Structural basis for substrate recognition in the *Phytolacca americana* glycosyltransferase PaGT3

Rakesh Maharjan,a‡ Yohta Fukuda,a Taisuke Nakayama,b Toru Nakayama,c Hiroki Hamada,d Shin-ichi Ozaki,e and Tsuyoshi Inoue* a

Received 1 October 2021
Accepted 24 January 2022

Edited by M. Czjzek, Station Biologique de Roscoff, France

Present address: Department of Pharmacology, School of Medicine, Case Western Reserve University, Cleveland, OH 44106, USA.

Keywords: UGT; glycosylation; polyphenols; protein crystallization; capsaicin; kaempferol; crown ethers.

PDB references: *Phytolacca americana* UGT3, complex with kaempferol and UDP-2-fluoroglucose, 7vej; complex with capsaicin and UDP-2-fluoroglucose, 7vek; complex with UDP-2-fluoroglucose, 7vel

Supporting information: this article has supporting information at journals.iucr.org/d

Capsaicinoids are phenolic compounds that have health benefits. However, the pungency and poor water solubility of these compounds limit their exploitation. Glycosylation is a powerful method to improve water solubility and reduce pungency while preserving bioactivity. *PaGT3*, a uridine diphosphate glycosyltransferase (UGT) from *Phytolacca americana*, is known for its ability to glycosylate capsaicinoids and other phenolic compounds. While structural information on several UGTs is available, structures of UGTs that can glycosylate a range of phenolic compounds are rare. To fill this gap, crystal structures of *PaGT3* with a sugar-donor analogue (UDP-2-fluoroglucose) and the acceptors capsaicin and kaempferol were determined. *PaGT3* adopts a GT-B-fold structure that is highly conserved among UGTs. However, the acceptor-binding pocket in *PaGT3* is hydrophobic and large, and is surrounded by longer loops. The larger acceptor-binding pocket in *PaGT3* allows the enzyme to bind a range of compounds, while the flexibility of the longer loops possibly plays a role in accommodating the acceptors in the binding pocket according to their shape and size. This structural information provides insights into the acceptor-binding mechanism in UGTs that bind multiple substrates.

1. Introduction

Capsaicinoids are compounds with a pungent taste that are produced by plants belonging to the genus *Capsicum*. Despite contact with capsaicinoids causing the inflammation of tissue, these compounds also provide health benefits. Capsaicin shows a cardioprotective effect through the activation of transient receptor potential vanilloid 1 (TRPV1) and inhibition of platelet aggregation (Mittelstadt *et al.*, 2012; Sharma *et al.*, 2013). Capsaicin supplements in a high-fat diet lower adipose tissue weight and serum triglyceride in rats (Kawada *et al.*, 1986). Capsaicinoids also possess antibacterial (Marini *et al.*, 2015), anti-inflammatory (Kim *et al.*, 2003), anticancer (Clark & Lee, 2016), antioxidant (Rosa *et al.*, 2002) and analgesic effects (Fusco & Alessandri, 1992). However, the pungency and poor water solubility of capsaicinoids limit their use as prodrug compounds.

Uridine diphosphate glycosyltransferases (UGTs), which are classified as family 1 glycosyltransferases, transfer sugar moieties from UDP-sugar donors to small lipophilic molecules (Lombard *et al.*, 2014). Glycosylation of lipophilic molecules improves their water solubility, membrane permeability, cellular absorption and localization, and biological half-life (Bowles *et al.*, 2005). UGTs also play a significant role in the biosynthesis of secondary metabolites and the elimination of...
xenobiotic compounds (Brazier-Hicks et al., 2007; Lim & Bowles, 2004; Radomska-Pandya et al., 2010). Besides in vivo functions, UGTs have garnered attention for the one-step enzymatic glycosylation of small lipophilic compounds, compared with the chemical glycosylation method, which requires a tedious and long process of protection/deprotection of functional groups (Shimoda et al., 2006; Dai et al., 2017). Small-molecule glycosylation provides advantages in several biotechnological applications, such as increasing the water solubility of poorly water-soluble compounds such as resveratrol (Lepak et al., 2015) and artepillin C (Shimoda et al., 2014), improving the stability of vitamin C (Muto et al., 1990), the production of indigo dye by an environmentally friendly process (Hsu et al., 2018), synthesis of the skin whitener α-arbutin from hydroquinone (Kurosu et al., 2002) and the producing of unnatural colours in flowers for decoration such as blue-coloured roses (Katsumoto et al., 2007). Accordingly, glycosylation is one of the methods that are used to improve the water solubility and decrease the pungency of capsaicinoids (Kometani et al., 1993). Capsaicinoid glycosides with improved solubility and reduced pungency show similar effects as the parent compounds and thus can find their way into preclinical trials as prodrugs.

UGTs share a conserved three-dimensional structure, known as a GT-B fold, consisting of two Rossmann-fold domains. These enzymes are characterized by the presence of a consensus plant secondary product glycosyltransferase (PSPG) motif, which contains most of the residues involved in UDP-sugar donor binding (Offen et al., 2006; Lim & Bowles, 2004). The sugar-acceptor binding pocket, including the acceptor-recognizing residues, varies significantly among different UGTs, although a His–Asp catalytic pair is highly conserved. This variation in the acceptor-binding pocket could allow different UGTs to recognize different aglycones and glycosylate at different positions (Li et al., 2007; Lairson et al., 2008). To understand the structure–function relationship, several UGT crystal structures have been determined with or without substrates. Most of these UGT structures are in complexes with flavonoid molecules, such as VvGT1 from Vitis vinifera with kaempferol/quercetin (Offen et al., 2006), UGT78G1 from Medicago truncatula with myricetin (Modolo et al., 2009) and UGT78K6 from Clitoria ternatea with delphinidin/petunidin/kaempferol (Hiromoto et al., 2015). Recently, crystal structures of some UGTs that glycosylate other phenolic compounds have also been determined in complex with the corresponding acceptor substrates, such as UGT76G1 from Stevia rebaudiana with rebaudioside A/ rubusoside (Yang et al., 2019), PaGT2 from Phytolacca americana with resveratrol/pterostilbene (Maharjan, Fukuda, Shimomura et al., 2020) and Os79UGT from Oryza sativa with trichotheccene (Wetterhorn et al., 2017). Surprisingly, some UGTs can recognize and glycosylate a range of compounds.

