ACS Synthetic Biology
Cite This: ACS Synth. Biol. 2019, 8, 2718−2725
Research Article
pubs.acs.org/synthbio

Enzymatic *O*-Glycosylation of Etoposide Aglycone by Exploration of the Substrate Promiscuity for Glycosyltransferases

Kai-Zhi Jia,[†] Li-Wen Zhu,[†] Xudong Qu,[§] Shengying Li,[‡] Yuemao Shen,[‡] Qingsheng Qi,[‡] Youming Zhang,[‡] Yue-Zhong Li,[‡] and Ya-Jie Tang*[,†,‡]

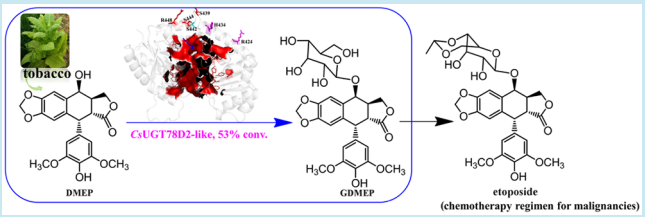
[†]Hubei Key Laboratory of Industrial Microbiology, Hubei Provincial Cooperative Innovation Center of Industrial Fermentation, Key Laboratory of Fermentation Engineering (Ministry of Education), Hubei University of Technology, Wuhan 430068, China

[‡]State Key Laboratory of Microbial Technology, Shandong University, Qingdao 266237, China

[§]Key Laboratory of Combinatorial Biosynthesis and Drug Discovery, Ministry of Education, School of Pharmaceutical Sciences, Wuhan University, Wuhan 430071, China

Supporting Information

ABSTRACT: The 4-*O*-*β*-D-glucopyranoside of DMEP ((−)-4′-desmethylepipodophyllotoxin) (GDMEP), a natural product from *Podophyllum hexandrum*, is the direct precursor to the topoisomerase inhibitor etoposide, used in dozens of chemotherapy regimens for various malignancies. The biosynthesis pathway for DMEP has been completed, while the enzyme for biosynthesizing GDMEP is still unclear. Here, we report the enzymatic *O*-glycosylation of DMEP with 53% conversion by exploring the substrate promiscuity and entrances of glycosyltransferases. Notably, we found 6 essential amino acid residues surrounding the putative substrate entrances exposed to the protein surface in UGT78D2, *Cs*UGT78D2, and *Cs*UGT78D2-like, and these residues may determine substrate specificity and high *O*-glycosylation activity toward DMEP. Our results provide an effective route for one-step synthesis of GDMEP. Identification of the key residues and entrances of glycosyltransferases will promote precise identification of glycosyltransferase biocatalysts for novel substrates and provide a rational basis for glycosyltransferase engineering.



KEYWORDS: *glycosyltransferase, substrate promiscuity, O-glycosylation, etoposide aglycone*

DMEP, (−)-4′-desmethylepipodophyllotoxin (Figure 1), is a naturally occurring lignin.[1] The biosynthetic pathway

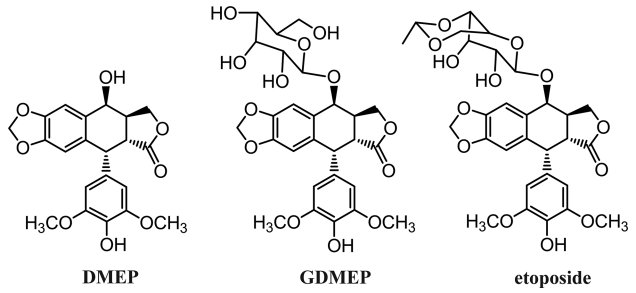


Figure 1. (−)-4′-Desmethylepipodophyllotoxin (DMEP) and its glycosides 4-*O*-*β*-D-glucopyranoside of DMEP (GDMEP) and etoposide.

of DMEP was reconstituted in *Nicotiana benthamiana* (tobacco) by overexpressing 6 pathway genes mined from the transcriptome of *Podophyllum hexandrum* and 4 previously discovered biosynthetic genes.[1] The 4-*O*-*β*-D-glucopyranoside of DMEP (GDMEP) (Figure 1), a natural product from *P. hexandrum*,[2] is the immediate precursor to the topoisomerase inhibitor etoposide (Figure 1) used in dozens of chemotherapy regimens for various malignancies.[3,4] The glycosylation of DMEP decreases the toxicity of this compound and increases the water-solubility, thereby improving the drug efficacy and pharmacokinetics,[5] and necessitates protection−deprotection steps in terms of chemical synthesis.[6,7] Notably, glycosyltransferases catalyze the glycosylation reaction in one step, transferring the sugar residue from various activated sugar donors to a variety of important acceptors.[8,9] However, the enzyme responsible for the glycosylation of DMEP remains unknown, mostly because the genome of *P. hexandrum* has not been sequenced, and the methods for mutant construction are laborious.[1,10]

The glycosylation patterns catalyzed by glycosyltransferases are usually associated with substrate promiscuity or specificity.[8] This promiscuity could contribute to the immense skeletal variations of small molecules in term of glycosylation pattern.[8,9] In contrast, glycosyltransferases can also be selective, and the specificities of some enzymes are defined by regiospecific or substrate-specific features of acceptors.[8,11,12] In this study, we designed and constructed the enzymatic *O*-glycosylation of DMEP using the substrate promiscuity of

