Supporting Information for

An ambidextrous polyphenol glycosyltransferase *Pa*GT2 from *Phytolacca americana*

Rakesh Maharjan,[†,⊥] Yohta Fukuda,[†,⊥] Naomichi Shimomura,[‡] Taisuke Nakayama,[§] Yuta Okimoto,[‡] Koki Kawakami,[∥] Toru Nakayama,[#] Hiroki Hamada,*[,∥] Tsuyoshi Inoue,*[,†,⊥] Shin-ichi Ozaki*[,‡]

[†]Department of Applied Chemistry, Graduate School of Engineering, Osaka University, Suita City, Osaka 565-0871, Japan

[⊥]Graduate School of Pharmaceutical Science, Osaka University, Suita City, Osaka 565-0871, Japan

[‡]Department of Biological Chemistry, Graduate School of Science and Technology for Innovations, Yamaguchi University, Yamaguchi 753-8515, Japan

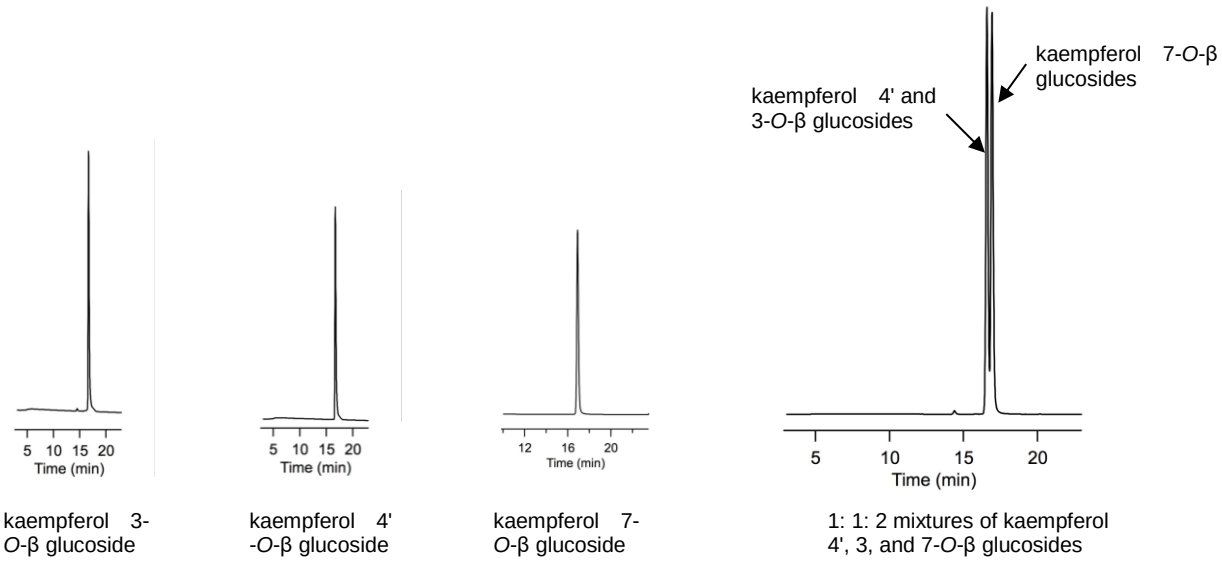
[§]National Institute of Biomedical Innovation, Health and Nutrition, Saito-Asagi, Ibaraki City, Osaka 567-0085, Japan

[#]Department of Biomolecular Engineering, Graduate School of Engineering, Tohoku University, Sendai, Miyagi 980-8579, Japan

[∥]Department of Life Science, Faculty of Science, Okayama University of Science, Okayama 700-0005, Japan

Correspondence to: hamada@dls.ous.ac.jp
t_inoue@phs.osaka-u.ac.jp
ozakis@yamaguchi-u.ac.jp

A



kaempferol 3-*O*-β glucoside

kaempferol 4'-*O*-β glucoside

kaempferol 7-*O*-β glucoside

1: 1: 2 mixtures of kaempferol 4', 3, and 7-*O*-β glucosides

B

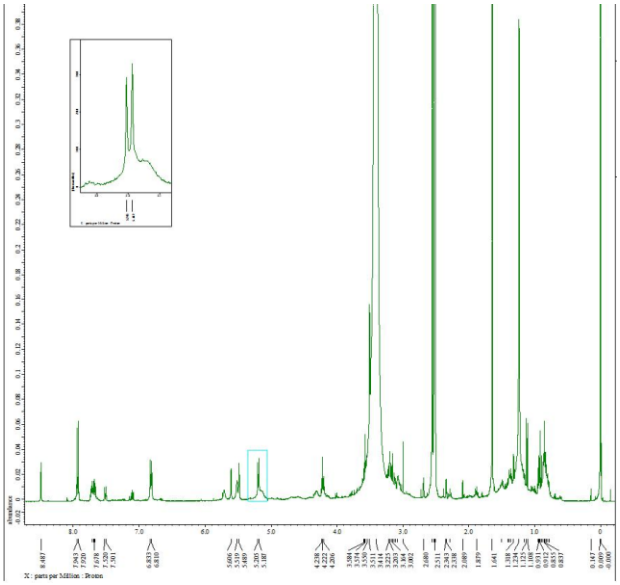


Figure S1. A. HPLC chromatogram of standard kaempferol 4'-*O*-β-glucoside, kaempferol 3-*O*-β-glucoside, kaempferol 7-*O*-β-glucoside, and mixture of all three glucosides as indicated. B. NMR spectra of the kaempferol glucoside product formed by WT *Pa*GT2. The peak at δ5.25 from the hydrogen at C1 position of glucose moiety confirms the product as kaempferol 3-*O*-β-glucoside.[2]

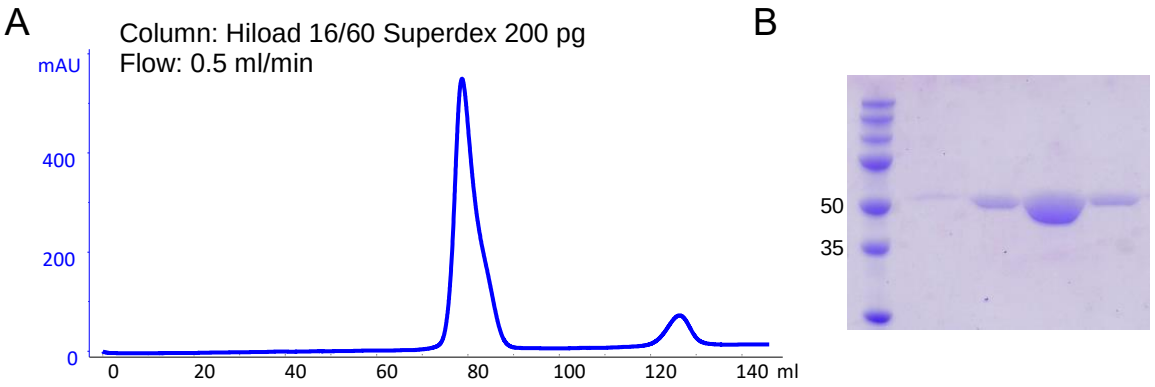


Figure S2. A. The size exclusion chromatogram of *Pa*GT2 indicate that *Pa*GT2 exist as monomer in solution. B. SDS-PAGE of *Pa*GT2 after SEC. Theoretical molecular mass of purified *Pa*GT2 is 51.3 kDa.

