorable in the mutant (Jeng *et al.*, 2012).

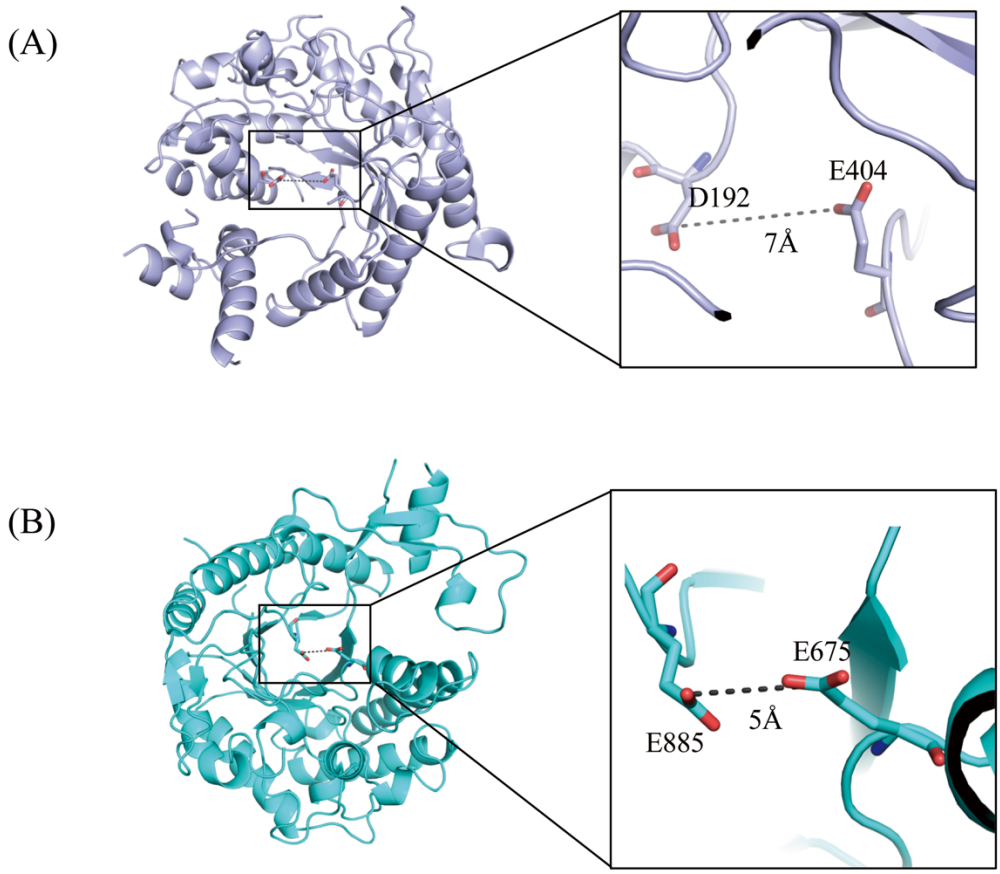



Fig 45. Conformation of two catalytic residues in two GH domains. The distance of carboxyl oxygens between the two residues is indicated. (A) GH1 domain. (B) GH2 domain.

There are some functional regions essential for the activity of $\beta$-glucosidase: glycone-binding site (GBS), aglycone-binding site (ABS) and catalysis-related residues (CR). GBS site forms hydrogen bond hydroxyl groups from the substrate saccharide and stacking interactions, locating at the bottom of the active site. ABS site is different for each enzyme, locating at the entrance of the active site. The ABS site is structurally less conserved than GBS site by the variable length of loop. The atomic positioning of CR and GBS residues is highly conserved (Tamaki *et al.*, 2016). Fig 46 shows the superposition of the two sites. We can see that most residues of the two regions share the same conformation. However, there are some residues of GH1 domain are mutated