P. americana is a toxic plant that is native to North America. Previously, three UGTs from the plant, namely PaGT1–PaGT3, have been isolated and characterized (Noguchi et al., 2009). These UGTs could be involved in the biosynthesis of flavonoid and/or triterpene glycoside derivatives, which have been isolated from different parts of P. americana (Bylka & Materlawska, 2001; Takahashi et al., 2001; Wang et al., 2008). Among these three UGTs, PaGT3 is capable of glycosylating capsicainicids (Fig. 1; Noguchi et al., 2009). Moreover, it has been shown that capsicain can increase the expression of PaGT3 in cultured callus tissue of P. americana. Additional studies show that PaGT3 glycosylates a range of substrates.

![Figure 1](image)

**Figure 1**

(a) PaGT3 catalyzes the glycosylation of capsaicin to form capsaicin glucoside. (b) With kaempferol, PaGT3 form a mixture of kaempferol glucosides, with kaempferol 3-O-glucoside as the major product.
which includes flavonols, stilbenoids, hydroxylavones, hydroxybenzoic acids, retinol, vitamin E and its derivatives, and artepillin C (Iwakiri, Imai et al., 2013; Ozaki et al., 2012; Shimoda et al., 2006, 2014; Iwakiri, Mase et al., 2013). However, PaGT3 forms multiple monoglycosylated products, rendering low regioselectivity with compounds that have more than one glycosylation site, such as kaempferol, artepillin C and resveratrol (Noguchi et al., 2009; Shimoda et al., 2014; Ozaki et al., 2012).

Although crystal structures of the promiscuous UGTs PaGT3 (Majarhan, Fukuda, Nakayama et al., 2020) and Bs-Yjc (Dai et al., 2021) are available, these structures do not contain acceptors. Thus, these structures do not provide sufficient information to understand the acceptor-recognition mechanisms in such promiscuous UGTs. Similarly, capsicain glycosides have been enzymatically synthesized using different cultured plant cells (Shimoda et al., 2007; Katsuragi et al., 2010, 2011). However, the enzyme or UGT that transforms capsicain in these plant cell cultures is not known and structural information is not available. Thus, to shed light on the mechanism of capsicain glycosylation and the acceptor-recognition mechanism in a promiscuous UGT, we report crystal structures of PaGT3 in complex with the sugar-donor analogue uridine-2-fluoroglucose (UDP-2FGlc) at 2.20 Å resolution as well as of PaGT3 with UDP-2FGlc and capsicain at 2.60 Å resolution. We also determined the crystal structure of PaGT3 with UDP-2FGlc and kaempferol at 1.85 Å resolution to understand the poor regioselectivity in the glycosylation of acceptors with multiple possible glycosylation sites. The structure of PaGT3 with capsicain provides a mechanistic overview of the recognition of long-chain phenolic compounds in UGTs, while the structure of the kaempferol complex elaborates the poor regioselective glycosylation of phenolic compounds with multiple possible glycosylation sites.

2. Materials and methods

2.1. Protein expression and purification

The gene expressing PaGT3 was cloned and expressed and the protein was purified as described previously (Majarhan, Fukuda, Nakayama et al., 2020). Briefly, the gene encoding PaGT3 (UniProt ID B5MGN9) was amplified by polymerase chain reaction (PCR) using the forward and reverse primers † and ‡, respectively. † The start and termination codons are underlined in the forward and reverse primers, respectively. ‡ The Tobacco etch virus (TEV) protease recognition site is underlined in the forward primer.

The supernatant was loaded onto a nickel–nitrotriacetic acid (Ni–NTA) column (HisTrap HP 5 ml, GE Healthcare) and the target protein was eluted using a buffer consisting of 20 mM Tris–HCl pH 8.5, 100 mM NaCl, 5 mM DTT, 300 mM imidazole. The fraction containing PaGT3 was pooled, mixed with TEV protease and dialyzed against 20 mM Tris–HCl pH 8.5, 100 mM NaCl, 5 mM DTT. The dialysed protein was passed through an Ni–NTA column to separate PaGT3 from the His6 tag and TEV protease. The protein was further purified by cation-exchange chromatography on a HiTrap Q column (GE Healthcare) and eluted with a buffer consisting of 20 mM Tris–HCl pH 8.5, 1 M NaCl. Finally, size-exclusion chromatography (SEC) was performed on a HiLoad 16/600 Superdex 200 pg column using a buffer consisting of 20 mM Tris–HCl pH 8.0, 100 mM NaCl, 5 mM dithiothreitol. The fractions containing PaGT3 were pooled, concentrated to ~10 mg ml⁻¹ and stored at ~80°C until crystallization. Macromolecule-production information is summarized in Table 1.

2.2. Crystallization

For crystallization experiments, UDP-2FGlc was purchased from Fuji Molecular Planning, Yokohama, Japan. Crystallization screening of PaGT3 with UDP-2FGlc and/or acceptors was performed in a 96-well sitting-drop crystallization plate (Viablow) using an automatic pipetting machine (Mosquito LCP, TTP Labtech). As for apo PaGT3, the cocrystallization of PaGT3 with ligands did not form diffraction crystals without 18-crown-6 ether. For crystallization screening, 100 nl protein solution consisting of 10 mg ml⁻¹ PaGT3, 50 mM 18-crown-6 ether, 5 mM UDP-2FGlc, 2 mM capsicain

| Table 1 | Macromolecule-production information. |
| Source organism | Phytolacca americana |
| DNA source | pQE30 PaGT3 (Ozaki et al., 2012) |
| Forward primer† | 5′-CTTTATTCTCAAGGCTGGTTGCTGAAC-3′ |
| Reverse primer† | 5′-AGCACAGAATCTCTGACATGATAACC-3′ |
| Cloning vector | pQE30 |
| Expression vector | pCold I |
| Expression host | Escherichia coli BL21 (DE3) |

