*J. Biochem.* **144,** 467–475 (2008)
doi:10.1093/jb/mvn095

# Computational and Experimental Analyses of Furcatin Hydrolase for Substrate Specificity Studies of Disaccharide-specific Glycosidases

**Hiromi Daiyasu[1,*], Hiromichi Saino[2], Hiroo Tomoto[2], Masaharu Mizutani[2], Kanzo Sakata[2] and Hiroyuki Toh[3]**

*[1]The Center for Advanced Medical Engineering and Informatics, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531; [2]Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011; and [3]Division of Bioinformatics, Medical Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka, Fukuoka 812-8582, Japan*

Received May 31, 2008; accepted July 5, 2008; published online July 30, 2008

**Disaccharide-specific glycosidases (diglycosidases) are unique glycoside hydrolases, as their substrate specificities differ from those of monosaccharide-specific β-glycosidases (monoglycosidases), in spite of similarities in their sequences and reaction mechanisms. Diglycosidases selectively hydrolyse the β-glycosidic bond between glycone and aglycone of disaccharide glycosides, but do not cleave the bond between two saccharides, and barely hydrolyse monosaccharide glycosides. We analysed the substrate recognition mechanisms of diglycosidases by computational and experimental methods, using furcatin hydrolase (FH) (EC 3.2.1.161) derived from *Viburnum furcatum*. Amino acid sequence comparisons and model structure building revealed two residues, Ala419 and Ser504 of FH, as candidates determining the substrate specificity. These residues were specifically conserved in the diglycosidases. The model structure suggested that Ala419 is involved in the aglycone recognition, whereas Ser504 recognizes the external saccharide of the glycone. Mutations at these sites drastically decreased the diglycosidase activity. The mechanism by which the diglycosidases acquired their substrate specificity is discussed, based on these observations.**

**Key words: β-glucosidase, disaccharide glycoside, furcatin hydrolase, homology modelling, mutagenesis.**

Abbreviations: diglycosidases, disaccharide-specific glycosidases; diglycosides, disaccharide glycosides; FH, furcatin hydrolase; GH, glycosyl hydrolase; GST, glutathione S-transferase; monoglycosidases, monosaccharide-specific β-glycosidases; monoglycosides, monosaccharide glycosides; NJ, neighbour-joining; *p*AP, *para*-allylphenyl; PD, β-primeverosidase; *p*NP, *para*-nitrophenyl; PSG, *para*-nitrophenyl-β-D-thioglucoside; TIM, triose phosphate isomerase; VH, vicianin hydrolase.

β-Glucosidases exist in almost all living organisms and are involved in a wide variety of biological reaction pathways, such as starch and sucrose metabolism, cyanoamino acid metabolism and phenylpropanoid biosynthesis, as described in the KEGG database (http://www.genome.ad.jp/kegg/). They hydrolyse the β-glycosidic bond of a β-glucoside between a glucose and an aglycone (non-carbohydrate moiety) (*1*, *2*) (Fig. 1). Most of the plant β-glucosidases are classified into glycosyl hydrolase (GH) family 1, based on the sequence similarity in the CAZY (Carbohydrate Active enZYmes) database (http://www.cazy.org) (*3*, *4*). The tertiary structures of GH family 1 proteins share an $(\alpha/\beta)_8$ triose phosphate isomerase (TIM) barrel structure, and their core secondary structures are highly conserved, even when the sequence identity is <25%. They have two conserved Glu residues, which are essential for hydrolysis via the retaining mechanism (*2*, *5*, *6*). These residues reside at the bottom of the slot-like barrel hole. Around the active site, a highly conserved loop, bearing several aromatic and polar residues, interacts with the glucose of the substrate via hydrogen bonds (*7*). Biological studies of numerous plant β-glucosidases revealed their diverse substrate specificity. Through the catalytic reaction, these enzymes release various kinds of aglycone moieties as bioactive metabolites for defense mechanisms. For example, hydrogen cyanide is generated from inactive β-glycosides via hydrolysis in several edible plants, such as cassava and black cherry (*5*, *8–10*).

In addition to the monosaccharide glycosides (monoglycosides), with a glucose unit for a glycone, the disaccharide glycosides (diglycosides), in which the glycones are glucose with another monosaccharide, exist in many plants. Several glycosidases that can utilize these diglycosides as their substrates have also been detected in plants (*11–15*). They release a disaccharide unit by the selective hydrolysis of the β-glycosidic linkage between the disaccharide glycone moiety and the aglycone (Fig. 1), and therefore, they do not hydrolyse the bond connecting the two sugars. Furthermore, they hardly hydrolyse monoglycosides, which distinguishes these glycosidases from other β-glycosidases.

*To whom correspondence should be addressed. Tel: +81-6-6850-6603, Fax: +81-6-6850-6602, E-mail: daiyasu@ist.osaka-u.ac.jp

 

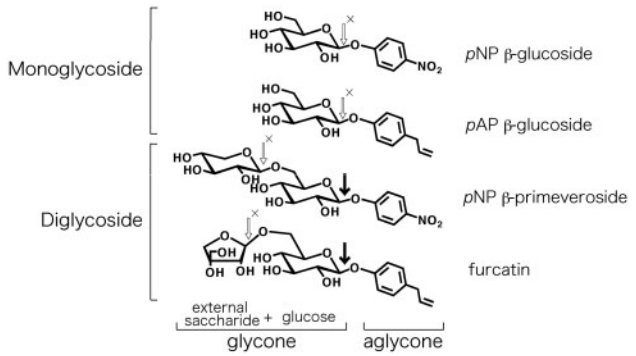


Fig. 1. **Schematic diagrams of monoglycosides and diglycosides.** An empty arrow indicates the β-glycosidic bond hydrolysed by monoglycosidases, whereas diglycosidases cleave the β-glycosidic bond indicated by a solid arrow. 'X' indicates a bond that is barely hydrolysed by diglycosidases.