to other residues: P45, Y147, D191, D452, K102, D192, S327, F329, responding to Q529, H630, N674, E934, R585, E675, N817, Y819 in GH2 domain, respectively. According to the analysis above, we suggest that the GH1 domain might does not possess the hydrolysis activity. Likewise, CjCEL1A has a two tandem structure of GH1 domains, which has high sequence identity(Sakamoto *et al.*, 2009). The result of sequence alignment showed that two conserved glutamic acids are conserved in the second GHF1 domain, and also the residues relating to substrate binding. However, the glutamic acid residues are not present in the first GFH1 domain and also the residues relating to substrate binding. Another similar protein, human small intestinal lactase-phlorizin hydrolase (LPH) consists of an N-terminal extracellular domain, comprising four homologues domains. The domain Ⅰ and domain Ⅱ are cleaved out during maturation process. Domain III functions as an intramolecular chaperone. Without this domain leads to a misfolded, enzymatically inactive conformation protein (Jacob *et al.*, 2002) (Amiri *et al.*, 2015). This may suggest that the GH1 domain of *aku*BGL might share a similar function, and also the first domain of cjCEL1A. This will need further more studies to elucidate it.

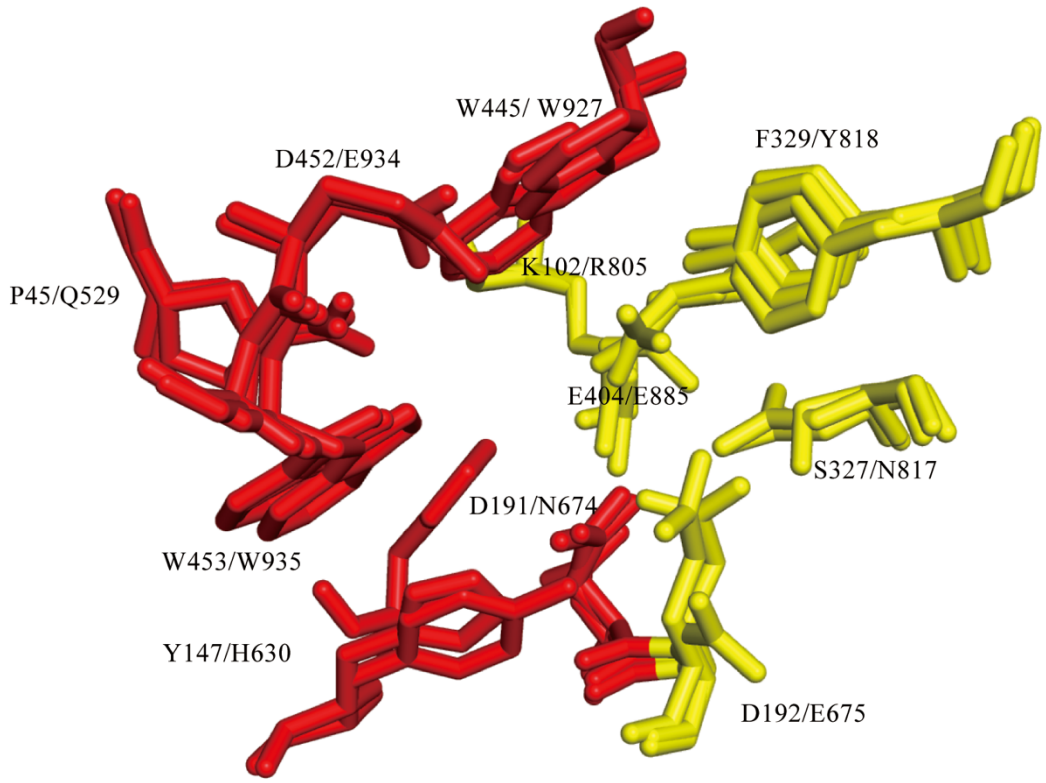



Fig 46. Superposition of active site conserved residues of *aku*BGL with Sf$\beta$gl. The number of D1 domain and D2 domain is labeled. GBS site is shown as red sticks. Catalysis-related residues (CR) are shown as yellow sticks.

### 4.2.3 Molecular docking

Because we did not get the complex structure of BGL-TNA up to now, molecular docking studies were performed to investigate the possible binding manner of BGL with TNA. As we discussed before, it seems likely that only D2 domain has the hydrolysis function, therefore, only the D2 domain was used for the docking study. The docking models were used to analyze the interactions between BGL with TNA. The docking result is shown in Fig 47a. Additionally, we superpose the structure of

rice BGlu1 complexed with cellopentaose (Chuenchor *et al.*, 2011). Comparison of the two structures, it can be visualized that tannic acid occupied the substrate binding site of BGL. This might explain the reason that tannic acid inhibits the activity of BGL.