MNHKVHHRHHHHLQENLNPQGQAEQPQQLVHVFFPMADGMFTPFLRAFNLANVRA
TTITTPMPAHTFTKAIEMKGMKNSGPTIH
LELFKFAQDGLPECGCENLQALGSSL
IEKFKRVGGLRLEQALEAYLEKTRPMLC
ADVFFPWATODSAKENFNLPRLEFWGTSFF
SLCALEVRLHEPKNVSSEELFDLPL
FPHEIDKMRMLQLPEDVNEKHEAKTRKRL
KLKESESKLYGVSGYFLEFYLPNAEYFFKRELGRKANVPSLCSNRTESDKQAR
GKTSQIDHECLKWXLSKRSNNAVIVYICF
GSMQAIAPQYELIAKMAMGSGQEVNWV
VRNMQNDDDCDSSNLRPGFQKGVEKGG
LIRGWPQVLLEHAIGAYFVTGCGWN
STLEGITAGVMVTPHFAEQRYEKLNV
QNIKLQGVPVGKNSRETSIDQVKKKD
AIEKALRLEIVMQEDAEERGSRKDKKLM
AKRAVEEGGSGYSYDSLALEREILGNYA

† The start and termination codons are underlined in the forward and reverse primers, respectively. ‡ The Tobacco etch virus (TEV) protease recognition site is underlined.
or kaempferol was mixed with 100 nl reservoir solution. Crystals appeared overnight using a reservoir solution consisting of 0.2 M potassium acetate, 20% (w/v) PEG 3350.

For diffraction experiments, crystals of PaGT3 with UDP-2FGlc were obtained by mixing 1 μl protein solution consisting of 10 mg ml⁻¹ PaGT3, 50 mM 18-crown-6 ether, 5 mM UDP-2FGlc with 1 μl reservoir solution consisting of 0.15–0.20 M potassium acetate, 20% (w/v) PEG 3350. Crystals of PaGT3 with UDP-2FGlc and kaempferol were obtained by mixing 1 μl protein solution consisting of 10 mg ml⁻¹ PaGT3, 50 mM 18-crown-6 ether, 5 mM UDP-2FGlc, 2 mM kaempferol with 1 μl reservoir solution consisting of 0.15–0.20 M potassium acetate, 20% (w/v) PEG 3350. Although crystals of PaGT3 with UDP-2FGlc and kaempferol (2–10 mM) were obtained under similar conditions, electron density for capsaicin was not observed during structure determination. Thus, to determine the capsaicin-bound PaGT3 structure, crystals of PaGT3 with UDP-2FGlc were soaked in reservoir solution containing an excess of capsaicin before harvesting. For data collection, all crystals were harvested by soaking in a cryoprotectant solution consisting of the reservoir solution supplemented with 15% ethylene glycol. Crystallization information is summarized in Table 2.

### 2.3. Data collection and structure determination

All data sets were collected on beamline BL44XU at SPring-8, Japan. The data sets were processed with the XDS package (Kabsch, 2010) and were scaled with AIMLESS (Evans, 2011) in the CCP4 package (Winn et al., 2011). The phases for each structure were determined by molecular replacement in MOLREP (Vagin & Teplyakov, 2010) using the structure of apo PaGT3 (PDB entry 6ly2; Maharjan, Fukuda, Nakayama et al., 2020) as the search model. Coot (Emsley et al., 2010) was used for manual model building, adding substrates into the corresponding electron-density maps and adding water molecules. The polder omit map for capsaicin was calculated using phenix.polder (Liebschner et al., 2017) in Phenix (Liebschner et al., 2019). Refinement was performed in REFMAC5 (Kovalveksiyi et al., 2018) and phenix.refine. The structures were validated using MolProbity (Williams et al., 2018). Images were prepared using PyMOL (version 1.8; Schrödinger). The data-collection and refinement statistics are given in Tables 3 and 4.

### 3. Results

#### 3.1. The overall structure of PaGT3 complexes

Recombinant PaGT3 was expressed and purified to near-homogeneity for crystallization as described previously (Maharjan, Fukuda, Nakayama et al., 2020). The crystal structure of PaGT3 with the sugar-donor analogue UDP-2FGlc was determined at 2.20 Å resolution (Supplementary Fig. S1a). The ternary complexes of PaGT3 with UDP-2FGlc and the sugar acceptors capsaicin or kaempferol were refined to 2.60 and 1.85 Å resolution, respectively (Fig. 2 and Supplementary Figs. S1b and S1c). The asymmetric unit of each crystal structure of PaGT3 consists of two molecules of the enzyme linked together with an 18-crown-6 metal-ion complex, which plays the role of a molecular glue during crystallization (Maharjan, Fukuda, Nakayama et al., 2020). Data-collection and refinement statistics are given in Tables 3 and 4, respectively.

The ligand-bound PaGT3 structures are nearly identical to the apo PaGT3 structure (Supplementary Fig. S2). Structural alignment of apo PaGT3 with the complexes of PaGT3 with UDP-2FGlc, with UDP-2FGlc and capsaicin, and with UDP-2FGlc and kaempferol shows root-mean-square deviations (r.m.s.d.s) of 0.71, 0.70 and 0.89 Å, respectively, for all Cα atoms. However, closer examination shows the displacement of some loops that are present around the substrate-binding cavity. Compared with the apo-PaGT3 structure, the loop Gly78–Gly91 shifts towards the acceptor-binding pocket in the kaempferol-bound structure (Supplementary Fig. S2). Compared with the apo PaGT3 structure the same loop is seen to shift outwards in the UDP-2FGlc-bound structure, and it shifts further outwards in the capsaicin-bound structure due to the binding of the larger capsaicin molecule. The loops Cys289–Ile297 and Val412–Lys429 shift towards the pocket in the substrate-bound structures compared with the apo PaGT3 structure. These results show that PaGT3 adopts similar conformations with subtle differences to accommodate acceptors with different shapes and sizes.

![Figure 2](image_url)

Overall structure of substrate-bound PaGT3. Crystal structure of PaGT3 colour-ramped from the N-terminus (blue) to the C-terminus (red). The donor (UDP-2FGlc, yellow sticks) and acceptor (capsaicin, wheat sticks) binding sites are highlighted in transparent red and blue colours, respectively.