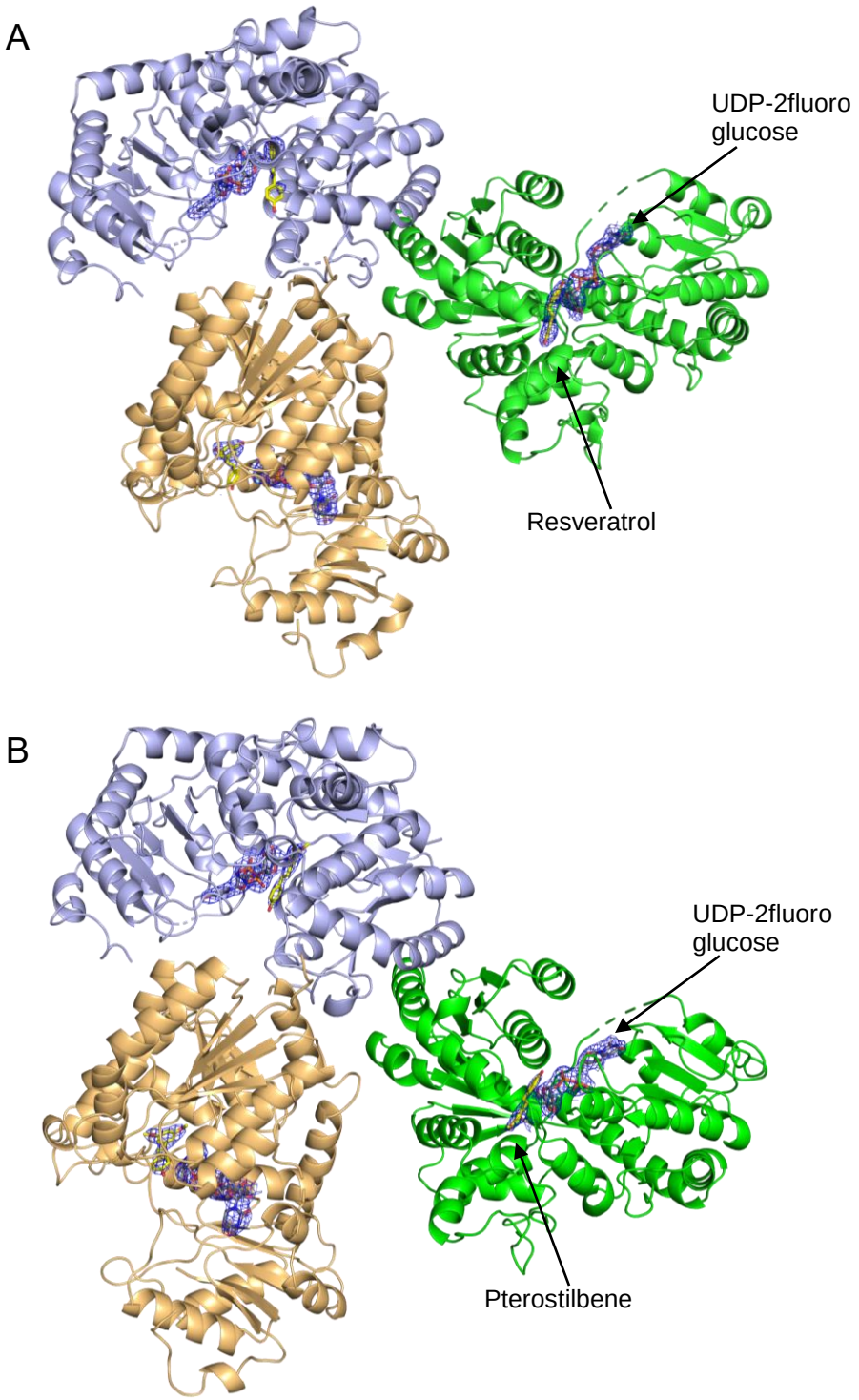
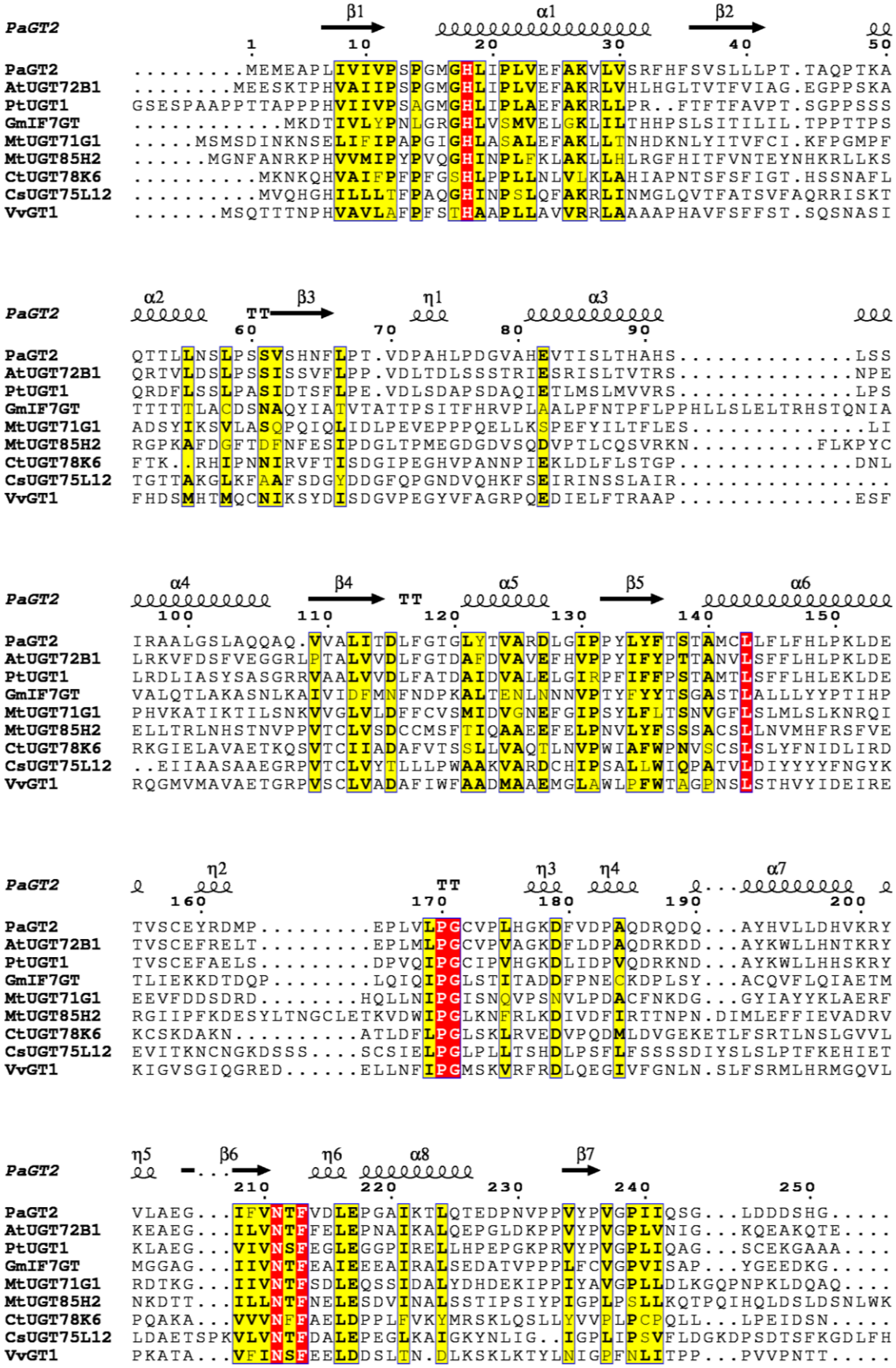


Figure S3. Molecules of *Pa*GT2 in the asymmetric unit of the crystal structures with UDP-2FGlc and (A) resveratrol and (B) pterostilbene. Each molecule in the asymmetric unit are separately colored. Sigma-A-weighted $2F_o$-F_c electron density maps contoured at 1σ for each substrate is shown as blue maps.