Such a disaccharide-specific glycosidase is hereafter called a diglycosidase, whereas a β-glycosidase that liberates a monosaccharide is referred to as a monoglycosidase. We identified two diglycosidases, β-primeverosidase (PD), derived from *Camellia sinensis*, and furcatin hydrolase (FH), from *Viburnum furcatum* and analysed their biological functions (*16*, *17*). They share the highest amino acid sequence similarity (64% identity), although their taxonomic relationship is not close. They are classified into GH family 1, based on the sequence similarity to plant monoglycosidases, and their reactions also follow the retaining mechanism. The natural substrates of PD are β-primeverosides (6-*O*-β-D-xylopyranosyl-β-D-glucopyranosides), and the reaction is involved in either aroma formation or defense mechanisms (*16*, *18*). FH hydrolyses furcatin [*para*-allylphenyl (*p*AP) 6-*O*-β-D-apiofuranosyl-β-D-glucopyranoside] into *p*-allylphenol and the disaccharide, and it works as a deterrent against insect attack by releasing a toxic aglycone (*17*). Although the substrate specificity differs somewhat between FH and PD, *para*-nitrophenyl (*p*NP) β-primeveroside is their common substrate (*17*, *18*). A previous molecular phylogenetic analysis suggested that these diglycosidases form a single subcluster in the plant monoglycosidases cluster (*17*). Although the substrate specificities and the evolutionary relationship of the diglycosidases have been investigated, the substrate recognition mechanism has not been clarified at the molecular level. The tertiary structures of the diglycosidases have not been solved yet. Recently, Han *et al*. (*19*) suggested a substrate recognition mechanism for PD, based on a molecular dynamics simulation with the model structure. In their model structure, however, the substrate bond targeted for hydrolysis was far from the catalytic site. Therefore, their model structure does not seem to represent the actual binding mode or explain the catalytic mechanism.

To investigate the reaction mechanism underlying the unique substrate specificities of the diglycosidases, we compared the amino acid sequences of the di- and mono-glycosidases, and constructed a model structure of FH including its substrate. Among the 538 residues of the FH sequence, we identified two candidates that contribute to the substrate specificity. Based on computational analyses, these residues were examined by site-directed mutagenesis. The diglycosidase activities of these mutants were remarkably reduced. We discuss the substrate recognition mechanism of the diglycosidases, based on a synergistic approach utilizing computational and experimental studies.

## MATERIALS AND METHODS

*Amino Acid Sequence Comparison and Phylogenetic Tree Construction*—The proteins with significant amino acid sequence similarity to FH were collected by a BLAST (*20*) search of the non-redundant protein sequence database at the NCBI site (http://www.ncbi.nlm.nih.gov/BLAST/). An *E*-value cut-off of $<10^{-125}$ was used as the criterion for significant sequence similarity. When some proteins thus collected showed >80% sequence identity to each other, one of them was selected for the following analysis as a representative of the group. In total, 35 sequences were obtained, and a multiple alignment of these sequences was produced with the alignment software MAFFT ver.5.8 (*21*).

Based on the alignment, an unrooted molecular phylogenetic tree was constructed by the neighbour-joining (NJ) method (*22*). The genetic distance between every pair of aligned sequences was calculated as a maximum likelihood estimate (*23*), using the JTT model (*24*) for the amino acid substitutions. The sites including gaps in the alignment were excluded from the calculation. The statistical significance of the NJ tree topology was evaluated by a bootstrap analysis (*25*) with 1,000 iterative tree reconstructions. Two software packages, PHYLIP 3.5c (*26*) and MOLPHY 2.3b3 (*27*), were used for the phylogenetic analysis.

*Homology Modelling*—A model structure of FH with the substrate was constructed, using the coordinates of *Trifolium repens* cyanogenic β-glucosidase (PDB ID: 1CBG; resolution: 2.15 Å) as a template, based on the phylogenetic analysis (see RESULTS section). The sequence identity between FH and 1CBG was 55%. *p*NP β-primeveroside was selected as a substrate model, because the compound is known as a good substrate for both FH and PD (*17*, *18*), and thus it seemed to be suitable to extract the common characteristics of these diglycosidases. To build the glycone moiety (xylose and glucose) of *p*NP β-primeveroside, the coordinates of *p*NP β-D-thioglucoside (PSG), an inhibitor included in the crystal structure of *Zea mays* β-glucosidase (PDB ID: 1E1F; resolution: 2.6 Å) were used. The two structures, 1CBG and 1E1F, were superimposed based on the sequence alignment, and the coordinates of PSG were transferred into the corresponding position of 1CBG. The inhibitor was then changed to *p*NP β-primeveroside in the following manner. The C6 of the glucose in PSG was modified with the xylose by a 6-*O*-glycosidic bond, and the *S*-linked β-glycosidic bond between glucose and *p*NP was changed to the *O*-linked bond. The coordinates of 1CBG with *p*NP β-primeveroside were used as an initial structure to build the model. All operations for the homology modelling were performed with the molecular modeling tool MOE (Molecular Operating Environment, ver. 2005.06, Chemical Computing Group,

Montreal, Canada). Energy minimization was performed with the MMFF94s force field (*28*), using an implicit solvent model. The quality of the model structure was checked with Verify3D (http://nihserver.mbi.ucla.edu/Verify_3D/).