---

**Table 2**

| Method                  | Hanging-drop vapour diffusion |
|------------------------|-------------------------------|
| Plate type             | Sample cups and siliconized cover glasses |
| Temperature (K)        | 293                           |
| Protein concentration (mg ml⁻¹) | 10                           |
| Buffer composition of protein solution | 20 mM Tris–HCl pH 7.5, 100 mM NaCl, 5 mM DTT |
| Composition of reservoir solution | 0.15–0.20 M potassium acetate, 20% (w/v) PEG 3350 |
| Volume ratio of drop   | 1 μl:1 μl                     |
| Volume of reservoir (μl) | 400                           |
The crystallization of PaGT3 requires 18-crown-6 ether as a crystallization additive. Similar to the crystal structure of apo PaGT3, a crown ether molecule is present between the two protomers of the protein in the asymmetric unit (Supplementary Fig. S1). The crown ether cavity consists of a metal ion coordinated through the six O atoms of the crown ether and the main-chain O atoms of Glu238 from the two molecules of PaGT3. Previously, we assigned the metal ion in the crown ether cavity as a sodium ion, because the apo PaGT3 crystallization solution contained sodium bromide. PaGT3–substrate complex crystals were obtained with a mother-liquor solution containing potassium acetate. Usually, the distances between potassium and oxygen in macromolecular crystal structures are >2.7 Å, while sodium–oxygen distances are between 2.4 and 2.5 Å (Zheng et al., 2017). In the PaGT3–capsaicin crystal structure the average distances from the central metal ion to the O atoms and C atoms of the 18-crown-6 ether are 2.9 and 3.6 Å, respectively. These values are comparable to the K–O and K–C distances reported in the crystal structure of a 18-crown-6 ether–potassium ion complex (Ozutsumi et al., 1989). The distances from the central potassium ion to the O and C atoms of the crown ether in the PaGT3–capsaicin crystal structure are presented in Supplementary Table S2. Moreover, it is known that 18-crown-6 ether has a higher affinity for potassium ion than for sodium ion. Thus, we assign the electron density in the crown ether cavity present in the PaGT3 complex structures as a potassium ion, which could be from the crystallization solution.

3.2. Sugar-donor binding in PaGT3

The C-terminal domain harbours the sugar-donor binding cavity in GT-B-fold UGTs. UDP-2FGLc occupies the sugar-donor binding cavities in all three PaGT3 crystal structures (Supplementary Fig. S1). Electron density for UDP-2FGLc is present in all three PaGT3 structures (Fig. 3a). Among our three substrate-bound PaGT3 structures, the UDP-2FGLc/kaempferol-containing structure has the highest resolution. Thus, we describe the features of UDP-2FGLc binding in PaGT3 with reference to this structure. The residues that interact with UDP-2FGLc mainly come from the C-terminal domain and are shown in Fig. 3(b). Most of these sugar-donor-recognizing residues are highly conserved in the GT-B-fold UGTs and come from the consensus PSPG motif, which extends from Trp352 to Glu395 in PaGT3 (Supplementary Fig. S3). Hence, the sugar-donor binding in PaGT3 is comparable to other GT-B-fold UGTs.

Acta Cryst. (2022). D78, 379–389

Rakesh Maharjan et al. · Substrate recognition in PaGT3
to other that in known plant UGT structures and has been discussed in a previous report.

Among the residues interacting with UDP-2FGlc, the side chains of Ser292, Trp352 and His370 show different configurations when compared with the apo PaGT3 structure (Fig. 3c). Although Ser292 is outside the PSPG motif, it is seen to form hydrogen bonds with the sugar-donor analogue. The movement of Ser292 comes from movement of the Cys289–Ile297 loop in substrate-bound structures (Supplementary Fig. S2). In the substrate-bound PaGT3 structures, the indole moiety of Trp352 flips to form a \( \pi \)-stacking interaction with the uracil ring of the sugar-donor analogue. Such a \( \pi \)-stacking interaction between the uridine moiety and the corresponding Trp residue has been observed in several UGT structures (Hiromoto et al., 2015; Brazier-Hicks et al., 2007; Yang et al., 2019). Similarly, the side chain of His370 rotates to form a hydrogen bond to the O3A atom on the \( \alpha \)-phosphate moiety of UDP-2FGlc. This histidine residue is highly conserved among UGTs and plays a remarkable role in sugar-donor binding. For example, mutation of His293 in the Streptomyces antibioticus UGT OleI (Bolam et al., 2007), corresponding to His370 in PaGT3, significantly diminishes the activity of the enzyme.

In addition to the residues from the C-terminal domain, UDP-2FGlc in PaGT3 structures is also stabilized through hydrogen bonds from residues in the N-terminal domain (Fig. 3c). The side chain of His18 is likely to contribute to stabilizing the sugar moiety by forming a hydrogen bond to the 6-OH of 2-fluoroglucose. His18, Gly19 and Glu87 stabilize the sugar-donor analogue through water-mediated hydrogen bonds. Interestingly, the side chain of Glu87 makes a large movement to form a hydrogen bond to a water molecule (HOH5) in the substrate-bound structure. The movement of Glu87 is a result of the shift of the Gly78–Gly91 loop in the substrate-bound structures (Supplementary Fig. S2).

3.3. Sugar-acceptor binding in PaGT3

The ternary complexes of PaGT3 with UDP-2FGlc and aglycones were prepared by either soaking or co-crystallization methods. For the capsaicin-bound complex, co-crystals of PaGT3 and UDP-2FGlc were soaked in reservoir solution containing an excess of capsaicin, whereas PaGT3 was co-crystallized with UDP-2FGlc and kaempferol to obtain the ternary complex with kaempferol.

Among the two protomers of PaGT3 in the asymmetric unit, molecule \( A \) does not show any possible electron density for capsaicin in the acceptor-binding site. However, molecule \( B \) shows an elongated \( mF_o - DF_c \) electron density in the acceptor-binding pocket. Initially, the electron-density map