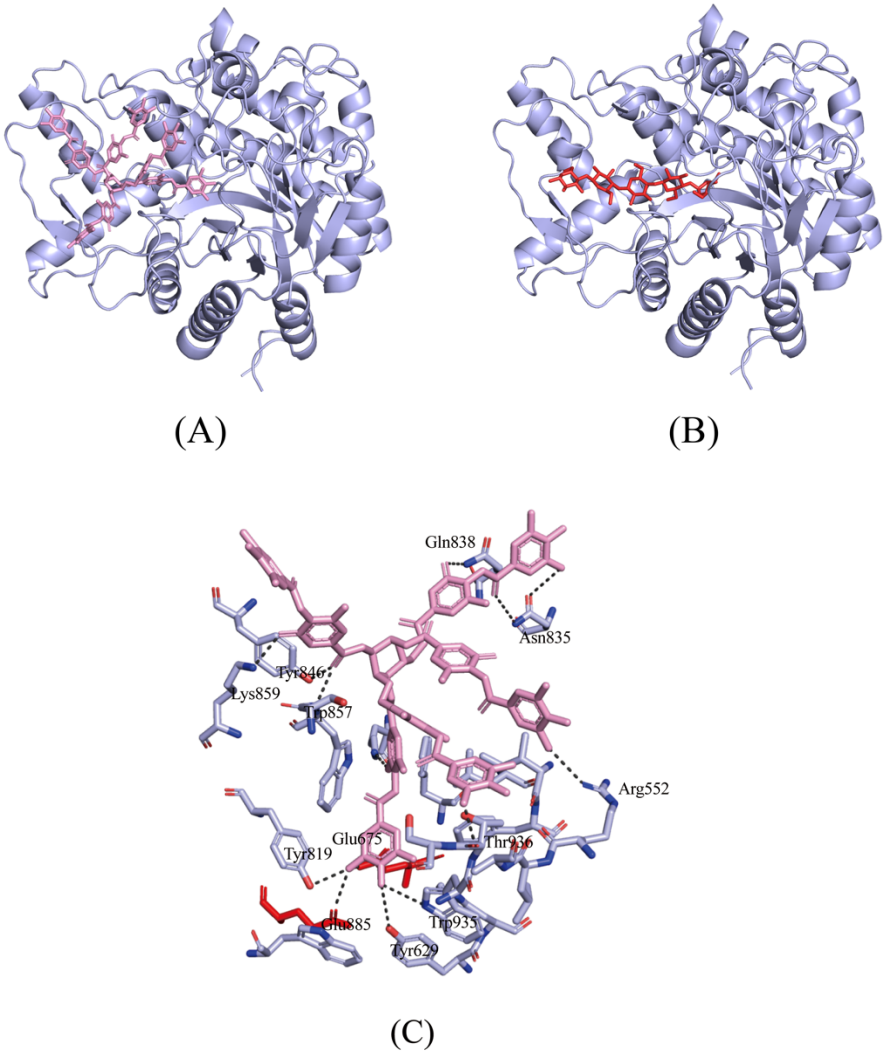



Fig 47 Structure of BGL complex. BGL is shown in light blue. (A) Docking result of BGL with tannic acid. Tannic acid is shown in pink and represented in stick. (B) Superpositions of BGL with rice BgGlu1 (PDB entry: 3F5K). Substrate is shown in red and represented in stick. (C) Interaction analysis of BGL with tannic acid. Hydrogen bond is shown in dash lines. The two catalytic residues Glu675 and Glu885 are colored in red.

The interaction analysis of tannic acid with BGL is shown in Fig 47C. Seven gallic acid rings form hydrogen bonds with BGL. It is worthy note that the two catalytic residues Glu675 and Glu 885 also involves the hydrogen bonding, which might explain why tannic acid has a highest glide score. Additionally, benzene rings form hydrophobic contact with several hydrophobic residues. Specifically, the stable **π-π** stacking is formed between residues Phe 676, Tyr629, Phe689, Tyr819, Tyr 846, Trp857, Trp927, Trp935.