Received: July 30, 2019
Published: November 27, 2019

DOI: 10.1021/acssynbio.9b00318
ACS Synth. Biol. 2019, 8, 2718−2725

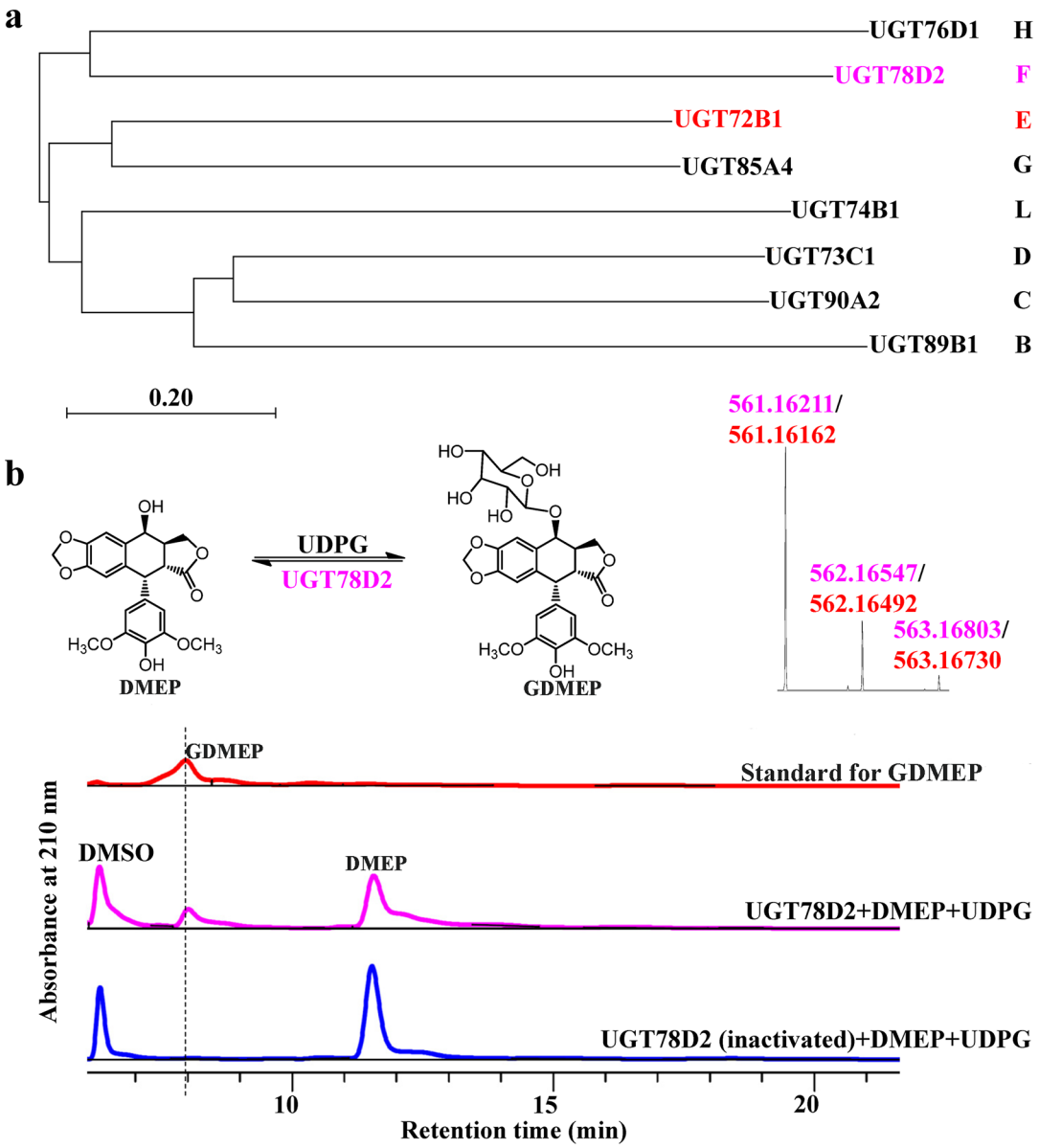


Figure 2. UGT78D2 exhibits high glycosylation activity toward DMEP. (a) Nonrooted molecular phylogenetic tree of the 8 UGTs. The tree was constructed from a ClustalW multiple sequence alignment using the neighbor-joining method in MEGA X 10.0.4. Bar = 0.2 amino acid substitutions/site. The GenBank accession numbers for the sequences are shown in parentheses: UGT76D1 (NP_180216.1), UGT78D2 (NP_197207), UGT72B1 (NP_192016), UGT85A4 (NP_177950), UGT74B1 (NP_173820), UGT73C1 (NP_181213), UGT90A2 (NP_172511.3), and UGT89B1 (NP_177529). (b) HPLC analysis of enzymatic products produced by UGT78D2, with retention times relative to that of the GDMEP standard. The standard was not detected in the control with denatured UGT78D2. The glycosylation reaction catalyzed by UGT78D2 was conducted at pH 8.5. DMSO, dimethyl sulfoxide.

glycosyltransferase and then identified glycosyltransferases with high *O*-glycosylation activity toward DMEP by exploring the phylogenetic relationships of glycosyltransferases. By combining chimera construction, alanine mutation, and entrance analysis, we identify the essential amino acid residues responsible for high *O*-glycosylation activity toward DMEP and provide a model for substrate recognition. This study will provide an effective route for the synthesis of GDMEP in one step, promote precise identification of glycosyltransferase biocatalysts for novel substrates, and provide a rational basis for glycosyltransferase engineering.

RESULTS AND DISCUSSION

UGT72B1 Catalyzes the *O*-Glycosylation of DMEP. UGTs (UDP-glycosyltransferases) catalyze the transfer of a sugar moiety from a UDP-activated sugar onto a broad range of endogenous and xenobiotic substrates and exhibit broad acceptor tolerance.[12−14] This promiscuity of UGTs is exploited by *Arabidopsis* plants to metabolize xenobiotic substrates such as herbicides, pesticides, and organic pollutants.[15] A total of 122 family-1 UGT-like sequences were found in the *Arabidopsis* genome (http://www.p450.kvl.dk/At_ugts/table.shtml).[15,16] These sequences were divided into 14 groups based on sequence similarity and evolutionary relatedness.[16] In general, UGT activity toward hydroxylated substrates was widely observed in groups B, C, D, E, L, F, G, and H.[15,16]

DOI: 10.1021/acssynbio.9b00318
ACS Synth. Biol. 2019, 8, 2718−2725

Table 1. UGT Activity toward DMEP

UGTs	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m ($M^{-1}\cdot min^{-1}$)	conversion[a] (%)
UGT78D2	21.64 ± 0.39	0.0080 ± 0.0001	369.69	39
*Cs*UGT78D2-like	0.99 ± 0.08	0.0034 ± 0.0000	3434.34	53
*Cs*UGT78D2	2.32 ± 0.15	0.0020 ± 0.0001	862.07	45

[a]Conversion after incubation at optimal pH (8.5) for 36 h.