---

**Figure 3**

Interaction between PaGT3 and UDP-2FGlc. (a) Residues of PaGT3 (magenta) from the kaempferol/UDP-2FGlc-bound structure showing the sugar-donor analog stabilized by a network of hydrogen bonds. Side chains of the corresponding residues in apo PaGT3 (pink) show that the residues around UDP-2FGlc shift towards the substrate, where the side chains of Glu87, Ser292, Trp352 and His370 shows large movements. Water molecules involved in the hydrogen-bond network are shown as red spheres. The possible hydrogen bonds are indicated with dashed lines. The inset shows a \( \sigma \)-weighted \( 2F_o - F_c \) electron-density map contoured at 1\( \sigma \) for UDP-2FGlc in the PaGT3/UDP-2FGlc/kaempferol structure. (b) 2D figure showing residues that interact with UDP-2FGlc in the crystal structure.
was not clear enough to determine a capsaicin molecule (Supplementary Fig. S4a). However, this electron density is large than an ethylene glycol molecule, which was used as a cryoprotectant. We assumed that the electron density is from a bound capsaicin molecule and that the poor electron density could possibly be due to low occupancy or/and the highly flexible alkyl chain of capsaicin. Thus, we modelled a capsaicin molecule in the acceptor-binding site in PaGT3 molecule B and refined it to an occupancy of 0.8. To confirm the presence of capsaicin, we calculated a \( mF_o - DF_c \) polder omit map (Liebschner et al., 2017) in the Phenix suite, which excludes bulk solvent from the selected area to calculate the omit map (Fig. 3a). The calculated polder map confirms the occupancy of capsaicin in PaGT3 molecule B. The \( 2mF_o - DF_c \) electron-density map contoured at 1σ for capsaicin is comparable to the calculated polder map (Supplementary Fig. S4b). We also added a capsaicin molecule in the acceptor-binding pocket of molecule A and similarly calculated an \( mF_o - DF_c \) polder map for it. However, the program is unable to calculate an interpretable omit map for capsaicin in PaGT3 molecule A, suggesting the absence of capsaicin in protomer A.

According to the calculated polder map, the 10-OH group of capsaicin, which is the putative glycosylation site, faces towards the catalytic histidine His20 (Fig. 4b). The distance from His20 to the 10-OH of capsaicin is 3.3 Å. The GT-B-fold UGTs contain a conserved His–Asp catalytic pair. From the PaGT3 crystal structures as well as from comparison with other UGTs, the His20–Asp124 pair has been identified as the conserved catalytic pair in PaGT3. Moreover, the mutation of His20 to Ala or Asp has been shown to completely impair the activity of the enzyme (Ozaki et al., 2012). Another UGT from \( P. \) americana, PaGT2, has been shown to possess two catalytic histidines: the conserved catalytic histidine His18 and the alternate catalytic residue His81 (Maharjan, Fukuda, Shimomura et al., 2020). The mutation of either of the catalytic histidines in PaGT2 was compensated by another catalytic histidine, which helped to retain the catalytic activity of the enzyme. However, no such residue that can catalyse glycosylation in the absence of His20 is observed around capsaicin in PaGT3.

Within 4.5 Å, capsaicin is mainly surrounded by hydrophobic side-chain and a few polar side-chain amino acids. These residues include His20, Met125, Phe126, His145, Thr147, Leu155, Val158, His164, Leu190, Pro191, Val194, Leu206, Ala393, Glu394, Tyr397 and Trp417. The shortest distances between the atoms of these residues and the atoms of capsaicin are listed in Supplementary Table S2. Although Arg205 and Ile209 are farther from capsaicin, the side chains of these residues are also involved in formation of the acceptor-binding site. The phenolic ring of capsaicin is stacked between the side chains of Met125-Phe126 and Ala393-Glu394. A hydrophobic cavity formed by the side chains of Leu155, Val158, Arg159, His164, Leu190, Pro191, Val194, Leu206, Tyr397 and Trp417 harbours the alkyl chain of capsaicin. Although the acceptor-binding pocket is formed by numerous residues, capsaicin can only possibly form hydrogen bonds to His20 and Glu394. This suggests that capsaicin and other acceptor molecules in the acceptor-binding pocket of PaGT3 are mainly stabilized by hydrophobic interactions. As capsaicin contains a single glycosylation site, it forms only a single glycosylated product. However, compounds such as artepillin C contain two possible glycosylation sites and PaGT3 can form both artepillin C 4-β-D-glucoside and artepillin C 9-β-D-glucoside (Shimoda et al., 2014). This suggests that due to the lack of extensive hydrogen bonds between the enzyme and acceptor molecules, PaGT3 can recognize a single acceptor molecule in different binding orientations to form multiple possible products. This observation is comparable to the sugar-acceptor binding pocket in \( S. \) rebaudiana UGT76G1, which also recognizes acceptor molecules through hydrophobic interactions and recognizes a single steviol acceptor molecule in different orientations to form different products (Yang et al., 2019).

The crystal structure of PaGT3 with UDP-2FGlC and kaempferol is similar to the capsaicin-complexed structure, with an r.m.s.d. of 0.34 Å for 855 C' atoms. The electron density of kaempferol in both molecules of the enzyme clearly indicates that ring A faces towards the inner side of the acceptor-binding pocket (Figs. 5a and 5b). Superimposition of the crystal structure ofPaGT3 with UDP-2FGlC and kaempferol (Fig. 4a) suggests that due to the lack of extensive hydrogen bonds between the enzyme and acceptor molecules, PaGT3 can recognize a single acceptor molecule in different binding orientations to form multiple possible products. This observation is comparable to the sugar-acceptor binding pocket in \( S. \) rebaudiana UGT76G1, which also recognizes acceptor molecules through hydrophobic interactions and recognizes a single steviol acceptor molecule in different orientations to form different products (Yang et al., 2019).

The crystal structure of PaGT3 with UDP-2FGlC and kaempferol is similar to the capsaicin-complexed structure, with an r.m.s.d. of 0.34 Å for 855 C' atoms. The electron density of kaempferol in both molecules of the enzyme clearly indicates that ring A faces towards the inner side of the acceptor-binding pocket (Figs. 5a and 5b). Superimposition of the crystal structure of PaGT3 with UDP-2FGlC and kaempferol (Fig. 4a) suggests that due to the lack of extensive hydrogen bonds between the enzyme and acceptor molecules, PaGT3 can recognize a single acceptor molecule in different binding orientations to form multiple possible products. This observation is comparable to the sugar-acceptor binding pocket in \( S. \) rebaudiana UGT76G1, which also recognizes acceptor molecules through hydrophobic interactions and recognizes a single steviol acceptor molecule in different orientations to form different products (Yang et al., 2019).