*Site-directed Mutagenesis*—The bacterial expression plasmid pGEX-FH-WT, which expresses the mature form of wild-type FH fused with glutathione S-transferase (GST) at the N-terminus, was used as the template (*17*). Primers for the mutations were designed as follows: 5′-ACGATTGGGAATGGAAC**GCT**GGTTTCA ACGTAAGG-3′ for S504A, 5′-ACGATTGGGAATGGAAC**T GG**GGTTTCAACGTAAGG-3′ for S504W, 5′-TCACTTCT AGACGATTGG**CWA**TGGAACTCTGGTTT-3′ for E501Q and E501L, 5′-TCACTTCTAGACGATTGG**GMC**TGGAA CTCTGGTTT-3′ for E501A and E501D, 5′-TCCGGCG ACTGGTACACC**GTG**GTTTTGTTTCTGTCCAGAA-3′ for A419V, and 5′-CCGGCGACTGGTACACCT**TGG**TTTTG TTTCTGTCCAGAA-3′ for A419W (the mutated codon is shown in bold). Site-directed mutagenesis was carried out with a QuikChange Multi site-directed mutagenesis kit (Stratagene, La Jolla, CA, USA), according to the manufacturer's instructions. *Escherichia coli* BL21 cells were transformed with either the wild-type plasmid (pGEX-FH-WT) or the mutated plasmid. A 60 ml culture of the transformed cells was grown to log phase ($A_{600} = 0.6$) at 37°C, and then isopropyl-β-D-thiogalactopyranoside was added to a final concentration of 1 mM. The induced culture was incubated at 25°C for an additional 15 h. The cells were harvested by centrifugation, and were suspended in 50 mM Tris–HCl buffer (pH 8.0) containing 0.2 mg/ml lysozyme (WAKO Pure Chemicals, Osaka, Japan). After incubation for 10 min at 30°C, the cells were sonicated 15 times for 30 s intervals at 4°C, and the lysate was centrifuged at 14,000*g* for 10 min. The recombinant protein in the supernatant was purified by passage through a glutathione-Sepharose column (0.7 × 2.5 cm) (GE Healthcare Bio-Science AB, Uppsala, Sweden). SDS–PAGE analysis of eluted fractions exhibited single band at 88 kDa, and the purity of FH mutants were equal to that of GST–FH fusion protein purified by previous method (*17*). The purified enzyme was utilized in the enzyme assay without cleavage of the GST-tag.

*Enzyme Assays*—The activities of the wild-type and mutant enzymes were measured with *p*NP β-primeveroside, *p*NP β-glucoside, furcatin and *p*AP β-glucoside. *p*NP β-glucoside were purchased from Sigma-Aldrich Co. (MO, USA). *p*NP β-primeveroside was kindly provided by Amano Enzyme Co. (Nagoya, Japan). Furcatin and *p*AP β-glucoside was prepared as described previously (*17*). Assays were carried out at 37°C using a final concentration of 2 mM substrate and 20 mM sodium citrate buffer (pH 5.0), in a total volume of 50 µl. The reaction was started by adding 5 µl of the enzyme, and was stopped by adding 50 µl of 0.2 M $Na_2CO_3$. The initial velocities were determined by monitoring the liberated aglycones, *p*-nitrophenol with a spectrophotometer at 405 nm and *p*-allylphenol with HPLC, as described previously (*17*).

The kinetic parameters of the wild-type and mutant enzymes were determined with *p*NP β-primeveroside. The assay conditions were the same as those described above, except the substrate concentration ranged from 1 to 25 mM. The initial velocities for *p*NP β-primeveroside at each concentration were fit to the Michaelis–Menten equation by a non-linear regression analysis using KaleidaGraph (Synergy Software, PA, USA), and the constant for the curve yielded $K_m$ and $k_{cat}$.

## RESULTS

*Computational Analysis*—Database searching yielded 1,114 proteins with sequence similarity to FH. The top of the list for the detected sequences was occupied by hundreds of plant glycosidases, including PD. The GH family 1 proteins derived from other eukaryotes, eubacteria and archaea followed the plant enzymes in the list. We expected that sequence and structure comparisons between FH and its closely related monoglycosidases would identify the residues involved in the unique substrate recognition. Therefore, we collected FH and the top 34 proteins sequences with *E*-values $<10^{-125}$, as described in the MATERIALS AND METHODS section. All of them were plant enzymes, and three of the proteins had solved tertiary structures (1CBG, 1E1F and 1V02 in Fig. 3). Among the selected enzymes, vicianin hydrolase (VH) from *Vicia angustifolia* was recently identified (*29*) in addition to two previous diglycosidases, PD and FH. As described below, the acquisition of diglycosidase activity by VH is considered to be independent from those of PD and FH. Therefore, we did not use this enzyme in the comparative analysis.

The obtained sequences were aligned, and the partial alignment data of the representative glycosidases are shown in Fig. 2. In this study, an alignment site at which >80% of the total sequences have the same or similar amino acids was defined as a conserved site. According to this definition, about half of the alignment sites were conserved, and these sites were mainly found in the core secondary structure of the TIM barrel fold. This observation suggested that the tertiary structures of the diglycosidases are similar to those of the monoglycosidases. Two catalytic Glu residues corresponded to the alignment sites 176 and 411, as shown in Fig. 2. In addition to them, the alignment sites 23, 127, 463, 470 and 471 in Fig. 2, which correspond to glucose binding residues of the monoglycosidases (*7*), were also conserved in FH and PD. In the case of FH, Gln88, His192, Trp494, Glu501 and Trp502 were respectively aligned to these sites. This conservation suggested that the diglycosidases interact with the glucose unit of the substrate, in a similar manner as the monoglycosidases.

A phylogenetic tree of GH family 1 was constructed, based on the multiple alignment of the 35 homologous sequences, to select a template structure for model building (Fig. 3). The tree suggested that 1CBG is the most closely related to FH among the three monoglycosidases with known structures, and thus, this β-glucosidase was chosen as the template. The tree also suggested that FH and PD evolved from a common ancestor of plant monoglycosidases. This observation is consistent with the phylogenetic analyses in the previous reports (*16*, *17*). On the other hand, VH was far from these diglycosidases, but close to the other monoglycosidases.