Because the activity of BGL is inhibited by phlorotannins, several chemical structures of phlorotannins, eckol, phlorofucofuroeckol and phloroglucinol were used to perform the docking studies with BGL. The docked binding manner is shown in Fig 48. Fig 48A shows the binding manner of BGL with eckol. Comparison with tannic acid, the binding with eckol has a lower binding energy. Eckol binds with BGL by the hydrogen bond and hydrophobic interactions. The hydroxyl group forms hydrogen bond with one catalytic residue Glu675. In addition, the benzene rings form **π-π** stacking with several residues. Phloroglucinol has a lowest binding energy. The binding manner of phloroglucinol with BGL is shown in Fig 48B. Even though three are only three hydrogen bonds between BGL with phloroglucinol, one of which is formed with catalytic residue Glu675. Likewise, the benzene rings also form hydrophobic interactions with Val678, Ile 736, His 818, Tyr 819 and Trp857. The binding mode of BGL with Phlorofucofuroeckol A is shown in Fig 48 C. Likewise, the complex is stabilized by hydrogen bond and hydrophobic interactions. The difference is that

neither of the two catalytic residues involves in forming the hydrogen bond with BGL.

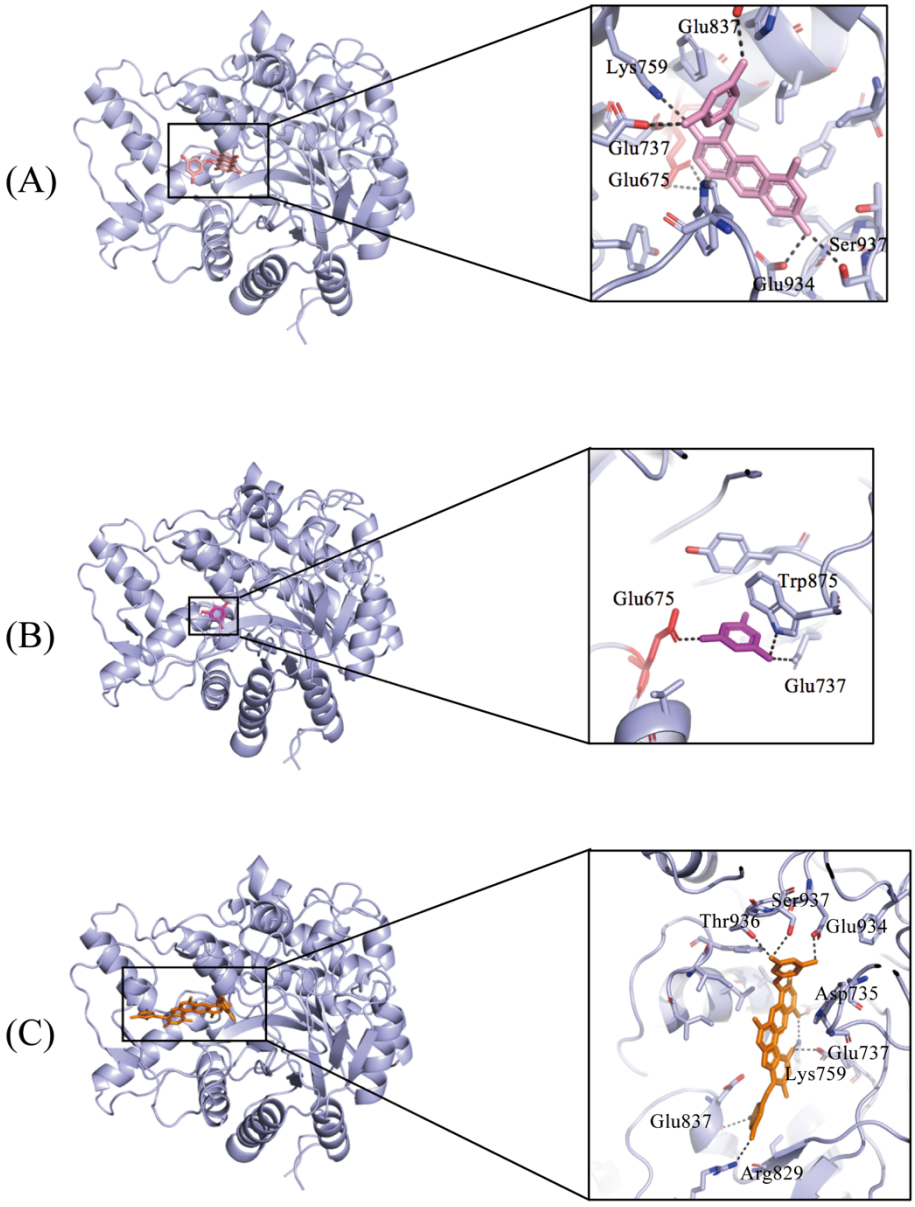