PaGT2
β1
α1
β2
1 10 20 30 40 50
PaGT2 .....MEMEAPLIVIVPSPGMGHLIPLVEFAKVLVSRFHFSVSLLLPT.TAQPTKA
AtUGT72B1 .....MEESKTPHVAIIPSPGMGHLIPLVEFAKRLVHLHGLTVTFVIAG.EGPPSKA
PtUGT1 GSESPAAPPTTAPPPHVIIVPSAGMGHLIPLAEFAKRLLPR..FTFTFAVPT.SGPPSSS
GmIF7GT .....MKDTIVLYPNLGRGHLVSMVELGKLILTHHPSLSITILIL.TPPTTPS
MtUGT71G1 .....MSMSDINKNSELIFIPAPGIGHLASALEFAKLLTNHDKNLYITVFCI.KFPGMPF
MtUGT85H2 ......MGNFANRKPHVVMIPYPVQGHINPLFKLAKLLHLRGFHITFVNTEYNHKRLLKS
CtUGT78K6 .....MKNKQHVAIFPFPFGSHLPLLNLVLKLAHIAPNTSFSFIGT.HSSNAFL
CsUGT75L12 .....MVQHGHILLLTFPAQGHINPSLQFAKRLINMGLQVTFATSVFAQRRISKT
VvGT1 .....MSQTTTNPHVAVLAFPFSTHAAPLLAVVRRLAAAAPHAVFSFFST.SQSNASI
α2
β3
η1
α3
60 70 80 90
PaGT2 QTTLLNSLPSSVSHNFLPT.VDPAHLPDGVAHEVTISLTHAHS.....LSS
AtUGT72B1 QRTVLDSLPSSISSVFLPP.VDLTDLSSSTRIESRISLTVTRS.....NPE
PtUGT1 QRDFLSSLPASIDTSFLPE.VDLSDAPSDAQIETLMSLMVVRS.....LPS
GmIF7GT TTTTTLACDSNAQYIATVTATTPSITFHRVPLAALPFNTPFLPPHLLSLELTRHSTQNIA
MtUGT71G1 ADSYIKSVLASQPQIQLIDLPEVEPPPQELLKSPEFYILTFLES.....LI
MtUGT85H2 RGPKAFDGFTDFNFESIPDGLTPMEGDGDVSQDVPTLCQSVRKN.....FLKPYC
CtUGT78K6 FTK..RHIPNNIRVFTISDGIPEGHVPANNPIEKLDLFLSTGP.....DNL
CsUGT75L12 TGTTAKGLKFAAFSDGYDDGFQPGNDVQHKFSEIRINSSLAIR.....
VvGT1 FHDSMHTMQCNIKSYDISDGVPEGYVFAGRPQEDIELFTRAAP.....ESF
α4
β4
α5
β5
α6
100 110 120 130 140 150
PaGT2 IRAALGSLAQQAQ.VVALITDLFGTGLYTVARDLGIPPYLYFTSTAMCLLFLFHLPKLDE
AtUGT72B1 LRKVFDSFVEGGRLPTALVVDLFGTDAFDVAVEFHVPPYIFYPTTANVLSFFLHLPKLDE
PtUGT1 LRDLIASYSASGRRVAALVVDLFATDAIDVALELGIRPFIFFPSTAMTLSFFLHLEKLDE
GmIF7GT VALQTLAKASNLKAIVIDFMNFNDPKALTENLNNNVPTYFYYTSGASTLALLLYYPTIHP
MtUGT71G1 PHVKATIKTILSNKVVGLVLDFFCVSMIDVGNEFGIPSYLFLTSNVGFLSLMLSLKNRQI
MtUGT85H2 ELLTRLNHSTNVPPVTCLVSDCCMSFTIQAAEEFELPNVLYFSSSACSLLNVMHFRSFVE
CtUGT78K6 RKGIELAVAETKQSVTCIIADAFVTSSLLVAQTLNVPWIAFWPNVSCSLSLYFNIDLIRD
CsUGT75L12 ..EIIAASAAEGRPVTCLVYTLLLPWAAKVARDCHIPSALLWIQPATVLDIYYYYFNGYK
VvGT1 RQGMVMAVAETGRPVSCLVADAFIWFAADMAAEMGLAWLPFWTAGPNSLSTHVYIDEIRE
η2
η3
η4
α7
160 170 180 190 200
PaGT2 TVSCEYRDMP.....EPLVLPGCVPLHGKDFVDPAQDRQDQ...AYHVLLDHVKRY
AtUGT72B1 TVSCEFRELT.....EPLMLPGCVPVAGKDFLDPAQDRKDD...AYKWLLHNTKRY
PtUGT1 TVSCEFAELS.....DPVQIPGCIPVHGKDLIDPVQDRKND...AYKWLLHHSKRY
GmIF7GT TLIEKKDTDQP.....LQIQIPGLSTITADDFPNECKDPLSY...ACQVFLQIAETM
MtUGT71G1 EEVFDDSDRD.....HQLLNIPGISNQVPSNVLPDACFNKDG...GYIAYYKLAERF
MtUGT85H2 RGIIPFKDESYLTNGCLETKVDWIPGLKNFRLKDIVDFIRTTNPN.DIMLEFFIEVADRV
CtUGT78K6 KCSKDAKN.....ATLDFLPGLSKLRVEDVPQDMLDVGEKETLFSRTLNSLGVVL
CsUGT75L12 EVITKNCNGKDSSS....SCSIELPGLPLLTSHDLPSFLFSSSSDIYSLSLPTFKEHIET
VvGT1 KIGVSGIQGRED.....ELLNFIPGMSKVRFRDLQEGIVFGNLN.SLFSRMLHRMGQVL
η5
β6
η6
α8
β7
210 220 230 240 250
PaGT2 VLAEG...IFVNTFVDLEPGAIKTLQTEDPNVPPVYPVGPIIQSG...LDDDSHG.....
AtUGT72B1 KEAEG...ILVNTFFELEPNAIKALQEPGLDKPPVYPVGPLVNIG...KQEAKQTE....
PtUGT1 KLAEG...VIVNSFEGLEGGPIRELLHPEPGKPRVYPVGPLIQAG...SCEKGAAA....
GmIF7GT MGGAG...IIVNTFEAIEEEAIRALSEDATVPPPLFCVGPVISAP...YGEEDKG.....
MtUGT71G1 RDTKG...IIVNTFSDLEQSSIDALYDHDEKIPPIYAVGPLLDLKGQPNPKLDQAQ....
MtUGT85H2 NKDTT...ILLNTFNELESDVINALSSTIPSIYPIGPLPSLLKQTPQIHQLDSLDSNLWK
CtUGT78K6 PQAKA...VVVNFFAELDPPLFVKYMRSKLQSLLYVVPLPCPQLL...LPEIDSN.....
CsUGT75L12 LDAETSPKVLVNTFDALEPEGLKAIGKYNLIG..IGPLIPSVFLDGKDPSDTSFKGDLFH
VvGT1 PKATA...VFINSFEELDDSLTN.DLKSKLKTYLNIGPFNLITPP...PVVPNTT.....