The crystal structure of PaGT3 with UDP-2FGlC and kaempferol is similar to the capsaicin-complexed structure, with an r.m.s.d. of 0.34 Å for 855 C' atoms. The electron density of kaempferol in both molecules of the enzyme clearly indicates that ring A faces towards the inner side of the acceptor-binding pocket (Figs. 5a and 5b). Superimposition of
the two molecules of PaGT3 in the asymmetric unit shows that the structures of these two protomers are highly similar (r.m.s.d. of 0.3 Å for 375 Cα atoms). The acceptor-binding pocket in PaGT3 appears to be large for a kaempferol molecule. Thus, in addition to kaempferol, some ethylene glycol molecules from the cryoprotectant solution and water molecules are present in the acceptor-binding pocket (Supplementary Fig. S5). Superposition of the kaempferol molecules in molecules A and B of the enzyme shows that the binding positions of the two kaempferol molecules are different (Supplementary Fig. S6b). In molecule A the 5-OH of kaempferol forms a hydrogen bond to Lys210 via a water molecule (HOH114). Due to the shift of kaempferol in molecule B and the absence of a water molecule, no such hydrogen bond is observed between the 5-OH of kaempferol and Lys210 (Supplementary Fig. S5). The 3-OH of kaempferol in molecule A can form a hydrogen bond to the main-chain carbonyl O atom of His145. In molecule B, the 3-OH of kaempferol is placed farther away from His145. Similarly, the 4’-OH of kaempferol in molecule A is stabilized by hydrogen bonds to two water molecules (HOH3 and HOH18). The side chain of Arg419 in molecule A flips away from kaempferol. In molecule B, Arg419 flips towards the kaempferol molecule and occupies a position relative to the water molecule (HOH18) to form a hydrogen bond to the 4’-OH of kaempferol. The importance of Arg419 is not known; however, this residue seems to be important for binding smaller aglycones such as kaempferol for stabilization and glycosylation by PaGT3. While a water molecule (HOH5) occupies the corresponding position to HOH3, this water molecule is too distant from the 4’-OH to form a hydrogen bond. Interestingly, the 3-OH groups on kaempferol are at distances of 2.6 and 2.9 Å from the catalytic His20 in molecules A and B of PaGT3, respectively. However, the distances between 3-OH of kaempferol and the C1 atom in the 2FGlc moieties of UDP-2FGlc in molecules A and B of PaGT3 are 5.5 and 3.8 Å, respectively. As PaGT3 glycosylates kaempferol to form kaempferol 3-O-glucoside as the major product, the binding orientation of kaempferol in molecule B is likely to be a close representation of the Michaelis complex that forms the major product. The binding of kaempferol in PaGT3 protomers in the asymmetric unit is an indication that sugar-acceptor molecules can bind in different orientations in the acceptor-binding pocket. Hence, PaGT3 forms more than one glycosylated product with sugar acceptors containing more than one glycosylation site (Noguchi et al., 2009; Shimoda et al., 2014).

Although the structures of PaGT3 with capsaicin and kaempferol show high similarity, there is a slight difference in their acceptor-binding pockets. Mainly, the loops Gly78–Gly91 and Val412–Lys429 are seen to shift towards the acceptor-binding site in the kaempferol-bound structure (Supplementary Fig. S2). This could be due to differences in the binding orientations as well as in the shapes and sizes of capsaicin and kaempferol. In addition to the residues binding capsaicin, the other residues Ala17, Gly19, Glu87, Leu86, Phe99, Phe392 and Arg419 also take part in forming the acceptor-binding pocket for kaempferol. The involvement of these extra residues in the stabilization of kaempferol is evident from the movement of the loops (mentioned above) towards the active site compared with the capsaicin-bound structure (Supplementary Fig. S2).

In the kaempferol-bound structure, some extra electron density is present in the acceptor-binding pocket. We modelled this electron density as molecules of ethylene glycol, which was used as a cryoprotectant (Supplementary Fig. S6a). The structural alignment of capsaicin- and kaempferol-bound PaGT3 structures shows that one of the ethylene glycol molecules aligns with the alkyl chain of capsaicin (Supplementary Fig. S6c). Similarly, the space occupied by the terminal methyl groups of capsaicin is occupied by two water molecules in the kaempferol-bound structure. The putative

Figure 5
Kaempferol binding in PaGT3. σ-Weighted 2mFo – DFo electron-density maps (blue mesh) contoured at 1σ for kaempferol in (a) molecule A (cyan sticks) and (b) molecule B (salmon sticks) in the asymmetric unit of the crystal structure of PaGT3. The distances from kaempferol to nearby residues of PaGT3, the C1 carbon of UDP-2FGlc and water molecules (red spheres) are indicated.
10-OH glycosylation site of capsaicin also aligns with the 3-OH of kaempferol in PaGT3 molecule B, although the 10-OH of capsaicin is a little away from the catalytic residue His20. This indicates that the orientation of capsaicin calculated by the polder maps is in good agreement. Thus, the capsaicin-bound PaGT3 structure shows that the enzyme can bind and catalyse the glycosylation of large phenolic compounds. However, due to the large acceptor-binding pocket and the lack of residues that can stabilize acceptor molecules with hydrogen bonds, PaGT3 shows relatively low glycosylation activity towards smaller molecules such as salicyl acid, trans-p-coumaric acid and m-hydroxybenzoic acid compared with the larger capsaicin or kaempferol molecules (Noguchi et al., 2009).

3.4. Catalytic mechanism of PaGT3

Similar to other plant UGTs, PaGT3 is an inverting glycosyltransferase that belongs to the GT1 family in the Carbohydrate Active Enzymes database. In these UGTs, glycosylation is catalysed by a conserved His–Asp pair in the active site. The highly conserved histidine residue acts as a catalytic base to remove the proton from the glycosylation site on the acceptor and the aspartate is thought to stabilize the protonated catalytic histidine. The generated nucleophilic acceptor then attacks the C1 carbon of the UDP-sugar to form the product, with displacement of UDP.

In the crystal structures of PaGT3 the putative glycosylation sites of capsaicin and kaempferol are close to His20–Asp124 pair. Structural and amino-acid sequence alignment show His20–Asp119 in VvGT1, His18–Asp115 in PaGT2, His25–Asp124 in UGT76G1 and His17–Asp114 in UGT78K6 to occupy equivalent positions. In the PaGT3–capsaicin structure the 10-OH group of capsaicin is about 3.3 and 4.5 Å away from the N atom of His20 and the C1 atom of UDP-2FGlc, respectively. Similarly, the 3-OH group of kaempferol in molecule B of the PaGT3–kaempferol crystal structure is about 2.8 and 3.8 Å away from His20 and the C1 carbon of UDP-2FGlc, respectively. In a previous study, His20Ala mutant PaGT3 failed to form glycosylated products. These results suggest His20 to be the catalytic base that abstracts a proton from the acceptor molecule to generate a nucleophile, which then attacks the C1 atom of UDP-glucose to form products. Asp119 is the possible catalytic pair which stabilizes the protonated His20.