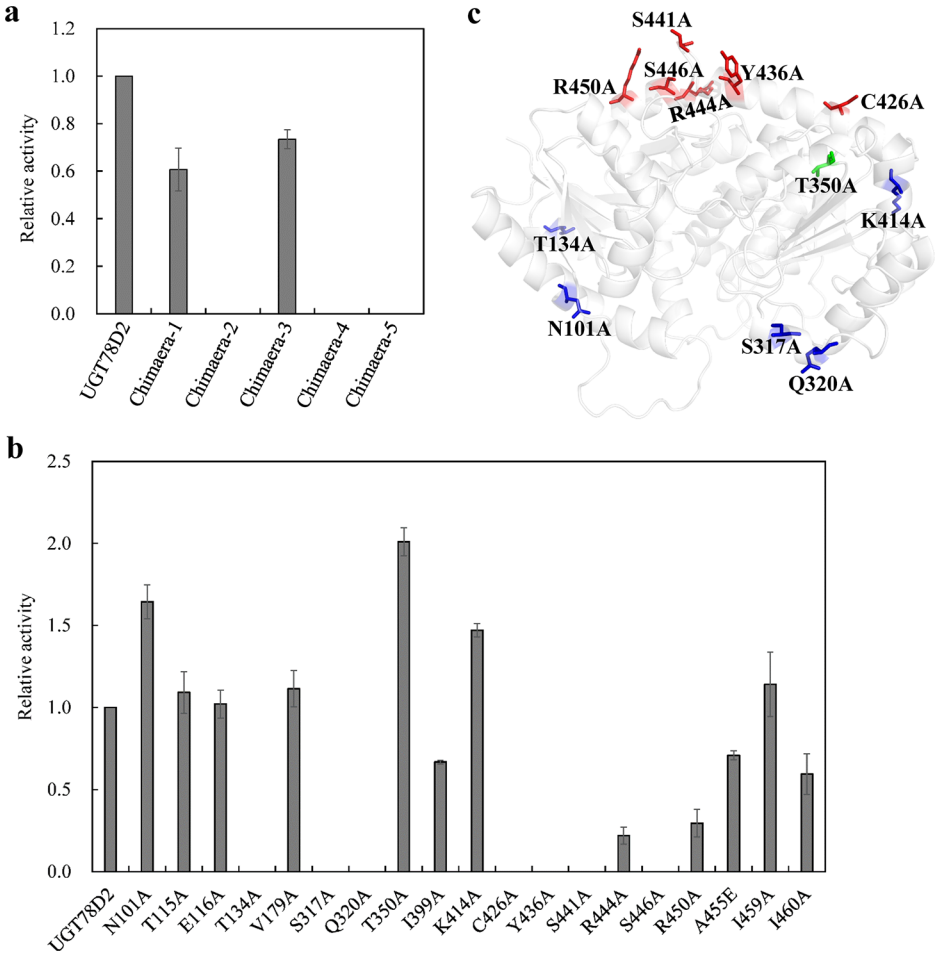


Figure 3. Essential amino acid residues implicated in DMEP glycosylation. (a) Activity analysis of chimeras for identification of key regions implicated in DMEP glycosylation. The specific glycosylation activity toward DMEP for UGT78D2 was 2.82 ± 0.15 nmol GDMEP·min^{-1}·mg recombinant $protein^{-1}$ and the relative activity for glycosylation of DMEP to GDMEP was set at 1.0. (b) Relative activity of purified UGT78D2 and its mutants using DMEP as the substrate. (c) Positions of the essential amino acid residues in the UGT78D2 model. The UGT78D2 model was constructed using *Vv*GT1 (PDB ID: 2C1X) as a template. T350A positioned in the PSPG box is marked as a green stick model, and C426A, Y436A, S441A, R444A, S446A, and R450A, surrounding the putative entrance, are marked as red stick models. The other essential amino acid residues are marked as blue stick models.

UGT72B1, belonging to group E, was the most active *O*-glycosyltransferase toward hydroxylated substrates, including 2,4,5-trichlorophenol, 2-chloro-4-trifluoromethylphenol, and triclosan, and exhibited high promiscuity in the catalysis of xenobiotic glycosylation.[15,16] Therefore, UGT72B1 was adopted for glycosylation of DMEP.

The ORF (open reading frame) of UGT72B1 was amplified with specific primers and cDNA (complementary DNA) obtained from *Arabidopsis thaliana* mRNA, cloned into pYES2 and then heterologously expressed in *Saccharomyces cerevisiae* (Table S1). Recombinant UGT72B1 was purified using Ni-nitrilotriacetic acid chromatography and size exclusion chromatography (Figure S1a). A protein with the predicted molecular weight (52.9 kDa) was detected by migration on an SDS-PAGE gel (Figure S1i), and then, the purified protein was identified as UGT72B1 by using MALDI-TOF MS (Table S2). The activity of UGT72B1 toward DMEP was examined. The purified UGT72B1 was incubated with UDPG as a sugar donor and DMEP as a sugar acceptor. The trace amount of reaction product was able to be detected by LC−MS (an Ultimate 3000 UHPLC (ultra high performance liquid chromatography) system coupled with a Q Exactive Focus mass spectrometer), but not observed by usual HPLC (high-performance liquid chromatography) with UV (ultraviolet−visible) detector. The reaction product was with the same retention time (6.15 min) and molecular formula of $C_{27}H_{29}O_{13}$ (m/z $[M-H]^-$ 561.16162, calculated 561.16136) as the purchased GDMEP standard (Figure S2a). Additionally,