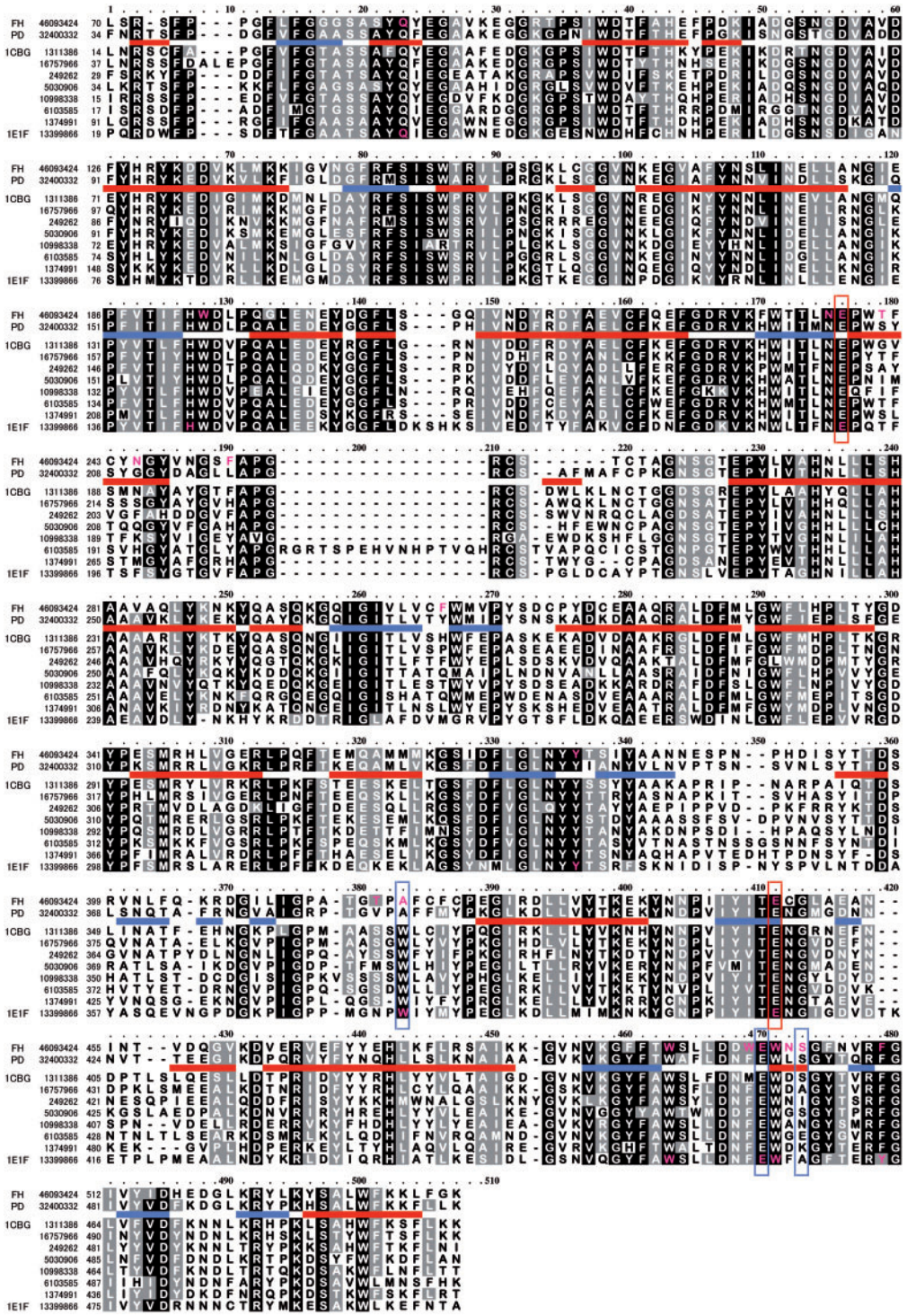

Fig. 2. **A multiple alignment of diglycosidases and monoglycosidases.** The alignment, including FH, PD and eight monoglycosidases, was reconstructed from the multiple alignment of 34 sequences. The GI number of each protein is shown in the left column. The number neighbouring the first residue in each line indicates the residue number in the sequence. A measure indicating the site is shown over the alignment. A conserved site with the same or physicochemically similar amino acid residues in >80% of the total sequences is indicated by an inverse character with a black or grey background, respectively. The criteria of Dayhoff *et al*. (*33*) were used for the classification of physicochemically similar amino acids: (i) hydrophobic group, L, I, M and V; (ii) aromatic group, F, Y and W; (iii) small hydrophilic group, S, T, P, A and G; (iv) negatively charged group and the relatives, D, E, Q and N; (v) positively charged group, K, R and H; and (vi) Cys residue, C. The catalytic glutamates are enclosed by red lines. The residues within 3.5 Å from the substrate in the FH model structure and 1E1F are coloured red. The substituted sites in the FH mutant are enclosed by blue lines. The α-helices and β-strands of 1CBG are indicated by red and blue bars, respectively. Gaps for insertions and/or deletions are indicated by hyphens.