3.5. Comparison of sugar-acceptor binding in PaGT3 with that in other plant UGTs

PaGT3 and its isoenzyme PaGT2, which are both UGTs, share the GT-B-fold structure; however, superimposition of the crystal structures of these two UGTs shows many differences (r.m.s.d. of 0.92 for 139 Cα atoms; Supplementary Fig. S7a). The loops Phe70–Gly91 (Supplementary Fig. S7a, blue box) and Tyr161–Ala200 (Supplementary Fig. S7a, black box) around the acceptor-binding pocket in PaGT3 shift outwards compared with the corresponding loops in PaGT2. Similarly, a loop in the C-terminal domain, Pro411–Lys429 (Supplementary Fig. S7a, red box), is longer in PaGT3 than in PaGT2. This loop extends up to the opening of the acceptor-binding pocket of PaGT3, while the corresponding loop in PaGT2 is much shorter. As a consequence, the acceptor-binding pocket in PaGT3 is much wider than that in PaGT2, which allows the binding of acceptors in different orientations in PaGT3 compared with PaGT2 to form multiple glycosylated products. This assumption correlates with the previous observation of lower product regioselectivity in PaGT3 compared with PaGT2 using the same compounds (Noguchi et al., 2009). Also, the wider acceptor-binding pocket in PaGT3 enables the enzyme to glycosylate larger molecules such as capsaicin and betanidin which are not glycosylated by PaGT2. On the other hand, the smaller acceptor-binding pocket in PaGT2 could be a reason why the enzyme is able to glycosylate smaller molecules such as p-hydroxybenzoic acid and hydroquinone that are not glycosylated by PaGT3.

Although the acceptor-binding pockets in plant UGTs are usually hydrophobic, sugar acceptors are also stabilized by hydrogen bonds between the acceptors and enzyme residues (Supplementary Fig. S8). In VvGT1 from V. vinifera, Ser18 forms a hydrogen bond to the O4 atom of quercetin (Offen et al., 2006). Similarly, Gln84, His150 and Gln188 form hydrogen bonds to the hydroxyl groups on C7, C3’ and C4’, respectively. In UGT78K6 from C. ternatea, Asp367, Asp181 and the main-chain carbonyl O atom of Pro78 stabilize kaempferol, forming hydrogen bonds to the C5, C7 and C4’ hydroxyl groups of the acceptor molecule, respectively (Hiromoto et al., 2015). Also, in PaGT2 it has been shown that the 3-OH and 4’-OH groups of the piceatannol molecule are stabilized by hydrogen bonds to His81 and Glu82, respectively (Maharjan, Fukuda, Shimomura et al., 2020). However, in PaGT3 the acceptor capsaicin is mainly stabilized by hydrophobic interactions and the only possible hydrogen bond is observed to the catalytic His20. On the other hand, the binding of kaempferol is assisted by a network of hydrogen bonds provided by water molecules, which could be due to the smaller size of kaempferol compared with the size of the acceptor-binding pocket.

Comparison of the PaGT3 structure with some other UGTs (Supplementary Fig. S7) shows that the Pro411–Lys429 loop is much longer in PaGT3 than in other UGTs. The Pro411–Lys429 loop in PaGT3 extends to the opening of the acceptor-binding pocket. In other UGTs the corresponding loops are comparatively short and do not extend to the acceptor-binding pocket. Therefore, we assume that the Pro411–Lys429 loop in PaGT3 could play a role in modulating acceptor recognition in PaGT3.

Polyphenols are plant secondary metabolites that have important roles in plant growth and in ensuring their survival in the environment (Kuhn et al., 2016). Capsaicin, a pungent compound produced by plants of the genus Capsicum, behaves as an allelochemical. Capsaicin is likely to be involved in inhibition of the germination of other competing plants (Kato-Noguchi & Tanaka, 2003) and is lethal to certain insects (Ahn et al., 2011). Although humans have long been using capsaicin as a component of spices, recent studies have shown that capsaicin has cardioprotective, antibacterial, anti-
inflammatory, anticancer and antioxidant functions. Similarly, kaempferol, along with other flavonoids, is known to protect plants with its antioxidant properties (Shimoji & Yamasaki, 2005) as well as to induce pollen-specific gene products (Pourcel & Grotewold, 2009). Kaempferol and its glucosides are also known to have various health benefits such as the prevention of cancer and cardiovascular diseases and to have neuroprotective, antidiabetic, antimicrobial and anti-inflammatory activities (Calderón-Montaño et al., 2011). Due to the poor water solubility of polyphenols, including capsaicin and kaempferol, it is difficult to administer the amounts of these compounds that are required to have a visible effect. Glycosylation is one of the methods that are used to overcome such problems and the utilization of plant UGTs to glycosylate polyphenols is an economic and environmentally friendly process.

The crystal structure of PaGT3 with capsaicin and UDP-2FGlc provides insight into the capsaicin recognition and glycosylation mechanism of PaGT3. PaGT3 can also glycosylate other long-chain phenolic compounds such as retinol and vitamin E derivatives (Shimoda et al., 2006), the structures of which are comparable with that of capsaicin. Thus, we assume that PaGT3 utilizes a similar mechanism for the recognition and glycosylation of these compounds as for capsaicin. The crystal structure of PaGT3 with kaempferol and UDP-2FGlc shows that smaller molecules can bind in different positions/conformations due to the large acceptor-binding pocket of the enzyme. The low regioselectivity of glycosylated products with more than one possible glycosylation site could be due to the binding of such molecules in different product-forming conformations. Overall, our crystal structures could be useful to understand the acceptor-recognition mechanism in promiscuous plant UGTs.