DOI: 10.1021/acssynbio.9b00318
ACS Synth. Biol. 2019, 8, 2718−2725

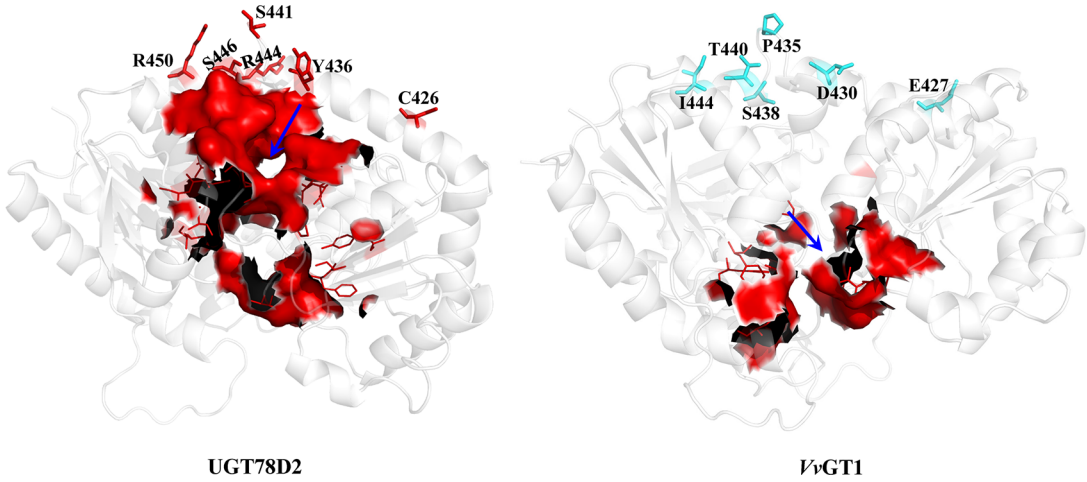


Figure 4. Comparison of the entrances of UGT78D2 and *Vv*GT1. The entrances were predicted by CAVER software, the amino acid residues located in the entrance and tunnel are shown as red lines, exhibited as surfaces and noted in Table S3. C426, Y436, S441, R444, S446, and R450, surrounding the putative entrance in UGT78D2, are marked as red stick models, and the corresponding amino acid residues in *Vv*GT1 are marked as cyan stick models.

a similar isotopic pattern distribution and MS2 spectrum were also observed (Figure S2b, S3a–c). These findings strongly indicated that UGT72B1 transfers a glycosyl moiety to the C4-OH of DMEP, albeit with very low activity.

UGT78D2 Exhibits High *O*-Glucosyltransferase Activity toward DMEP. Enzyme promiscuity was regarded as the starting point for the divergent evolution of enzymes, and the evolved enzymes tend to be specific to a single substrate and reaction.[17−20] The UGTs in groups B, C, D, E, L, F, G, and H exhibited *O*-glycosylation activity toward hydroxylated substrates. Therefore, 8 UGTs were selected from each group and represented in a phylogenetic tree (Figure 2a). In addition to UGT72B1, the other 7 UGTs were heterologously expressed in *Saccharomyces cerevisiae*, purified using Ni-nitrilotriacetic acid chromatography and size exclusion chromatography, detected by migration on an SDS-PAGE gel, and then identified using MALDI-TOF MS (Figure S1b–i, Table S2). The *in vitro* enzymatic activity assay indicated that UGT78D2 exhibited the strongest capacity for DMEP glycosylation, and the main product had the same retention time (8.02 min, detected by usual HPLC), molecular formula of $C_{27}H_{29}O_{13}$ (m/z $[M-H]^-$ 561.16211, 561.16136 calculated), and the similar isotopic pattern distributions (detected by UHPLC system coupled with a Q Exactive Focus mass spectrometer) as the purchased GDMEP standard (Figure 2b). The NMR data (Figure S4) of the main product were in accordance with those reported for GDMEP.[21] UGT89B1, UGT90A2, UGT73C1, UGT74B1, UGT85A4, and UGT76D1 were inactive, with no DMEP glycosylation detected even by high-resolution mass spectrometry. UGT78D2 produced the highest amount of GDMEP at pH 8.5, exhibiting 39% conversion (Figure 2b, Table 1).

Essential Amino Acid Residues Responsible for the High *O*-Glucosyltransferase Activity of UGT78D2 toward DMEP. Amino acid sequence alignment indicated that the catalytic residue histidine, acting as a Bronsted base, and the plant secondary product glycosyltransferase (PSPG) box, representing the UDP-sugar (UDPG) binding motif in the C-terminal domain, were highly conserved in the 8 UGTs (Figure S5). Ile86, Leu118, Phe119, Phe148, Leu183, and Leu197 (UGT72B1 numbering) constituted the acceptor-binding pocket positioned in the N-terminal domain,[15] but 3 binding sites (Ile86, Phe119, and Leu197) were conserved (Figure S5). UGT78D2 contains 460 amino acids and shares 28% identity with UGT72B1. Thus, we could not determine which domain determines the high *O*-glucosyltransferase activity of UGT78D2 toward DMEP.