Fig 48 Docking result of BGL with different phlorotannins. BGL is colored in light cyan, represented in cartoon. Hydrogen bond is shown in gray dash. (A) Dock of eckol into BGL. Eckol is colored in pink. (B) Dock of phloroglucinol into BGL. Phloroglucinol is colored in magenta. (C) Dock of phlorofucofuroeckol A into BGL. Phlorofucofuroeckol A is colored in orange.

The interaction analysis of the BGL with phlorotannins, combined with the interaction analysis of BGL with tannic acid shows that they share a similar binding manner. All the complex is stabilized by the hydrogen bond and hydrophobic interactions. Specifically, there are some residues involve all the binding. This might partly explain why all these compounds can inhibit the BGL activity.

Based on the structure analysis and molecular docking study result, the mechanism of tannic acid inhibits BGL activity and mechanism of EHEP protect BGL from inhibition were proposed. In conditions which phlorotannins were not present, the substrate bind to the active site. If there are phlorotanins are present, phlorotannins bind to the substrate binding site, thus preventing the substrate binding, inhibiting the BGL activity (Fig 49). When there is EHEP present, EHEP binds with phlorotannins, releasing the BGL substrate binding site. Therefore, the substrate binds to the active site of BGL, and BGL becomes active (Fig 50).

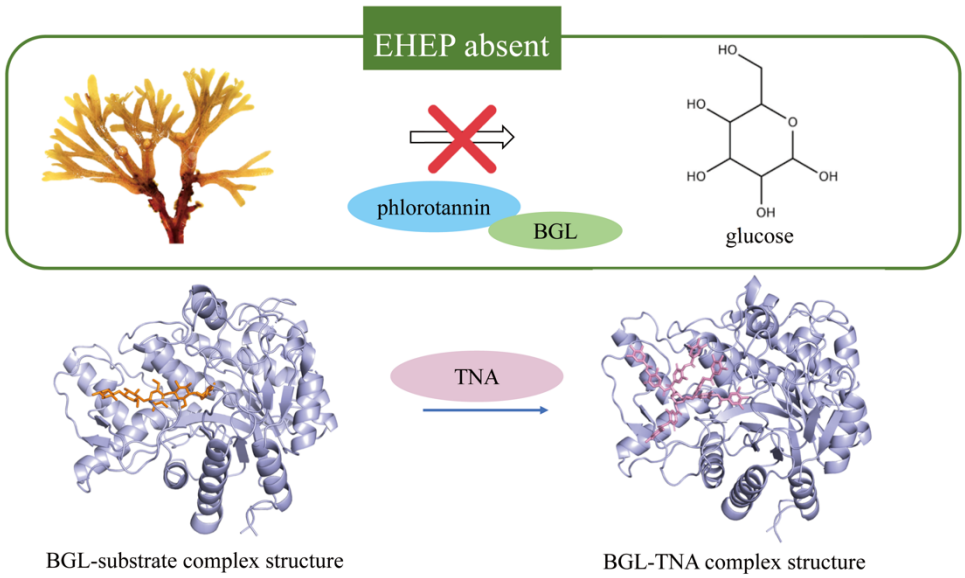



Fig 49 Proposed mechanism if TNA inhibit BGL activity.