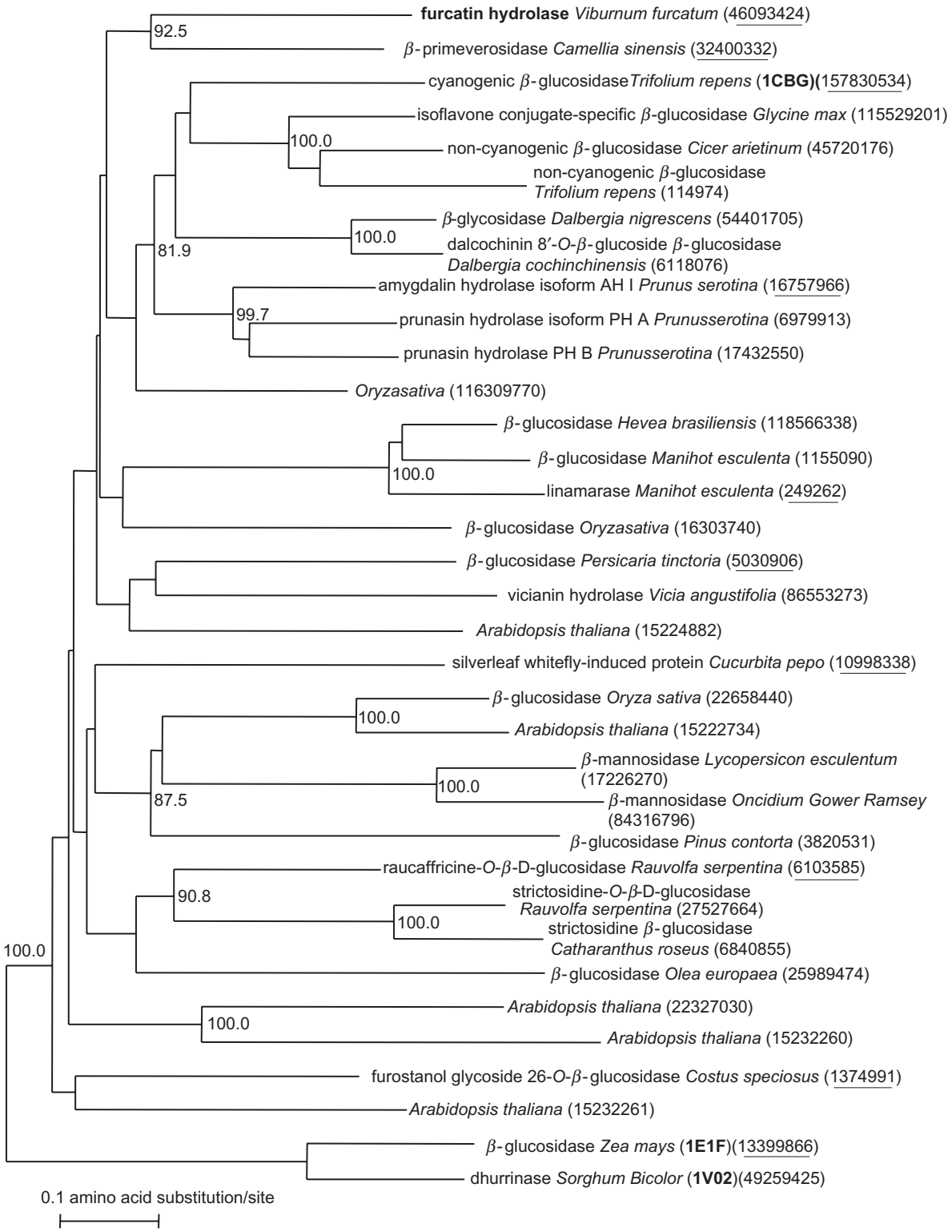


Fig. 3. **An unrooted phylogenetic tree of diglycosidases and monoglycosidases.** The protein name, source name and GI number or PDB ID of each sequence are indicated at each leaf. The GI numbers for the 10 sequences indicated in Fig. 2 are underlined. For the hypothetical sequences derived from genome data, only the source names and the GI numbers are indicated. Bootstrap probabilities of clustering at a node of >80.0% are indicated.

We built a model structure of FH by homology modelling including *p*NP β-primeveroside, as described in the MATERIALS AND METHODS section. This model structure was specifically constructed to represent the substrate binding mode, as described below. As shown in Fig. 4A, it formed a typical TIM barrel fold. The two deleted regions of FH, as compared to 1CBG (the alignment sites 214–218 and 424–425 in Fig. 2), corresponded to the loops, which were far from the substrate. Except for these regions, no global conformational change was found. The carboxylate groups of the catalytic Glu residues, Glu238 and Glu447 (the alignment sites 176 and 411 in Fig. 2), were ~5 Å apart from each other in the model structure. This location is suitable for catalysis via the retaining catalytic mechanism (*6*). An oxygen in the carboxylate of Glu447, which functions as the catalytic nucleophile, was close to both C1 (4.4 Å) and the hydrogen of O2 (2.4 Å) in the glucose unit of the substrate, whereas an oxygen in the carboxylate of Glu238, which functions as the acid/base catalyst, was near the glycosidic oxygen between the aglycone and the glycone (2.9 Å). Thus, the locations of the catalytic residues and the substrate binding mode were reproduced well in the model structure.

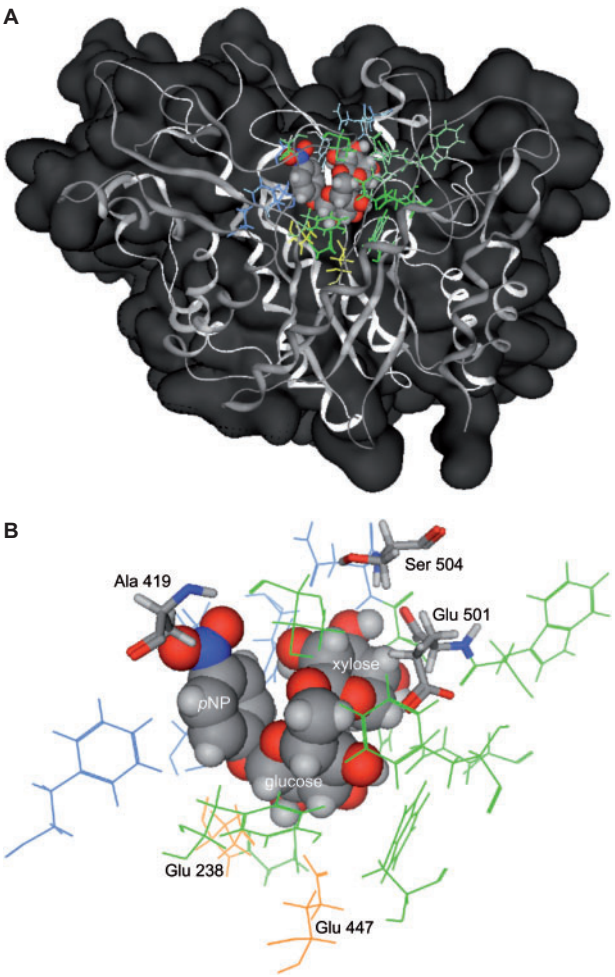


Fig. 4. **(A) The model structure of FH with the substrate.** The backbone of the model structure of FH is shown by a grey tube. The residues within 3.5Å from the substrate atoms are indicated by line models, which include two catalytic Glu residues (yellow), the conserved residues within GH family 1 (green) and others (blue). *p*NP β-primeveroside is indicated by a space filling model, coloured by atom type. (B) Active site of the FH model structure. *p*NP β-primeveroside and the amino acids within 3.5 Å from the substrate atoms are shown. The three mutated sites are indicated by stick models. Other residues and the substrate are shown in the same manner as in (A).