To investigate the unusual catalytic feature of UGT78D2, we compared the glycosylation activity of this protein to that of other members of group F, including UGT78D1 and UGT78D3, which share 72% and 76% identity, respectively, with UGT78D2 at the amino acid level (Figure S6–S7). The red grape enzyme *Vv*GT1 shared 59% identity with UGT78D2 (Figure S6–S7), and the three-dimensional structure of this protein has been determined.[22] However, *Vv*GT1 produced only trace amounts of GDMEP, as detected by high-resolution mass spectrometry (Figure S3a,b,d). UGT78D1 and UGT78D3 were inactive because these proteins were specific for UDP-rhamnose and arabinose but not UDPG as a sugar donor.[23−25]

UGT78D2 exhibited higher catalytic efficiency for transferring the glucose moiety to its acceptor than UGT78D3 did for the arabinose moiety.[26] To locate the key regions that determine the high *O*-glucosyltransferase activity of UGT78D2 to DMEP, we constructed 5 chimeras by replacing the peptide segments of UGT78D2 with the corresponding parts of UGT78D3 and compared the ability of these chimeras to produce GDMEP (Figure S6, Table S1). Interestingly, chimeras 2, 4, and 5 abolished *O*-glycosyltransferase activity, which suggested that the corresponding peptide segments of UGT78D2 determine the high *O*-glucosyltransferase activity toward DMEP (Figure 3a). Subsequently, we compared the amino acid sequences of UGT78D2 with the corresponding sequences of *Vv*GT1, UGT78D1, and UGT78D3 *via* ClustalW multiple sequence alignment and found that 19 amino acid residues were different from those in *Vv*GT1, UGT78D1, and UGT78D3 (Figure S6). These differences are assumed to be associated with the DMEP glycosylation-specific enzyme activity of UGT78D2.

To test the functional importance of these 19 amino acid residues, we mutated these residues to Ala and examined the ability of these mutants to glycosylate DMEP. As seen in

DOI: 10.1021/acssynbio.9b00318
ACS Synth. Biol. 2019, 8, 2718−2725

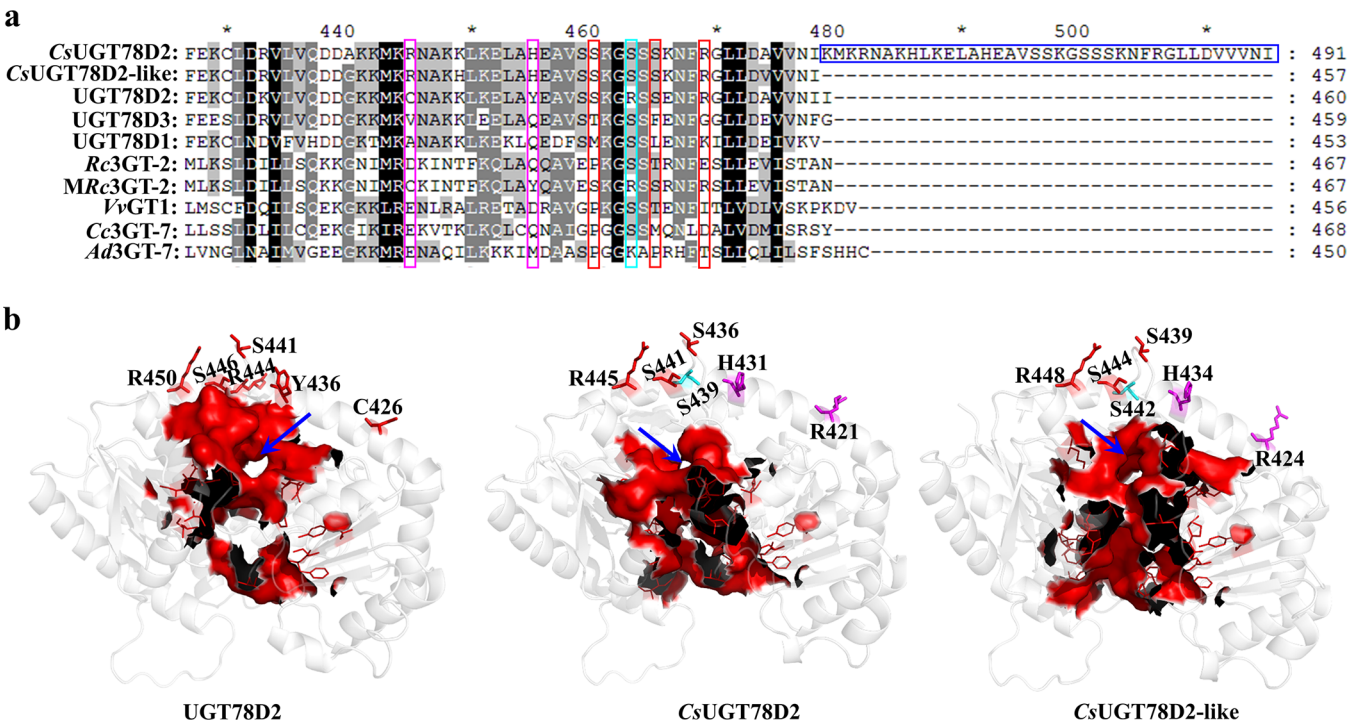


Figure 5. Essential amino acid residues associated with substance recognition. (a) Conservation analysis of the 6 essential amino acid residues in UGT78D2 and the other UGTs. The abbreviations are as follows: *Cs*UGT78D2, UDP-glycosyltransferase 78D2 (*Camelina sativa*); *Cs*UGT78D2-like, UDP-glycosyltransferase 78D2-like (*C. sativa*); *Rc*3GT-2, anthocyanidin 3-*O*-glucosyltransferase 2 (*Rosa chinensis*); *MRc*3GT-2, mutant of anthocyanidin 3-*O*-glucosyltransferase 2 (*R. chinensis*); *Cc*3GT-7, anthocyanidin 3-*O*-glucosyltransferase 7 (*Citrus clementina*); *Ad*3GT-7, anthocyanidin 3-*O*-glucosyltransferase 7 (*Arachis duranensis*). (b) The 6 essential amino acid residues positioned in UGT78D2, *Cs*UGT78D2, and *Cs*UGT78D2-like. The amino acid residues in *Cs*UGT78D2 and *Cs*UGT78D2-like that coincided with or differed from the 6 essential amino acid residues in UGT78D2 are marked in red or purple, respectively. Ser positioned at 439 and 442 in *Cs*UGT78D2 and *Cs*UGT78D2-like, respectively, coincided with *Vv*GT1 and is marked in cyan.