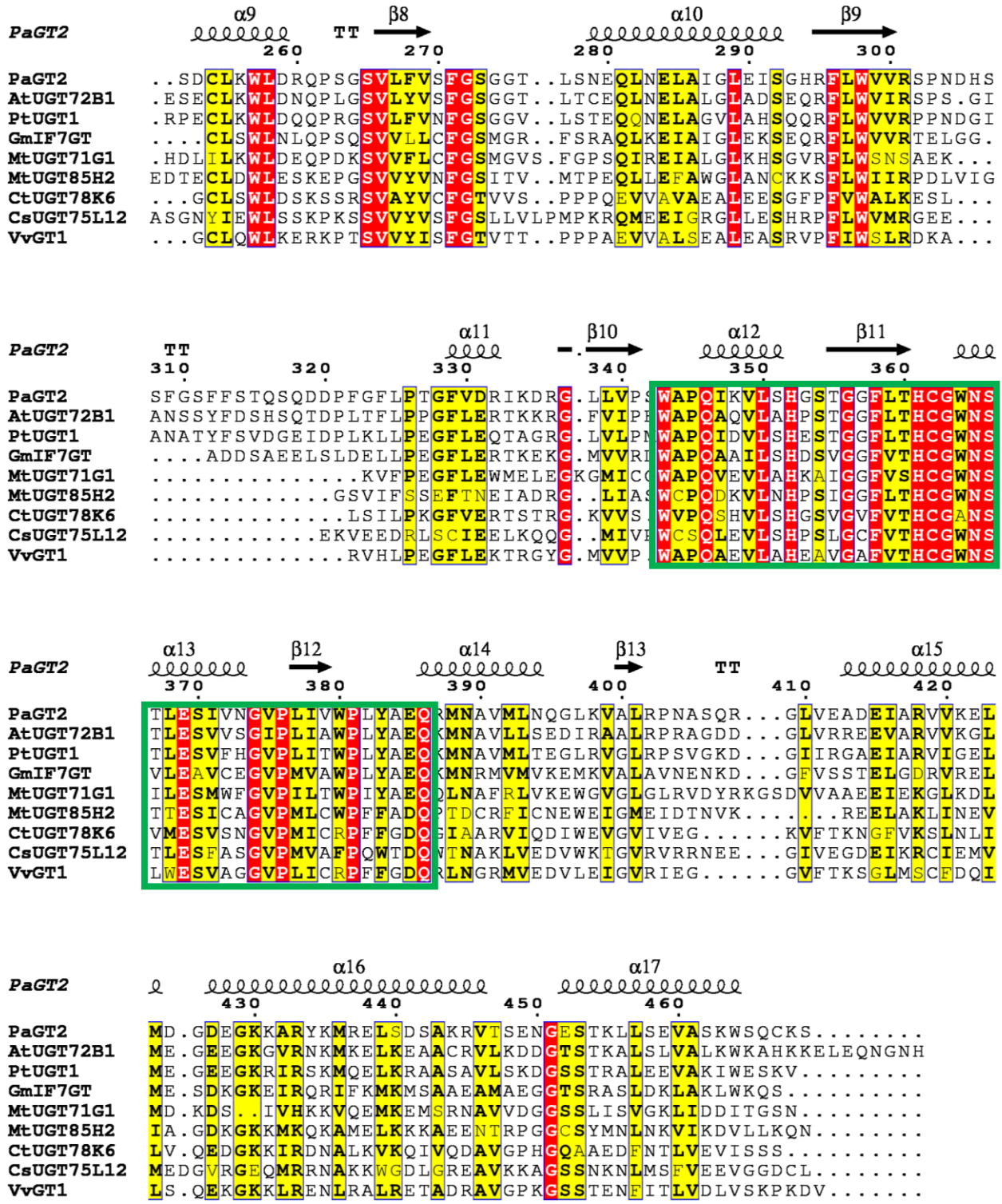


Figure S4. *Pa*GT2 is aligned with other plant UGTs with known structures. The UGTs are placed in the descending order of sequence similarity with *Pa*GT2. The conserved catalytic pair is indicated with red arrows, His81 is indicated with blue arrow, and PSPG motif is indicated with green box.

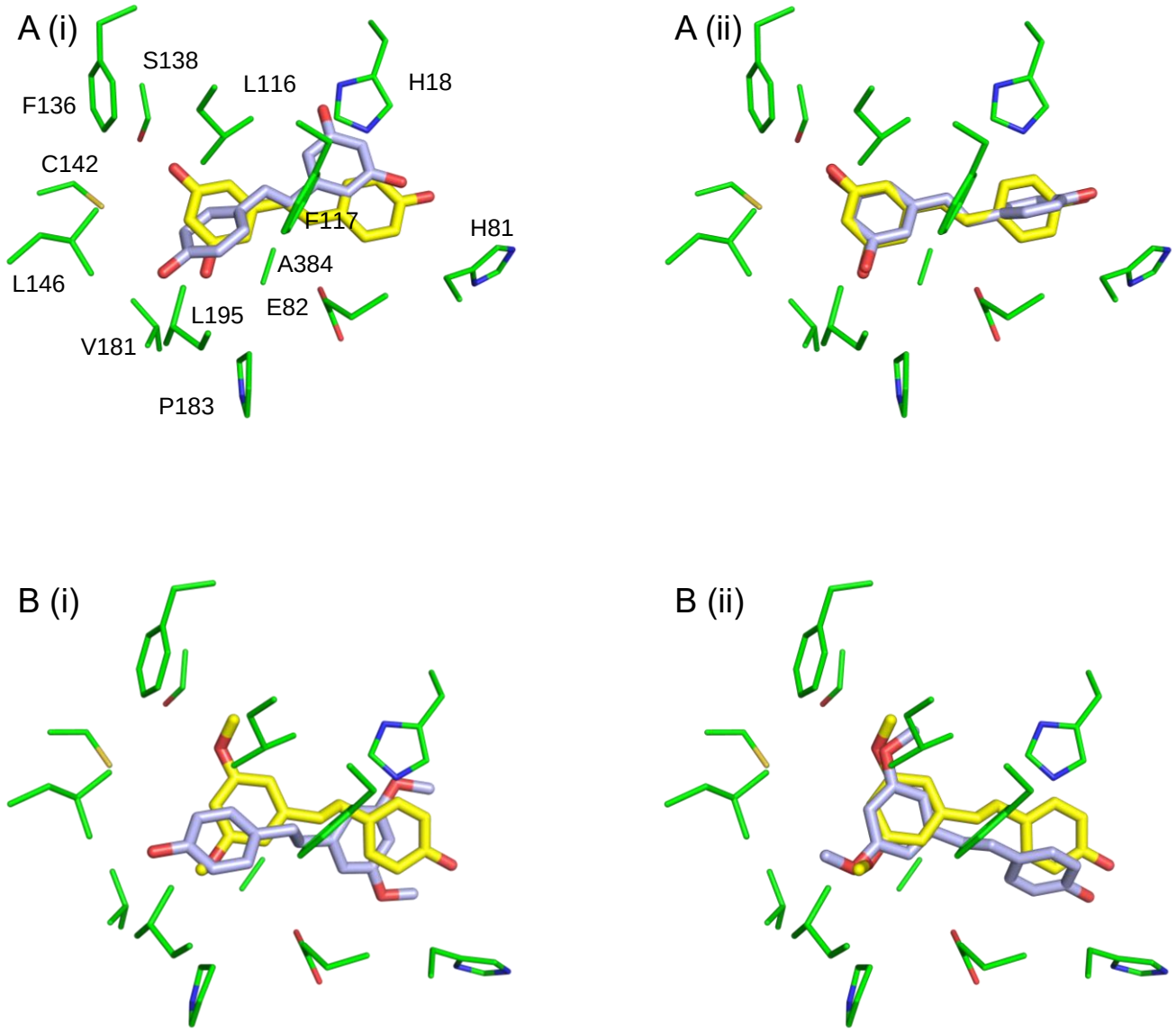


Figure S5: Molecular docking of resveratrol and pterostilbene. Comparison of the substrates in crystal structure (yellow) and from molecular docking (light blue) A (i), (ii). Resveratrol (binding energy: -6.0 kcal/mol and -5.2 kcal/mol, respectively) and B (i), (ii). Pterostilbene (-6.2 kcal/mol and -5.9 kcal/mol, respectively). The molecular docking results shows possibility of binding of substrates in opposite orientations as well as those seen in the crystal structures. Residues of *Pa*GT2 are shown in green. Residues are labelled only in fig A (i).