The structural comparison between the di- and monoglycosidases suggested that the difference in the substrate specificity of these enzymes was not caused by global conformational changes. Therefore, we selected the residues surrounding the substrate in the model structure as the candidates to determine the substrate specificity of FH. There were 19 residues within 3.5 Å from the substrate in the model structure, which included the pair of catalytic Glu residues (Fig. 4B). Among these 19 candidates, two residues, Ala419 at alignment site 383 and Ser504 at site 473, were further selected, as they were specifically conserved within FH and PD. In the monoglycosidases, alignment site 383 is occupied by a conserved Trp (Fig. 2). This aromatic residue anchors the aglycone at the entrance of the slot, placing the substrate in the proper position for catalysis in the structure of *Zea mays* β-glucosidase (*7*, *30*). In the FH model structure, Ala419 was also present near the entrance, and was located very close to the aglycone of the substrate (Fig. 4B). On the other hand, alignment site 473 was not conserved within the monoglycosidases. In the model structure, Ser504 was present at the bottom of the catalytic hole and was close to the O3 of xylose, the external saccharide (Fig. 4B). Hereafter, we focused on Ala419 and Ser504 of FH as the candidates to explain the substrate specificity toward diglycosides. These two amino acids of FH were examined by site-directed mutagenesis, in order to investigate their roles in the substrate specificity.

*Site-directed Mutagenesis of FH*—We constructed two mutants for each target site. The hydrolytic activities of these mutants toward diglycosides (furcatin and *p*NP β-primeveroside) and monoglycosides (*p*AP β-glucoside and *p*NP β-glucoside) were measured at the same substrate concentration (2 mM), and the activities were compared with those of wild-type FH (Table 1).

At first, Ala419 was replaced with Trp, which is conserved in monoglycosidases. This A419W mutant exhibited reduced activity toward all substrates, at 24–43% relative to the wild-type. Ala419 was also replaced with a hydrophobic Val, and A419V exhibited 40–127% activities relative to the wild-type. It is notable that the negative effect of the mutation on the hydrolytic activities toward all di- and mono-glycosides is greater in A419W than in A419V, suggesting that larger side chains at this site cause deeper reductions in the hydrolytic activity. As compared to A419W, the difference in the aglycone seemed to be more clearly demonstrated in A419V. Specifically, the decrease in the activity of A419V relative to the wild-type was more remarkable for the substrates with *p*AP, the same aglycone moiety as the natural substrate, than for those with *p*NP.

Ser504 was then replaced with a bulky Trp and also with a small Ala, lacking the hydroxy group as compared

Table 1. **Relative activities of wild-type and mutants of FH.**

| Enzyme | Relative activity (%) | | | |
|---|---|---|---|---|
| | Furcatin | *p*NP β-primeveroside | *p*AP β-glucoside | *p*NP β-glucoside |
| Wild-type | 100 | 100 (91) | 100 (11) | 100 (2) |
| A419W | 26 ± 1 | 24 ± 1 | 43 ± 1 | 39 ± 1 |
| A419V | 40 ± 1 | 78 ± 3 | 56 ± 4 | 127 ± 1 |
| S504W | 1.4 ± 0.030 | 1.3 ± 0.02 | 20 ± 0.5 | 28 ± 0.5 |
| S504A | 113 ± 1 | 62 ± 1 | 87 ± 3 | 87 ± 4 |
| E501L | 0.1 ± 0.002 | ND | ND | ND |
| E501D | 0.2 ± 0.010 | ND | ND | ND |
| E501Q | 0.1 ± 0.003 | ND | ND | ND |
| E501A | 0.1 ± 0.060 | ND | ND | ND |

The numbers in parentheses indicate the relative activity in comparison to the activity of the wild-type enzyme toward furcatin (0.27 μmol *p*-allylphenol released/min/mg protein). ND, not detectable.

Table 2. **Comparison of the kinetic parameters toward *p*NP β-primeveroside.**

| Enzyme | $K_m$ (mM) | $k_{cat}$ ($min^{-1}$) | $k_{cat}/K_m$ ($min^{-1} \cdot mM^{-1}$) | Relative $k_{cat}/K_m$ (%) |
|---|---|---|---|---|
| Wild-type | 4.0 ± 0.5 | 5.9 ± 0.3 | 1.5 | 100 |
| A419W | 1.7 ± 0.2 | 1.0 ± 0.03 | 0.59 | 39 |
| A419V | 1.9 ± 0.4 | 2.5 ± 0.2 | 1.3 | 87 |
| S504W | 23.0 ± 2.4 | 0.3 ± 0.02 | 0.013 | 1 |
| S504A | 7.5 ± 0.8 | 5.6 ± 0.3 | 0.75 | 50 |

to Ser. The activities of S504W toward diglycosides were severely decreased, to 1.4 and 1.3% relative to those of the wild-type (Table 1). In contrast, the effect of the mutation toward monoglycoside hydrolysis was relatively modest, with 20 and 28% activities relative to the wild-type. On the other hand, the activity of S504A toward furcatin increased 113% relative to the wild-type, and the activities toward the other substrates slightly decreased as compared to the wild-type. Thus, when Ser504 was substituted with an amino acid bearing a bulky side chain, a remarkable difference in the activity with the glycone, rather than the aglycone, was detected. However, such dependence on the glycone was not observed when Ser504 was substituted with an amino acid bearing a similar side-chain size.