Figure 3b, T134A, S317A, Q320A, C426A, Y436A, S441A, and S446A were inactive; R444A and R450A exhibited lower DMEP glycosylation activity than wild-type UGT78D2 (22% and 30% relative activity), while N101A, T350A, K414A exhibited stronger DMEP glycosylation activity than wild-type UGT78D2, indicating that the 12 amino acid residues are essential for DMEP glycosylation.

To gain further insights into the role of the 12 essential amino acid residues in DMEP glycosylation, we positioned them in the predicted 3D structure of UGT78D2 and found that the 12 essential amino acid residues, except T350A, were not located near the active site of UGT78D2 (Figure 3c). The T350A mutation was positioned in the PSPG box, and this mutant enhanced the relative activity by 101%; this mutation was inferred to influence the binding of UDPG. The other essential amino acid residues are part of the backbone helices and were inferred to be associated with enzyme stability. Moreover, C426A, Y436A, S441A, R444A, S446A, and R450A are contained in the backbone helix, its corresponding helices in 5GT-Cha from *V. vinifera* cv. Regent and UGT71A33 from strawberry were thought to be related with the entrance to the active site by deletion of amino acid sequences and mutagenesis of amino acid residues.[27,28]

Substrate Recognition of UGTs with High *O*-Glucosyltransferase Activity toward DMEP. The active site is connected with the surrounding environment by various channels. Accordingly, to access the active site, acceptors must pass through the protein *via* an entrance tunnel that acts as a gate keeper (molecular filter) and impacts substrate selectivity and sometimes even activity.[29−31] To investigate whether an entrance existed in the vicinity of the backbone helix, we adopted CAVER software to analyze the entrance of UGT78D2.[29] Interestingly, an entrance was predicted to exist in this region, which was surrounded by C426, Y436, S441, R444, S446, and R450 and exposed to the surface of UGT78D2 (Figure 4). Moreover, the entrance of *Vv*GT1, with relatively low glycosylation activity toward DMEP, was buried within *Vv*GT1 and was distant from the amino acid residues corresponding to C426, Y436, S441, R444, S446, and R450 in UGT78D2 (Figure 4). Comparison of the entrances of UGT78D2 and *Vv*GT1 suggested that the substrate DMEP may have much easier access to the entrance of UGT78D2 than to that of *Vv*GT1.

Most of the residues constituting the entrance, tunnel, and PSPG box are located between residues 250 and 460. These residues, together with the essential amino acid residues identified above, were designated as X in the bait sequence of 250−460 to perform pattern matching and similarity searches in the NCBI database. A total of 135 homologous glycosyltransferases that shared 43−83% identity with UGT78D2 were obtained. After homologous model building and entrance analysis, 19 UGTs with entrances exposed to the protein surface were selected (Table S3). Among these proteins, the predicted entrance of *Rc*3GT-2 was very close to the protein surface, while *Ad*3GT-7 and *Cc*3GT-7 were predicted to have two entrances and one tunnel that enclosed the PSPG box (Table S3). While these three UGTs had negligible glycosylation activity toward DMEP, this finding

DOI: 10.1021/acssynbio.9b00318
ACS Synth. Biol. 2019, 8, 2718−2725

indicated that the exposure of the entrance to the protein surface could not adequately explain the high glycosylation activity toward DMEP for UGT78D2.

The residues surrounding the predicted entrance in *Rc*3GT-2, *Ad*3GT-7, and *Cc*3GT-7 completely differed from the six essential residues (C426, Y436, S441, R444, S446, and R450) in UGT78D2 but appeared to be conserved to the corresponding residues in UGT78D1, UGT78D3, and *Vv*GT1 (Figure 5a). Phylogenetic analysis of all the UGTs used in this study indicated that UGT78D2 formed a distinct subgroup with *Cs*UGT78D2 and *Cs*UGT78D2-like, which are annotated as unknown proteins, and are distant from *Rc*3GT-2, *Cc*3GT-7, *Ad*3GT-7, *Vv*GT1, and the other UGTs (Figure S7). These results suggested that the 6 essential amino acid residues may play an important role in the glycosylation of DMEP. *Cs*UGT78D2 and *Cs*UGT78D2-like contain 3 out of the six essential residues in UGT78D2 (Figure 5b). *Cs*UGT78D2 and *Cs*UGT78D2-like exhibited higher glycosylation activity toward DMEP than UGT78D2, and the conversion increased by 6% and 14%, respectively (Table 1). This increase was caused by enhancement of the DMEP binding ability of *Cs*UGT78D2 and *Cs*UGT78D2-like by 8.33- and 20.86-fold, respectively (Table 1, Figure S8). Subsequently, the 6 essential residues of UGT78D2 were adopted to replace the corresponding amino acid residues of *Rc*3GT-2, which exhibited negligible activity toward DMEP. Distinct accumulation of GDMEP was observed in the enzymatic system catalyzed by the *Rc*3GT-2 mutant, and the conversion reached 13%. These data suggested that the six essential residues are closely associated with substrate recognition. Interestingly, two 37-amino-acid repeats containing the six essential residues were present near the C-terminus of *Cs*UGT78D2. However, DMEP binding ability of this protein is close to that of *Cs*UGT78D2-like, containing a single 37-amino-acid repeats. On the basis of the analysis of the homologous 3-D model, we speculated that the additional 37-amino-acid repeat slips over the entrance of *Cs*UGT78D2, and does not strengthen the ability for DMEP binding. All these results indicated that the 6 essential amino acid residues are involved in substrate recognition and determine substrate specificity. Glycosyltransferases from *Podophyllum hexandrum* were analyzed by phylogenetics and amino acid sequence alignment (Figure S9). PH000489, PH000719, PH045342, and PH009645 got together with *Ad*3GT-7 and *Rc*3GT-2, which had negligible glycosylation activity toward DMEP (Figure S9a), and the six amino acid residues were not conserved to the corresponding essential residues in UGT78D2 except S495 in PH000489 (Figure S9b). Thus, we inferred that PH000489, PH000719, PH045342, and PH009645 may have low glycosylation activity toward DMEP. Glycosylation patterns were previously predicted by coupling physicochemical features with isozyme recognition patterns over the entire glycosyltransferase superfamily 1 (GT1) of the plant *A. thaliana* without considering substrate specificity of glycosyltransferase.[32] The identification of the 6 essential amino acid residues will promote the precise identification of GT1 biocatalysts for novel substrates and annotation of uncharacterized GT1 enzymes and provide a rational basis for engineering GT1.