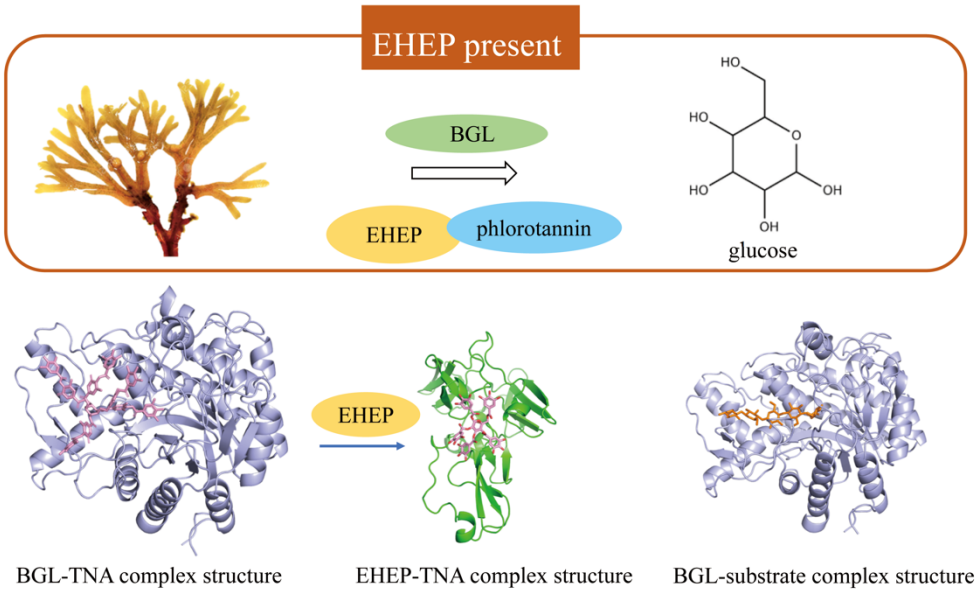



Fig 50 Proposed mechanism of EHEP protects BGL from TNA inhibition.

# Reference


Adamczyk, B., Simon, J., Kitunen, V., Adamczyk, S. & Smolander, A. (2017). *ChemistryOpen* **6**, 610-614.

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 213-221.

Amiri, M., Diekmann, L., von Kockritz-Blickwede, M. & Naim, H. Y. (2015). *Nutrients* **7**, 7209-7230.

Canon, F., Pate, F., Cheynier, V., Sarni-Manchado, P., Giuliani, A., Perez, J., Durand, D., Li, J. & Cabane, B. (2013). *Langmuir* **29**, 1926-1937.

Chuenchor, W., Pengthaisong, S., Robinson, R. C., Yuvaniyama, J., Oonanant, W., Bevan, D. R., Esen, A., Chen, C. J., Opassiri, R., Svasti, J. & Cairns, J. R. (2008). *J Mol Biol* **377**, 1200-1215.

Chuenchor, W., Pengthaisong, S., Robinson, R. C., Yuvaniyama, J., Svasti, J. & Cairns, J. R. (2011). *J Struct Biol* **173**, 169-179.

Deaville, E. R., Green, R. J., Mueller-Harvey, I., Willoughby, I. & Frazier, R. A. (2007). *J Agric Food Chem* **55**, 4554-4561.

Ellabban, O., Abu-Rub, H. & Blaabjerg, F. (2014). *Renewable and Sustainable Energy Reviews* **39**, 748-764.

Emsley, P. & Cowtan, K. (2004). *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132.

Enquist-Newman, M., Faust, A. M., Bravo, D. D., Santos, C. N., Raisner, R. M., Hanel, A., Sarvabhowman, P., Le, C., Regitsky, D. D., Cooper, S. R., Peereboom, L., Clark, A., Martinez, Y., Goldsmith, J., Cho, M. Y., Donohoue, P. D., Luo, L., Lamberson, B., Tamrakar, P., Kim, E. J., Villari, J. L., Gill, A., Tripathi, S. A., Karamchedu, P., Paredes, C. J., Rajgarhia, V., Kotlar, H. K., Bailey, R. B., Miller, D. J., Ohler, N. L., Swimmer, C. & Yoshikuni, Y. (2014). *Nature* **505**, 239-243.

Eom, S. H., Lee, S. H., Yoon, N. Y., Jung, W. K., Jeon, Y. J., Kim, S. K., Lee, M. S. & Kim, Y. M. (2012). *J Sci Food Agric* **92**, 2084-2090.