*Kinetic Analysis of the Mutants*—To further investigate the roles of Ala419 and Ser504 in the substrate specificity of FH, the kinetic parameters of their mutants were determined with *p*NP β-primeveroside (Table 2). The relative $k_{cat}/K_m$ values were similar to the relative activities toward *p*NP β-primeveroside in Table 1. The $K_m$ values of A419W and A419V were decreased to half of the wild-type value, indicating enhanced affinity of the mutants for *p*NP β-primeveroside. However, the lower $k_{cat}$ values (1/6 and 1/2 relative to the wild-type) indicated that the catalytic rates were decreased by the mutation. Thus, the lower overall activity ($k_{cat}/K_m$) was caused by the reduction of $k_{cat}$.

The $K_m$ value of S504W was increased 5-fold relative to that of the wild-type, and the $k_{cat}$ value was decreased to 1/20 of the wild-type. The increase in $K_m$ indicates reduced affinity of the mutant toward the substrate, suggesting that the bulky Trp prevents the binding of *p*NP β-primeveroside by steric hindrance. On the other hand, S504A exhibited a high catalytic rate, with a $k_{cat}$ comparable to that of the wild-type, but its $K_m$ value was slightly increased. The high $k_{cat}$ value suggested that the substrate was bound to S504A at an accurate position in the active site.

## DISCUSSION

Both computational studies and mutagenesis experiments suggested that at least two residues of FH are involved in the substrate specificity. Ala419 was located close to the aglycone of the diglycoside in the model structure. The kinetic study with *p*NP β-primeveroside revealed that the mutations at this site did not disturb the binding, but decreased the catalytic rate (Table 2). From these observations, the hydrophobic interaction between the aglycone and Trp or Val may enhance the affinity in these mutants, whereas the bulky side chain may shift the glycosidic oxygen away from the catalytic glutamates, resulting in the lower overall activity. The low relative activity toward furcatin (Table 1) may be explained by the same mechanism as with *p*NP β-primeveroside. However, there was a subtle difference in the relative activity between A419W and A419V. As described above, the decrease in the activity of A419V relative to the wild-type was more remarkable when the substrates included *p*AP, rather than *p*NP. This observation suggested that Ala419 is adapted to the interaction with the aglycone of the natural substrate. In contrast, this difference in the relative activity associated with the difference in the aglycone was not observed in A419W. The low catalytic rate of A419W (Table 2) may have obscured the different activities with aglycones. The corresponding site in monoglycosidases is occupied by Trp, which can interact with the aglycone. The small side chain of Ala at this site may enable the catalytic slot of the enzymes to expand, which would facilitate the accommodation of large substrates with two saccharides. That is, FH and PD may have acquired diglycosidase activity by the substitution of the conserved Trp with Ala, to create the space for diglycosides. As shown in Table 1, however, A419W hardly hydrolysed monoglycosides. In addition to the single mutation of Ala419 to Trp, additional amino acid substitutions at other positions

may be required to alter the diglycosidase activity into the monoglycosidase activity.

Ser504 is located at the inner wall of the hole, and is close to the external saccharide of the diglycosides, rather than the aglycone, in the model structure. As described above, the drastic increase in $K_m$ and the decrease in $k_{cat}$ of S504W may be caused by the collision between the bulky side chain of Trp and the substrate, which would prohibit the substrate from binding to the enzyme. Ser504 is considered to play an important role for the interaction with an external saccharide of the glycone, since the decrease in the relative activity of S504W is more remarkable for diglycosides than monoglycosides (Table 1). On the other hand, the slight increase in $K_m$ and the lack of a significant change in $k_{cat}$ of S504A toward *p*NP β-primeveroside suggested that the substrate can bind to this mutant protein in the proper position for the catalysis, although the affinity is slightly decreased. Since Ser and Ala share similar physicochemical properties, such as size and polarity, the mutation from Ser to Ala may not have introduced a large change in the structural environment around the substrate, especially xylose.

We also examined Glu501 by a mutation study, since Glu501 is close to both the glucose and xylose with distances of 2.5–3.5 Å in the model structure. As described above, the corresponding site (alignment site 470 of Fig. 2) is completely conserved with Glu in both the mono- and di-glycosidases. The residue interacts with the hydroxy groups at positions C4 and C6 of glucose in monoglycosidases (*7*), and stabilizes the substrate–enzyme intermediate (*31*). However, the role of the residue in diglycosidases has not been determined. Since the C6-hydroxy group of glucose is linked by the external saccharide in the diglycosides, the Glu in the diglycosidases may interact with both saccharides. Therefore, we mutated Glu501 of FH to different amino acids and examined their catalytic activities. As shown in Table 1, all of the Glu501 mutants practically lost their activities toward both the diglycosides and monoglycosides. This observation indicated that Glu501 is essential for the GH family 1 proteins, including the diglycosidases, to recognize the glycone of the substrate.

In GH family 1, the interaction with the aglycone and/or the whole-substrate structure has not been fully clarified, even among the monoglycosidases (*7*); for example, a single mutation is not sufficient to exchange the substrate specificity between a pair of monoglycosidases with high sequence similarity (*32*). Our study provides the first clue toward revealing how the diglycoside recognition mechanism has evolved in GH family 1. We suggest that changes occurring at two different locations were critical for the evolution of diglycosidases from monoglycosidases. However, the phylogenetic tree suggests that diglycosidase activity has been independently acquired several times during the evolution of the GH family 1 proteins (Fig. 3). The diglycosidases that evolved independently from PD and FH may have acquired the activities by different mechanisms. In fact, VH has Trp and Ala at the alignment sites 383 and 473, respectively, like the monoglycosidases. Therefore, the substrate recognition by VH is considered to be different from those of FH and PD. Further accumulation of sequence and structure data of diglycosidases will reveal the various mechanisms for the same function.

ACKNOWLEDGEMENTS

We thank Dr Narutoshi Kamiya of Osaka University for helpful discussions about the simulation study.

FUNDING

Grants-in-Aid for Scientific Research on Priority Areas 'Membrane Interface' from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

CONFLICT OF INTEREST

None declared.