METHODS

Heterologous Expression and Purification of UGTs. Total RNA was isolated from *A. thaliana* plants using the Total RNA Kit II R6934 (Omega Biotek, USA) and reverse transcribed to cDNA using the SMARTer PCR cDNA Synthesis Kit (Clontech, USA). Subsequently, the full-length *UGT72B1*, *UGT78D2*, *UGT78D1*, *UGT78D3*, *UGT89B1*, *UGT76D1*, *UGT85A4*, *UGT74B1*, *UGT73C1*, and *UGT90A2* were amplified, sequenced, and cloned into the yeast expression plasmid pYES2. Additionally, fusion PCR with the primers detailed in Supplementary Table S1 was adopted to generate the UGT78D2-UGT78D3 chimera genes, which were then cloned into pYES2. For purification of these UGTs, a His-tag-encoding fragment was included after the second codon (TCT). The resulting plasmids were transformed into *S. cerevisiae* strain INVSc1, and then, the UGTs were expressed and purified using nickel affinity chromatography and size exclusion chromatography, as described.[33] The genes encoding *Vv*GT1, *Rc*3GT-2, *Ad*3GT-7, *Cc*3GT-7, *Cs*UGT78D2, and *Cs*UGT78D2-like were synthesized and cloned into pET28a by GenScript (Nanjing, China), and then, the expressed proteins were purified by nickel affinity chromatography, as described.[33] All the UGTs were subjected to MALDI-TOF/TOF analysis by a 4700 Proteomics Analyzer (Applied Biosystems, Framingham, MA) and identified based on a 95% or higher confidence interval of their scores in the MASCOT V2.0 search engine (Matrix Science, London, U.K.).

Glycosyltransferase Activity Determination. Glycosyltransferase activity was assayed (total volume of 1 mL) in 50 mM Tris-HCl (pH ranging from 7.0 to 9.0) containing 0.25 mM DMEP (purchased from Shanxi Huisheng Medicament Technology Company, Ltd. (Shanxi, China), over 98% pure), 0.50 mM UDPG (purchased from Sigma-Aldrich, over 98% pure) and 23 μg of the purified enzymes. The reaction was performed at 30 °C for 24 h and then terminated by double the amount of methanol. The activity of the UGTs and their mutants was determined by measuring the amount of GDMEP (the standard was purchased from Toronto Research Chemicals Inc., >95% pure) produced by the enzymatic reaction system. Chromatographic separation was performed on a Reprosil-Pur Basic C18 column (4.6 mm × 250 mm × 5 μm) from Dr. Maisch GmbH (Germany) using usual HPLC (DGU-20A$_{5R}$ liquid chromatography system, Shimadzu Corporation, Japan) with 50/50 methanol/water as the mobile phase. The column oven temperature was set to 40 °C, and the flow rate was 0.5 mL/min. The detection wavelength was 210 nm. The retention time for DMEP was 8.02 min (Figure 2b). The DMEP concentrations used were 0.5 to 500 μM, with UDPG at 500 μM for DMEP kinetics. The kinetic parameters were calculated based on the method described by Asada *et al.*[34] Mean values ± SD of three independent experiments are shown.

To identify the glycosylated products, an Ultimate 3000 UHPLC system coupled with a Q Exactive Focus mass spectrometer was employed for comparing the MS and MS2 spectra with the standard GDMEP, and the retention time for GDMEP was 6.15 min (Figure S2a). The product of glycosylation catalyzed by UGT78D2 was extracted with CH_2Cl_2. The organic layer was dried over $MgSO_4$, evaporated, and the crude residue was passed through the Amethyst C18−H column (10 mm × 250 mm × 5 μm) from Sepax Technologies (Newark, DE). Targeted samples were evaporated, freeze-dried, and analyzed by using 1D and 2D NMR (nuclear magnetic resonance) (Figure S4).

The product of glycosylation catalyzed by UGT78D2 (4-*O*-β-D-glucopyranoside of 4′-desmethylepipodophyllotoxin, white powder, yield: 12%; purity: 94%): ^{1}H NMR (400 MHz,

DOI: 10.1021/acssynbio.9b00318
ACS Synth. Biol. 2019, 8, 2718−2725

DMSO-d_6) δ 8.25 (s, 1H), 7.06 (s, 1H), 6.54 (s, 1H), 6.18 (s, 2H), 6.02 (d, J = 4.4 Hz, 2H), 5.04 (d, J = 3.2 Hz, 1H), 5.01 (d, J = 4.4 Hz, 1H), 4.94 (d, J = 5.2 Hz, 2H), 4.66 (t, J = 5.8 Hz, 1H), 4.50 (d, J = 5.4 Hz, 1H), 4.39−4.25 (m, 3H), 3.79−3.72 (m, 1H), 3.61 (s, 6H), 3.50−3.43 (m, 1H), 3.40−3.35 (m, 1H), 3.16−3.08 (m, 2H), 3.06−2.98 (m, 2H), 2.92−2.83 (m, 1H); ^{13}C NMR (101 MHz, DMSO-d_6) δ 174.96, 147.68, 147.12, 146.07, 134.64, 133.04, 130.46, 128.26, 110.13, 110.03, 108.35, 101.25, 99.85, 77.08, 76.59, 73.62, 70.48, 70.43, 67.74, 61.34, 55.98, 42.99, 40.50, 37.37; HRMS (m/z) [M−H]$^-$ calcd. for 561.16136, $C_{27}H_{29}O_{13}^-$, found 561.16211.