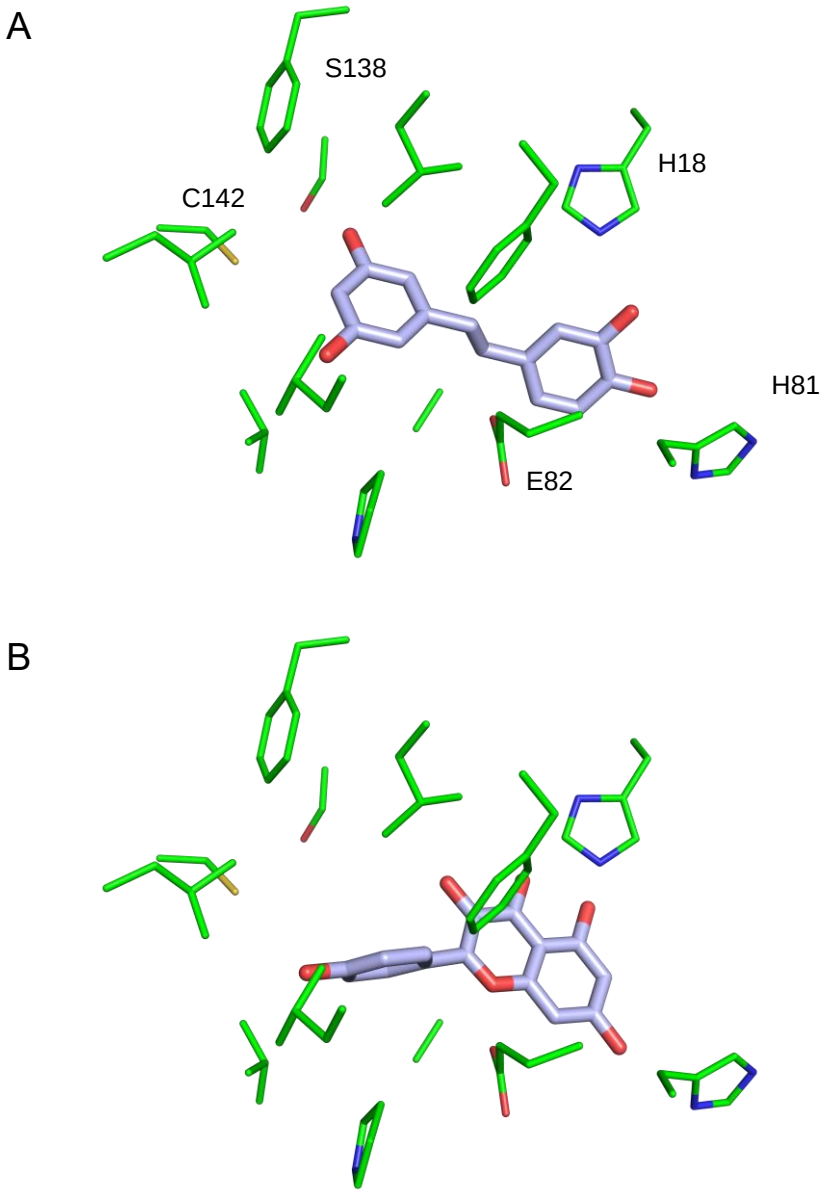


Figure S6: Molecular docking of piceatannol and kaempferol. The molecular docking of A. piceatannol (binding energy: -5.9 kcal/mol) and B. kaempferol (-7.0 kcal/mol) also predicts the binding of substrates in different orientations. The binding orientation as shown in figure A could be the suitable orientation for the production of kaempferol 5-*O*-glucoside and kaempferol 7-*O*-glucoside by the mutant enzymes.

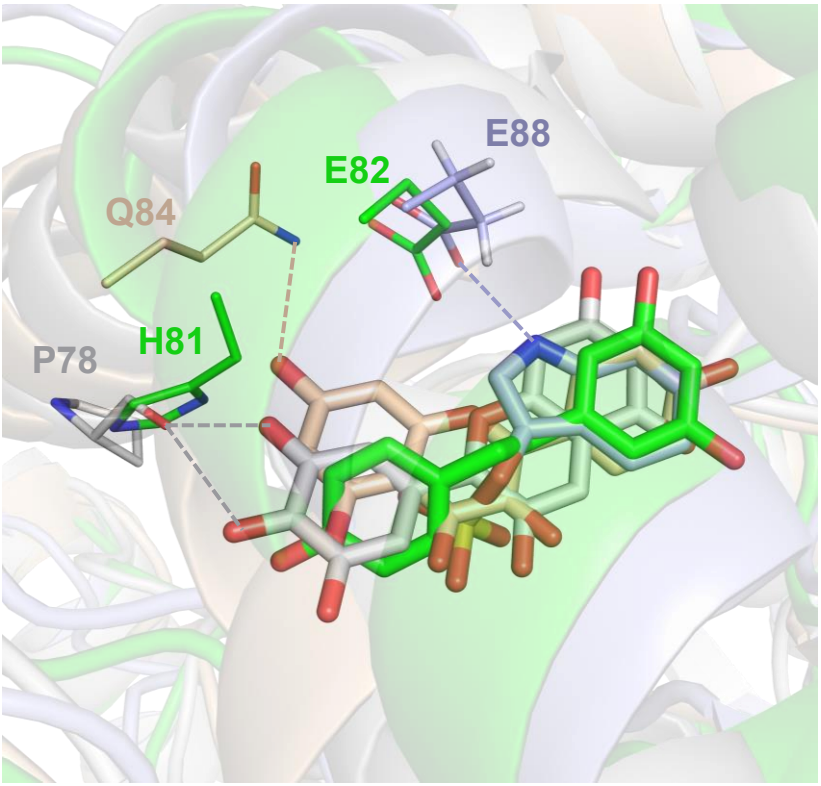


Figure S7. H81 and E82 in *Pa*GT2 (green) involve in acceptor stabilization similar to stabilization of substrates by E88 in *Pt*UGT1 (light blue), Q84 in *Vv*GT1, and P78 *Ct*UGT78K6 (light grey). Hydrogen bonds formed by these residues with respective substrates are shown by dotted lines using same color used for indicating the proteins.