REFERENCES

1. Davies, G. and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853–859
2. Zechel, D.L. and Withers, S.G. (2000) Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc. Chem. Res.* **33**, 11–18
3. Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309–316
4. Henrissat, B. and Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* **316**, 695–696
5. Keresztessy, Z., Kiss, L., and Hughes, M.A. (1994) Investigation of the active site of the cyanogenic β-D-glucosidase (linamarase) from *Manihot esculenta* Crantz (cassava). II. Identification of Glu-198 as an active site carboxylate group with acid catalytic function. *Arch. Biochem. Biophys.* **315**, 323–330
6. Rye, C.S. and Withers, S.G. (2000) Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* **4**, 573–580
7. Czjzek, M., Cicek, M., Zamboni, V., Burmeister, W.P., Bevan, D.R., Henrissat, B., and Esen, A. (2001) Crystal structure of a monocotyledon (maize ZMGlu1) β-glucosidase and a model of its complex with *p*-nitrophenyl β-D-thioglucoside. *Biochem. J.* **354**, 37–46
8. Poulton, J.E. (1990) Cyanogenesis in plants. *Plant Physiol.* **94**, 401–405
9. Hughes, M.A., Brown, K., Pancoro, A., Murray, B.S., Oxtoby, E., and Hughes, J. (1992) A molecular and biochemical analysis of the structure of the cyanogenic β-glucosidase (linamarase) from cassava (*Manihot esculenta* Cranz). *Arch. Biochem. Biophys.* **295**, 273–279
10. Zhou, J., Hartmann, S., Shepherd, B.K., and Poulton, J.E. (2002) Investigation of the microheterogeneity and aglycone specificity-conferring residues of black cherry prunasin hydrolases. *Plant Physiol.* **129**, 1252–1264
11. Imaseki, H. and Yamamoto, T. (1961) A furcatin hydrolyzing glycosidase of *Viburnum furcatum* Blume. *Arch. Biochem. Biophys.* **92**, 467–474
12. Lizotte, P.A. and Poulton, J.E. (1988) Catabolism of cyanogenic glycosides by purified vicianin hydrolase from squirrel's foot fern (*Davallia Trichomanoides* Blume). *Plant Physiol.* **86**, 322–324
13. Yasuda, T. and Nakagawa, H. (1994) Purification and characterization of the rutin-degrading enzymes in tartary buckwheat seeds. *Phytochemistry* **37**, 133–136
14. Wirth, J., Guo, W., Baumes, R., and Günata, Z. (2001) Volatile compounds released by enzymatic hydrolysis of glycoconjugates of leaves and grape berries from *Vitis*

*vinifera* Muscat of Alexandria and Shiraz cultivars. *J. Agric. Food Chem.* **49**, 2917–2923
15. Nakanishi, F., Nagasawa, Y., Kabaya, Y., Sekimoto, H., and Shimomura, K. (2005) Characterization of lucidin formation in *Rubia tinctorum* L. *Plant Physiol. Biochem.* **43**, 921–928
16. Mizutani, M., Nakanishi, H., Ema, J., Ma, S.J., Noguchi, E., Inohara-Ochiai, M., Fukuchi-Mizutani, M., Nakao, M., and Sakata, K. (2002) Cloning of β-primeverosidase from tea leaves, a key enzyme in tea aroma formation. *Plant Physiol.* **130**, 2164–2176
17. Ahn, Y.O., Mizutani, M., Saino, H., and Sakata, K. (2004) Furcatin hydrolase from *Viburnum furcatum* Blume is a novel disaccharide-specific acuminosidase in glycosyl hydrolase family 1. *J. Biol. Chem.* **279**, 23405–23414
18. Ma, S.J., Mizutani, M., Hiratake, J., Hayashi, K., Yagi, K., Watanabe, N., and Sakata, K. (2001) Substrate specificity of β-primeverosidase, a key enzyme in aroma formation during oolong tea and black tea manufacturing. *Biosci. Biotechnol. Biochem.* **65**, 2719–2729
19. Han, W., Li, Z., Zheng, Q., and Sun, C. (2006) Toward a blueprint for β-primeverosidase from the tea leaves structure/function properties: homology modeling study. *J. Theor. Comp. Chem.* **5**, 433–446
20. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402
21. Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518
22. Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425
23. Felsenstein, J. (1996) Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol.* **266**, 418–427
24. Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282
25. Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791
26. Felsenstein, J. (1993) *PHYLIP (Phylogeny Inference Package), Version 3.5c*. Department of Genetics, University of Washington, Seattle, WA
27. Adachi, J. and Hasegawa, M. (1996) *MOLPHY (Programs for Molecular Phylogenetics), Version 2.3b3*. Institute of Statistical Mathematics, Tokyo
28. Halgren, T.A. (1999) MMFF VI. MMFF94s option for energy minimization studies. MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *J. Comput. Chem*. **20**, 720–729, 730–748.
29. Ahn, Y.O., Saino, H., Mizutani, M., Shimizu, B., and Sakata, K. (2007) Vicianin hydrolase is a novel cyanogenic β-glycosidase specific to β-vicianoside (6-*O*-α-L-arabinopyranosyl-β-D-glucopyranoside) in seeds of *Vicia angustifolia*. *Plant Cell Physiol.* **48**, 938–947
30. Zouhar, J., Vevodova, J., Marek, J., Damborsky, J., Su, X.D., and Brzobohaty, B. (2001) Insights into the functional architecture of the catalytic center of a maize β-glucosidase Zm-p60.1. *Plant Physiol.* **127**, 973–985
31. Marana, S.R. (2006) Molecular basis of substrate specificity in family 1 glycoside hydrolases. *IUBMB Life* **58**, 63–73
32. Verdoucq, L., Czjzek, M., Moriniere, J., Bevan, D.R., and Esen, A. (2003) Mutational and structural analysis of aglycone specificity in maize and sorghum β-glucosidases. *J. Biol. Chem.* **278**, 25055–25062
33. Dayhoff, M.O., Schwartz, R.M., and Orcutt, B.C. (1978) *Atlas of Protein Sequence and Structure* (Dayhoff, M.O. ed.) pp. 345–358, National Biomedical Research Foundation, Washington, DC