Evans, P. R. & Murshudov, G. N. (2013). *Acta Crystallogr D Biol Crystallogr* **69**, 1204-1214.

Fadel, F., Zhao, Y., Cousido-Siah, A., Ruiz, F. X., Mitschler, A. & Podjarny, A. (2016). *PLoS One* **11**, e0154190.

Frazier, R. A., Papadopoulou, A., Mueller-Harvey, I., Kissoon, D. & Green, R. J. (2003). *Journal of Agricultural and Food Chemistry* **51**, 5189-5195.

Guillen, D., Sanchez, S. & Rodriguez-Sanoja, R. (2010). *Appl Microbiol Biotechnol* **85**, 1241-1249.

Hakulinen, N., Paavilainen, S., Korpela, T. & Rouvinen, J. (2000). *J Struct Biol* **129**, 69-79.

Hayashi, Y., Okino, N., Kakuta, Y., Shikanai, T., Tani, M., Narimatsu, H. & Ito, M. (2007). *J Biol Chem* **282**, 30889-30900.

Hendrickson, W. A. & Teeter, M. M. (1981). *Nature* **290**, 107-113.

Hollebeke, J., Van Damme, P. & Gevaert, K. (2012). *Biol Chem* **393**, 291-298.

Holm, L. & Rosenstrom, P. (2010). *Nucleic Acids Res* **38**, W545-549.

Hurlburt, N. K., Chen, L. H., Stergiopoulos, I. & Fisher, A. J. (2018). *PLoS Pathog* **14**, e1007263.

Jacob, R., Peters, K. & Naim, H. Y. (2002). *J Biol Chem* **277**, 8217-8225.

Jeng, W. Y., Wang, N. C., Lin, C. T., Chang, W. J., Liu, C. I. & Wang, A. H. (2012). *Acta Crystallogr D Biol Crystallogr* **68**, 829-838.

Jie Fu, J. x., Zhang, Y. & Lu, X. (2015). *Asian Journal of Chemistry* **27**, 3328-3332.

Jung, H. A., Oh, S. H. & Choi, J. S. (2010). *Bioorg Med Chem Lett* **20**, 3211-3215.

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

Kanakis, C. D., Hasni, I., Bourassa, P., Tarantilis, P. A., Polissiou, M. G. & Tajmir-Riahi, H. A. (2011). *Food Chem* **127**, 1046-1055.

Kayukawa, C. T. M., de Oliveira, M. A. S., Kaspchak, E., Sanchuki, H. B. S., Igarashi-Mafra, L. & Mafra, M. R. (2019). *Food Chem* **275**, 346-353.

Kellogg, J., Grace, M. H. & Lila, M. A. (2014). *Marine Drugs* **12**, 5277-5294.

Ketudat Cairns, J. R. & Esen, A. (2010). *Cell Mol Life Sci* **67**, 3389-3405.

Kitago, Y., Watanabe, N. & Tanaka, I. (2005). *Acta Crystallogr D Biol Crystallogr* **61**, 1013-1021.

Kopetz, H. (2013). *Nature* **494**, 29-31.

La Barre, S., Potin, P., Leblanc, C. & Delage, L. (2010). *Mar Drugs* **8**, 988-1010.

Lange, P. F. & Overall, C. M. (2011). *Nat Methods* **8**, 703-704.

Lee, S., Choi, J., Mohanty, J., Sousa, L. P., Tome, F., Pardon, E., Steyaert, J., Lemmon, M. A., Lax, I. & Schlessinger, J. (2018). *Nature* **553**, 501-505.

Li, X. & Ni, T. (2016). *J Biol Phys* **42**, 415-434.

Li, Y., Fu, X., Duan, D., Liu, X., Xu, J. & Gao, X. (2017). *Mar Drugs* **15**.

Lin, Y. & Tanaka, S. (2006). *Appl Microbiol Biotechnol* **69**, 627-642.

Luo, Q., Zeng, S., Shu, Y., Fu, Z., Zhao, H. & Su, S. (2018). *RSC Advances* **8**, 17151-17158.

Marana, S. R., Mendonca, L. M., Andrade, E. H., Terra, W. R. & Ferreira, C. (2003). *Eur J Biochem* **270**, 4866-4875.