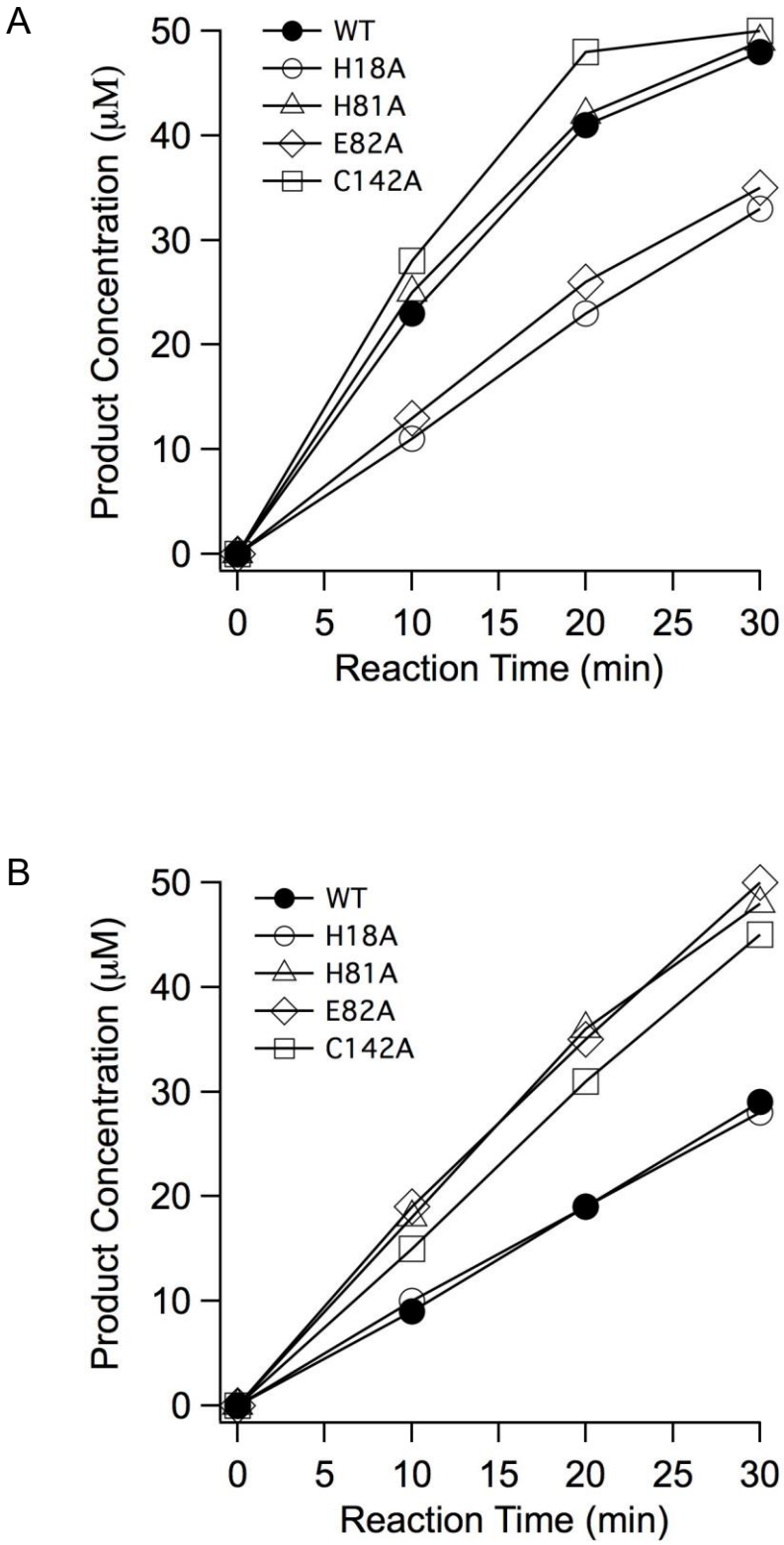


Figure S8: Kinetic traces for (A) piceatannol and (B) kaempferol glucosylation by wildtype *Pa*GT2 and mutants. A liner relationship between reaction time and product formation was observed in 20 min. The reactions were carried out at 37°C with 50 μM acceptor, 100 μM UDP-glucose, and 5 μM mutant enzymes in 50 mM potassium phosphate buffer (pH 7.4).