Acknowledgements

The authors wish to thank Dr K. Fujimoto, Fuji Molecular Planning for the synthesis of UDP-2FGlc. This work was performed using synchrotron beamline BL44XU at SPring-8. Harima, Japan under the Collaborative Research Program of the Institute for Protein Research, Osaka University. Diffraction experiments were performed on the Osaka University beamline BL44XU at SPring-8 (Proposal Nos. 2018A6844, 2018B6844, 2019A6943 and 2019B6943). The authors would like to acknowledge the beamline staff at beamline BL44XU for their support during data collection. The authors declare no conflicts of interest. Author contributions were as follows. Rakesh Maharjan designed the project, prepared proteins, performed experiments and data analysis, and wrote the paper. Yohta Fukuda performed experiments and data analysis, and wrote the paper. Taisuke Nakayama prepared proteins and performed experiments. Toru Nakayama prepared proteins. Hiroki Hamada designed the project and wrote the paper. Shin-ichi Ozaki designed the project and wrote the paper. Tsuyoshi Inoue supervised the project.

Funding information

This work was partly supported by Grant-in-Aid for Young Scientists 20K15971 to YF, Grant-in-Aid for Scientific Research (B) 18H02004 to TI and Grant-in-Aid for Scientific Research on Innovative Areas (C) 20K05708 to SO.

References

Ahn, S.-J., Badenes-Pérez, F. R. & Heckel, D. G. (2011). J. Insect Physiol. 57, 1212–1219.
Bolam, D. N., Roberts, S., Proctor, M. R., Turkenburg, J. P., Dodson, E. J., Martinez-Fleites, C., Yang, M., Davis, B. G., Davies, G. J. & Gilbert, H. J. (2007). Proc. Natl Acad. Sci. USA, 104, 5336–5341.
Bowles, D., Isayenko, J., Lim, E.-K. & Poppenberger, B. (2005). Curr. Opin. Plant Biol. 8, 254–263.
Brazier-Hicks, M., Offen, W. A., Gershater, M. C., Revett, T. J., Lim, E.-K., Bowles, D. J., Davies, G. J. & Edwards, R. (2007). Proc. Natl Acad. Sci. USA, 104, 20238–20243.
Bylka, W. & Materławka, I. (2001). Acta Pol. Pharm. 58, 69–72.
Calderón-Montaño, J. M., Burgos-Morón, E., Pérez-Guerrero, C. & López-Lázaro, M. (2011). Mini Rev. Med. Chem. 11, 298–344.
Clark, R. & Lee, S.-H. (2016). Anticancer Res. 36, 837–843.
Dai, L., Li, J., Yao, P., Zhu, Y., Men, Y., Zeng, Y., Yang, J. & Sun, Y. (2017). J. Biotechnol. 248, 69–76.
Dai, L., Qin, L., Hu, Y., Huang, J.-W., Hu, Z., Min, J., Sun, Y. & Guo, R.-T. (2021). Biochem. Biophys. Res. Commun. 534, 73–78.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Acta Cryst. D66, 486–501.
Evans, P. R. (2011). Acta Cryst. D67, 282–292.
Fusco, B. M. & Alessandri, M. (1992). Anesth. Analg. 74, 375–377.
Hasegawa, T., Honjo, E., Noda, N., Tamada, T., Kazuma, K., Suzuki, M., Blaber, M. & Kuroki, R. (2015). Protein Sci. 24, 395–407.
Hsu, T. M., Weilner, D. H., Russ, Z. N., Cervantes, B., Prathuril, R. L., Adams, P. D. & Dueber, J. E. (2018). Nat. Chem. Biol. 14, 256–261.
Iwakiri, T., Imai, H., Hamada, H., Nakayama, T. & Ozaki, S.-I. (2013). Nat. Prod. Commun. 8, 119–120.
Iwakiri, T., Mase, S., Murakami, T., Matsumoto, M., Hamada, H., Nakayama, T. & Ozaki, S.-I. (2013). J. Mol. Catal. B Enzym. 90, 61–65.
Kabsch, W. (2010). Acta Cryst. D66, 125–132.
Kato-Noguchi, H. & Tanaka, Y. (2003). Biol. Plant. 47, 157–159.
Katsumoto, Y., Fukuchi-Mizutani, M., Fukui, Y., Brugliera, F., Holton, T. A., Karan, M., Nakamura, N., Yonekura-Sakakibara, K., Togami, J., Pigearle, A., Tao, G.-Q., Nehra, N. S., Lu, C.-Y., Dyson, B. K., Tsuda, S., Asahikari, T., Kusumi, T., Mason, J. G. & Tanaka, Y. (2007). Plant Cell Physiol. 48, 1589–1600.
Katsuragi, H., Shimoda, K., Ohira, A. & Hamada, H. (2010). Acta Biol. Hung. 61, 449–456.
Katsuragi, H., Shimoda, K., Yamamoto, R., Ishihara, K. & Hamada, H. (2011). Biochem Insights, 4, BCI.S6682.
Kawada, T., Hagihiara, K. & Iwai, K. (1986). J. Nutr. 116, 1272–1278.
Kim, C.-S., Kawada, T., Kim, B.-S., Han, I.-S., Choe, S.-Y., Kurata, T. & Yu, R. (2003). Cell. Signal. 15, 299–306.
Kometani, T., Tanimoto, H., Nishimura, T., Kanbara, I. & Okada, S. (1993). Biosci. Biotechnol. Biochem. 57, 2192–2193.
Kovalcovskiy, O., Nicholls, R. A., Long, F., Carlson, A. & Murshedov, G. N. (2018). Acta Cryst. D74, 215–227.
Kuhn, B. M., Errafi, S., Bucher, R., Dobrev, P., Geisler, M., Bigler, L., Gázamalová, E. & Ringli, C. (2016). J. Biol. Chem. 291, 5385–5395.
Kurosu, J., Sato, T., Yoshiida, K., Tsugane, T., Shimura, S., Kirimura, K., Kino, K. & Usami, S. (2002). J. Biol. Chem. 277, 328–330.
LaRison, L. L., Henrisat, B., Davies, G. J. & Withers, S. G. (2008). Annu. Rev. Biochem. 77, 521–555.
Lepak, A., Gutmann, A., Kulmer, S. T. & Nidecky, B. (2015). ChemBioChem, 16, 1870–1874.
Li, L., Modolo, L. V., Escamilla-Trevino, L. L., Achnine, L., Dixon, R. A. & Wang, X. (2007). J. Mol. Biol. 370, 951–963.