Matthews, B. W. (1968). *J.Mol.Biol* **33**, 491-497.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J Appl Crystallogr* **40**, 658-674.

Mueller, G. A., Randall, T. A., Glesner, J., Pedersen, L. C., Perera, L., Edwards, L. L., DeRose, E. F., Chapman, M. D., London, R. E. & Pomes, A. (2016). *Clin Exp Allergy* **46**, 365-376.

Ozdal, T., Capanoglu, E. & Altay, F. (2013). *Food Research International* **51**, 954-970.

Papadopoulou, A., Green, R. J. & Frazier, R. A. (2005). *J Agric Food Chem* **53**, 158-163.

Poncet-Legrand, C., Gautier, C., Cheynier, V. & Imberty, A. (2007). *Journal of Agricultural and Food Chemistry* **55**, 9235-9240.

Rosano, G. L. & Ceccarelli, E. A. (2014). *Front Microbiol* **5**, 172.

Roy, D., Dutta, S., Maity, S. S., Ghosh, S., Singha Roy, A., Ghosh, K. S. & Dasgupta, S. (2012). *Journal of Luminescence* **132**, 1364-1375.

Sabbadin, F., Pesante, G., Elias, L., Besser, K., Li, Y., Steele-King, C., Stark, M., Rathbone, D. A., Dowle, A. A., Bates, R., Shipway, J. R., Cragg, S. M., Bruce, N. C. & McQueen-Mason, S. J. (2018). *Biotechnol Biofuels* **11**, 59.

Sakamoto, K., Uji, S., Kurokawa, T. & Toyohara, H. (2009). *Gene* **435**, 72-79.

Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. (2013). *Journal of Computer-Aided Molecular Design* **27**, 221-234.

Sheldrick, G. M. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 479-485.

Shibata, T., Fujimoto, K., Nagayama, K., Yamaguchi, K. & Nakamura, T. (2002). *International Journal of Food Science and Technology* **37**, 703-709.

Silva, R. D. & Martinho, R. G. (2015). *Proteomics* **15**, 2402-2409.

Suetake, T., Tsuda, S., Kawabata, S., Miura, K., Iwanaga, S., Hikichi, K., Nitta, K. & Kawano, K. (2000). *J Biol Chem* **275**, 17929-17932.

Tamaki, F. K., Souza, D. P., Souza, V. P., Ikegami, C. M., Farah, C. S. & Marana, S. R. (2016). *PLoS One* **11**, e0167978.

Tsuji, A., Kuwamura, S., Shirai, A. & Yuasa, K. (2017). *PLoS One* **12**, e0170669.

Tsuji, A., Tominaga, K., Nishiyama, N. & Yuasa, K. (2013). *PLoS One* **8**, e65418.

Van Alstyne, K. L. & Paul, V. J. (1990). *Oecologia* **84**, 158-163.

Watrelot, A. A., Renard, C. M. G. C. & Le Bourvellec, C. (2015). *Lwt-Food Science*

*and Technology* **63**, 376-382.

Wei, N., Quarterman, J. & Jin, Y. S. (2013). *Trends Biotechnol* **31**, 70-77.

Wu, X., Dey, R., Wu, H. U. I., Liu, Z., He, Q. & Zeng, X. (2013). *International Journal of Dairy Technology* **66**, 7-13.

# Acknowledgements

Completing the PhD course is a challenging experience for me, and maybe it would not have been possible to do without the support and guidance that many people offered.

I would like to firstly express my sincere gratitude to my supervisor, Professor Min Yao for all her continuous guidance and encouragement she gave me during these years. She devoted herself to the scientific research, which really inspired and encourage me.

I greatly appreciate Professor Akihiko Tsuji (Tokushima University) from for his valuable suggestions. I would also like to thank the Associate Professor Toyoyuki Ose, Assistant Professor Jian Yu, Koji Kato for your numerous suggestions to my research. I am also grateful to research assistant Naofumi Sakurai for his help to my research.

I gratefully acknowledge the IGP program for their finance support.

I thank all the members in the lab, for the stimulating discussions, for the funs we have had in the last three years. It was great sharing lab with all of you during last three years.

Last but not least, I would like to thank my family, for supporting me spiritually throughout during my life.