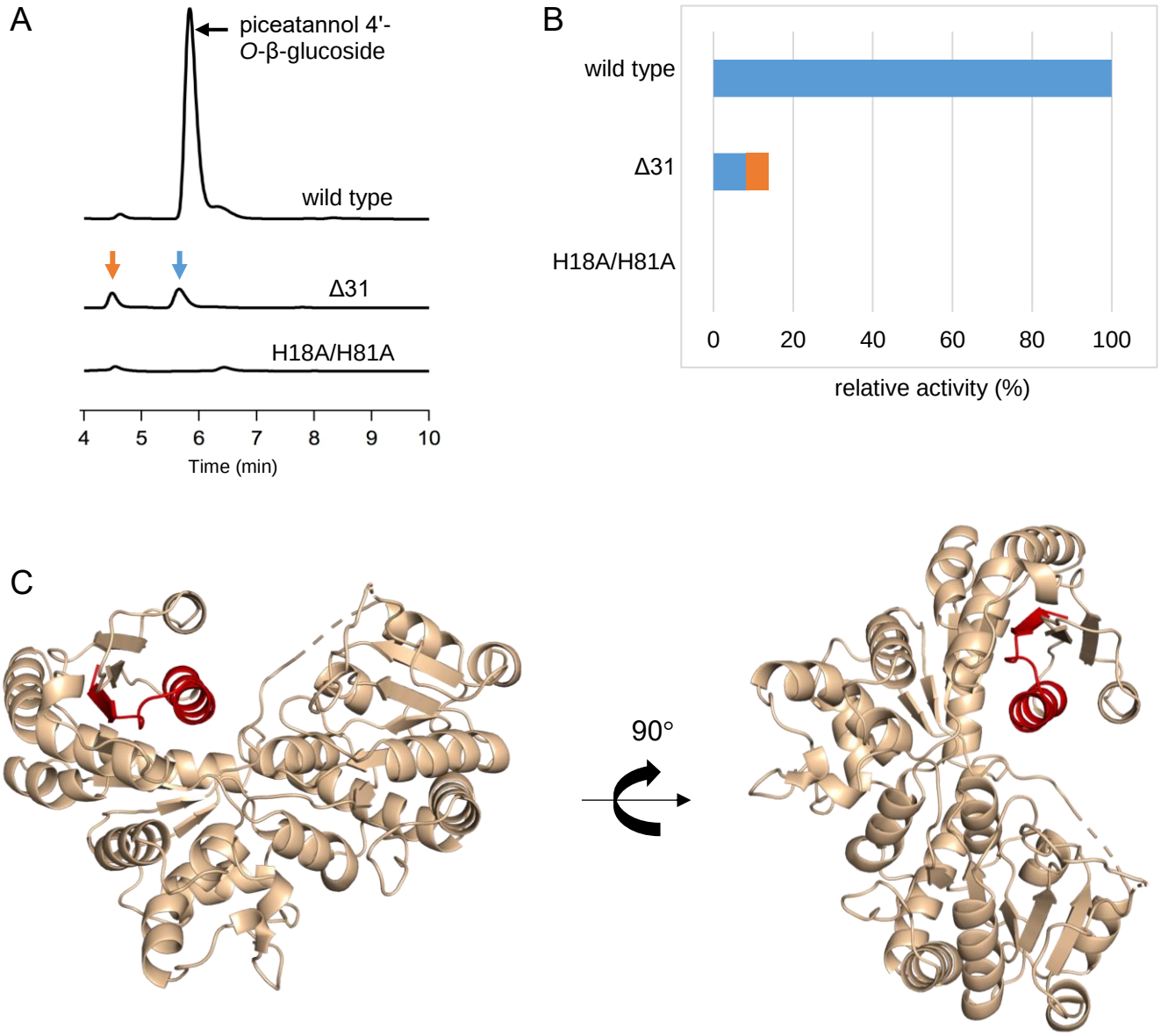
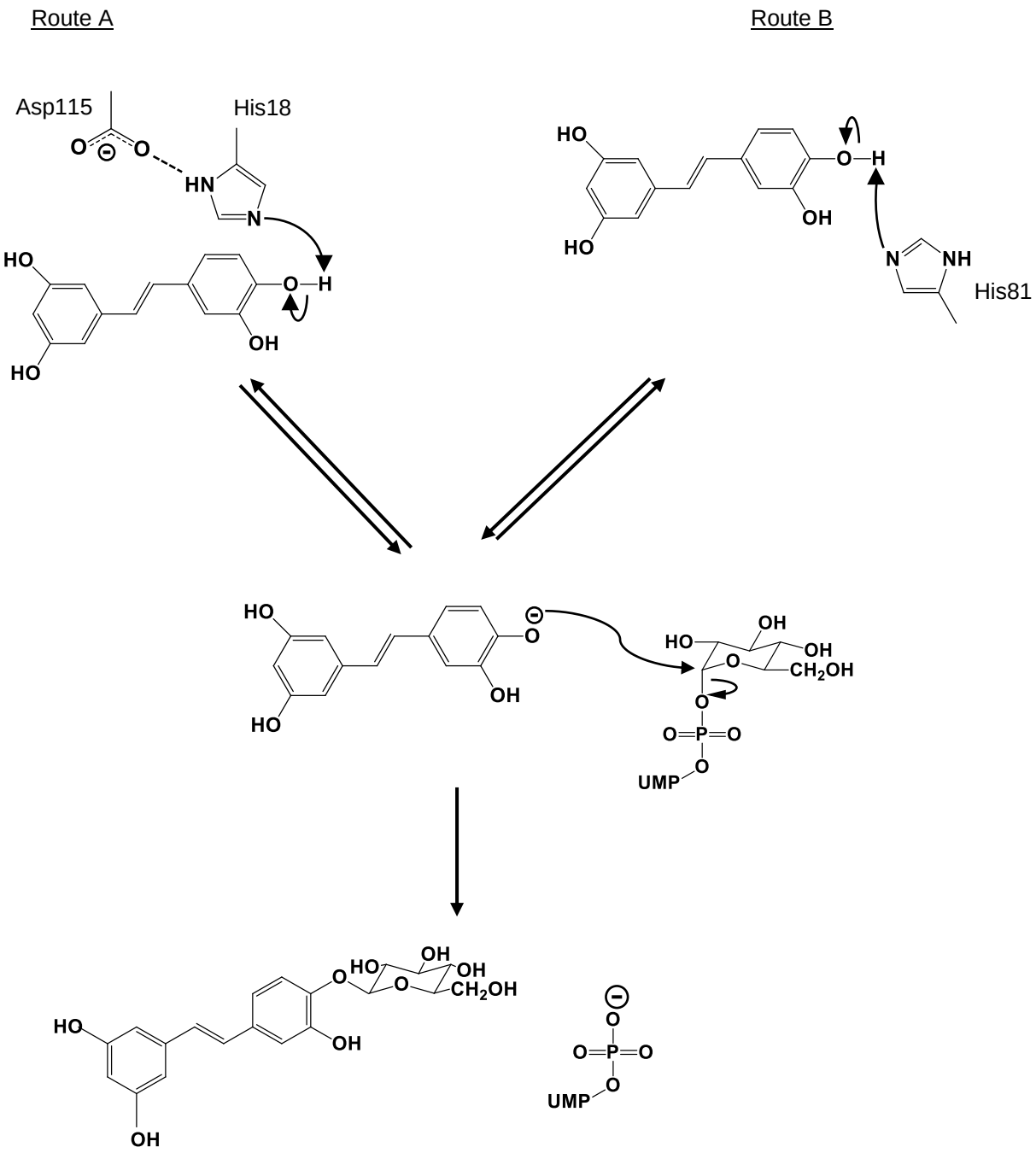
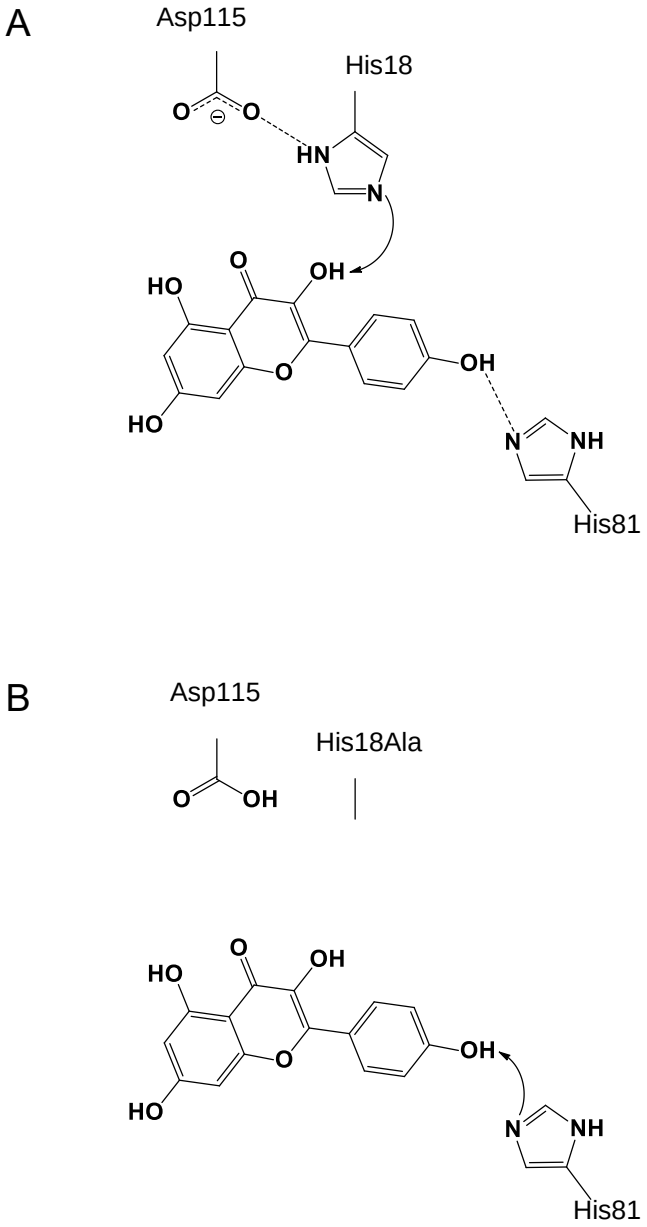


Figure S9: Glycosylation activity of *Pa*GT2 Δ31. A. HPLC profiles of piceatannol glucosylation by *Pa*GT2, *Pa*GT2 Δ31, and *Pa*GT2 H18A/H81A. B. Comparison of the piceatannol glucosylation activity. Relative activity was calculated based on the production of piceatnnnol 4′-*O*-β-glucoside (blue) and the side product (orange) shown by arrows in (A). C. Structure of *Pa*GT2 indicated with deleted 31 N-terminal residue in red color. The deletion of an α-helix and a β-sheet at the N-terminus could have distorted the structure of the enzyme.



Scheme S1. Piceatannol glycosylation in *Pa*GT2 can proceed via two routes, one catalyzed by the conserved catalytic pair (A) and the second catalyzed by non-conserved histidine (B).



Scheme S2: Kaempferol nucleophile generation by *Pa*GT2: A. The conserved mechanism in UGTs, where the conserved His-Asp pair generates the nucleophile for glycosylation. B. The complementary mechanism in *Pa*GT2, where the non-conserved catalytic base (His81) generates the nucleophile in absence of the conserved catalytic base (His18Ala). Possibly, His81 also generates the nucleophile in other mutants.

Table S1. Forward primers used for mutagenesis of *Pa*GT2

Mutant	Primers	Primer sequences
H18A	Forward	5'- atgggc**gcg**ctcatcccccta -3' *
	Reverse	5'- tccagggcttggaactatgactatgagtgg -3'
H81A	Forward	5'- gtggcc**gcg**gaggtcacaatctcc -3'
	Reverse	5'- gccgtcaggtaaatgggccgggtc -3'
E81A	Forward	5'- gtggcccac**gcg**gtcacaatctcc -3'
	Reverse	5'- gccgtcaggtaaatgggccgggtc -3'
C142A	Forward	5'- gctatg**gca**ttgctcttccttttt -3'
	Reverse	5'- cgtggatgtgaaatacaagtaaggtg -3'

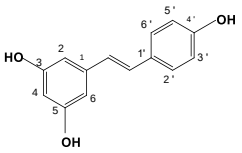
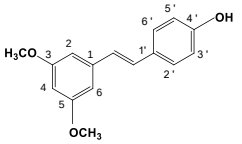
* Codon for the mutant residues are indicated with bold letters

Table S2. Data collection and refinement

	*Pa*GT2 (apo)	*Pa*GT2 + resveratrol + UDP-2FGlc	*Pa*GT2 + pterostilbene + UDP-2FGlc
Data collection			
X-ray source	SPring-8 BL44XU	SPring-8 BL44XU	SPring-8 BL44XU
Detector	Rayonix MX300HE	Rayonix MX300HE	EIGER X 16M
Wavelength (Å)	0.9	0.9	0.9
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions *a*, *b*, *c* (Å) *α*, *β*, *γ* (°)	 91.6 94.8 115.2 90.0 90.0 90.0	 57.3 137.6 208.9 90.0 90.0 90.0	 56.7 136.8 205.1 90.0 90.0 90.0
Resolution (Å)	50.00-2.30 (2.34-2.30) *	50.00-2.60 (2.64-2.60)	50.00-2.65 (2.74-2.65)
Total reflections	335208	349853	311499
Unique reflections	45171 (2216)	51665 (2535)	47340 (4530)
R_{merge}† (%)	7.2 (95.9)	8.6 (>100)	7.1 (>100)
R_{meas}‡ (%)	7.8 (103.0)	4.0 (51.4)	7.7 (>100)
$I/\sigma(I)$	28.5 (2.2)	25.9 (1.5)	15.7 (2.0)
$CC_{1/2}$§	(0.757)	(0.689)	(0.694)
Completeness (%)	99.9 (100.0)	99.0 (100.0)	99.8 (98.6)
Redundancy	7.4 (7.5)	6.8 (7.2)	6.6 (6.7)
Refinement			
Resolution (Å)	36.64-2.30 (2.36-2.30)	44.85-2.60 (2.66-2.59)	49.68-2.65 (2.71-2.65)
R_{work}/R_{free} (%)	21.99/25.11 (30.5/35.0)	19.27/25.22 (31.2/37.4)	19.44/24.76 (34.8/38.9)
RMSD bond length (Å)	0.0058	0.0085	0.0085
RMSD bond angles (°)	1.034	1.5581	1.5509
Ramachandran plot (%) Favored Allowed Outliers	 98.44 1.56 0.00	 97.28 2.72 0.00	 96.28 3.72 0.00
Average *B*-factor (Å^2) Protein Water UDP-2FGlc Resveratrol/Pterostilbene	 51.27 47.71	 81.43 80.64 136.23	 85.45 82.80 118.60
Molprobity score	1.30	1.85	1.96
Molprobity clash score (percentile)	4.81 (99th)	3.8 (99th)	3.9 (99th)
PDB ID	6JEL	6JEM	6JEN

Note: * Values in parentheses are for the highest resolution shell.
† $R_{merge} = \Sigma_{hkl}\Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$
‡ $R_{meas} = \Sigma_{hkl} \{N(hkl) / [N(hkl)-1]\}^{1/2} \times | \Sigma_{hkl}\Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the i^{th} observation of reflection (hkl) and N is the redundancy.
§ $CC_{1/2} = (\sigma_y^2 - \frac{1}{2}\sigma_\varepsilon^2) / (\sigma_y^2 + \frac{1}{2}\sigma_\varepsilon^2)$, where, σ_y^2 the variance of the average intensities across the unique reflections of a resolution shell and, σ_ε^2 the average of all sample variances of the averaged (merged) intensities across all unique reflections of a resolution shell.

Table S3. Distance between acceptors and polar residues of the enzyme in the designated molecule in the acceptor binding site of *Pa*GT2

Acceptor	Distance between	Chain A (Å)	Chain B (Å)	Chain C (Å)	Mean±SD*
Resveratrol 	His18 (NE2) - 4'OH	4.8	5.4	4.8	5.00 ± 0.34
	His81 (ND1) - 4'OH	4.3	4.3	4.6	4.40 ± 0.17
	Glu82 (OE1) - 4'OH	6.5	7.0	7.3	6.93 ± 0.40
	Ser138 (OG) - 3OH	3.8	3.8	3.7	3.76 ± 0.05
	Cys142 (SG) - 3OH	4.6	3.5	4.0	4.03 ± 0.55
Pterostilbene 	His18 (NE2) - 4'OH	6.0	6.1	6.1	6.06 ± 0.05
	His81 (ND1) - 4'OH	4.9	4.2	3.8	4.30 ± 0.55
	Glu82 (OE1) - 4'OH	8.8	8.0	7.9	8.23 ± 0.49
	Ser138 (OG) - 3OCH_3	4.3	4.1	3.9	4.10 ± 0.20
	Cys142 (SG) - 3OCH_3	4.6	4.6	3.9	4.36 ± 0.40

*SD: standard deviation.

Table S4. Binding affinity in the ascending order obtained for acceptors from molecular docking of ligands in PaGT2 structures

Acceptors	Binding mode	WT *Pa*GT2	*Pa*GT2 H18/81A (model)
		Binding affinity (kcal/mol)	Binding affinity (kcal/mol)
Resveratrol	(1)	-6.0	-6.1
	(2)	-5.9	-6.1
	(3)	-5.9	-6.0
	(4)	-5.9	-6.0
	(5)	-5.2	-5.8
	(6)	-5.0	-5.6
	(7)	-4.5	-5.6
	(8)	-3.1	-5.4
Pterostilbene	(1)	-6.2	-5.3
	(2)	-6.1	-5.9
	(3)	-5.9	-5.9
	(4)	-5.9	-5.8
	(5)	-4.5	-5.7
	(6)	-4.3	-5.7
	(7)	---	-5.6
	(8)	---	-5.5
	(9)	---	-5.3
Piceatannol	(1)	-5.9	-6.0
	(2)	-5.9	-6.0
	(3)	-5.9	-5.9
	(4)	-5.8	-5.6
	(5)	-5.6	-5.6
	(6)	-5.5	-5.6
	(7)	-5.4	-5.4
	(8)	-5.4	-5.4
	(9)	-5.3	-5.4
Kaempferol	(1)	-7.0	-6.8
	(2)	-6.9	-6.7
	(3)	-6.3	-6.7
	(4)	-5.8	-6.5
	(5)	-5.7	-6.4
	(6)	-5.4	-6.4
	(7)	-5.1	-5.8
	(8)	---	-5.8
	(9)	---	-5.8

Table S5. The specific activity of piceatannol glycosylation by *Pa*GT2 Cys142 mutants

	Specific activity (%)
Wild Type	100
C142Q	118.86
C142F	128.30
C142A	103.77