Generation of Mutant UGTs. For generation of mutant UGT78D2, *UGT78D2* was synthesized and cloned into pET28a, and the expressed UGT78D2 was purified following the above-mentioned method. To identify the essential amino acid residues in UGT78D2, the Mut Express II Fast Mutagenesis Kit V2 (Vazyme, Nanjing, Jiangsu Province, China) was adopted to mutate the selected amino acid residues to alanine (the alanine of UGT78D2 was mutated to the corresponding amino acid of UGT78D3) by using the primers detailed in Table S1.[35−37]

Molecular Modeling of Glycosyltransferases. The models of glycosyltransferases were established by using SWISS-MODEL (https://swissmodel.expasy.org/).[38] The crystal structure of *Vv*GT1 (PDB ID: 2C1X), which carried over 40% sequential identity with the 135 homologous glycosyltransferases mentioned above and has the higher GMQE (global model quality estimation) value, was adopted as the template for building the 3D structure of the 135 glycosyltransferases. The entrances were predicted by CAVER software and exhibited using PyMOL software (https://pymol.org/2/).[29]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.9b00318.

Supporting Table S1: Primers used for cloning and overexpression of genes encoding UGTs and construction of mutants and chimeras of UGTs; Supporting Table S2: Identification of UGTs by MALDI-TOF MS; Supporting Table S3: Analysis of entrance and tunnel of UGTs; Supporting Figure S1: Purification of the 8 UGTs; (a−h) Purification of the 8 UGTs (UGT72B1, UGT90A2, UGT85A4, UGT73C1, UGT76D1, UGT89B1, UGt74B1, and UGT78D2) by size exclusion chromatography; (i) SDS-PAGE analysis of the 8 UGTs; Supporting Figure S2: Glycosylation activity of UGT72B1 toward DMEP; High-resolution MS analysis of enzymatically formed products produced by UGT72B1 based on retention times (extracted ion chromatograms) (a), precise molecular weight and isotopic pattern distributions (b) relative to the GDMEP standard; Supporting Figure S3: Identification of enzymatically formed products produced by UGT72B1 and *Vv*GT1 from DMEP; Total ion currents for the standard and UGT72B1, *Vv*GT1-catalyzed reaction mixtures (a), MS/MS data for the standard (b) and the glycosylation products produced by UGT72B1 (c) and *Vv*GT1 (d); Supporting Figure S4: ^{1}H, ^{13}C NMR and 2D NMR (^{1}H−^{1}H COSY, HMBC, HSQC) spectra of the glycosylation product; Supporting Figure S5: Amino acid sequence alignment of the 8 UGTs (UGT72B1, UGT90A2, UGT85A4, UGT73C1, UGT76D1, UGT89B1, UGt74B1, and UGT78D2); Supporting Figure S6: Amino acid sequence alignment of UGT78D2, *Vv*GT1, UGT78D1, and UGT78D3; Supporting Figure S7: Nonrooted molecular phylogenetic tree of UGTs, including the 8 UGTs, UGT78D1, UGT78D3, and the UGTs listed in Supplementary Table S3; Supporting Figure S8: Determination of kinetic constants for UGT reaction catalyzed by UGT78D2 (a), *Cs*UGT78D2-like (b), and *Cs*UGT78D2 (c); Supporting Figure S9: Glycosyltransferases from *Podophyllum hexandrum* were analyzed by phylogenetics (a) and amino acid sequence alignment (b) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Tel./Fax: +86-532-58632365. E-mail: yajietang@sdu.edu.cn.

ORCID

Xudong Qu: 0000-0002-3301-8536
Shengying Li: 0000-0002-5244-870X
Yue-Zhong Li: 0000-0001-8336-6638
Ya-Jie Tang: 0000-0002-8342-5201

Author Contributions

K.-Z.J.: Data curation, formal analysis, investigation, methodology, writing (original draft), review and editing. L.-W.Z., X.-D.Q.: Formal analysis, methodology. S.L., Y.S., Q.Q., Y.Z., Y.L.: Writing (review and editing). Y.-J.T.: Supervision, project administration, writing (review and editing).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation for Distinguished Young Scholars (Grant No. 21625602), the National Natural Science Foundation of China (Grant Nos. 21838002 and 31570054), and Hubei Provincial Science and Technology Innovation Major Project (2017ACA173).

■ REFERENCES

(1) Lau, W., and Sattely, E. S. (2015) Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science 349*, 1224−1228.

(2) Yu, X., Che, Z. P., and Xu, H. (2017) Recent advances in the chemistry and biology of podophyllotoxins. *Chem. - Eur. J. 23*, 4467−4526.

(3) Gaspar, N., Occean, B.-V., Pacquement, H., Bompas, E., Bouvier, C., Brisse, H. J., Castex, M.-P., Cheurfa, N., Corradini, N., Delaye, J., Entz-Werle, N., Gentet, J.-C., Italiano, A., Lervat, C., Marec-Berard, P., Mascard, E., Redini, F., Saumet, L., Schmitt, C., Tabone, M.-D., Verite-Goulard, C., Le Deley, M.-C., Piperno-Neumann, S., and Brugieres, L. (2018) Results of methotrexate-etoposide-ifosfamide based regimen (M-Ei) in osteosarcoma patients inculded in the French OS2006/sarcome-09 study. *Eur. J. Cancer 88*, 57−66.

(4) Smith, S. J., Tyler, B., Gould, T. W. A., Veal, G. J. V., Gorelick, N. L., Rowlinson, J., Serra, R., Ritchie, A. A., Berry, P., Otto, A., Choi, J., Skuli, N., Estevez-Cebrero, M. A., Shakesheff, K. M., Brem, H., Grundy, R. G., and Rahman, R. (2019) Overall survival in malignant glioma is significantly prolonged by nerosurgical delivery of etoposide and Temozolomide from a thermo-responsive biodegradable paste. *Clin. Cancer Res. 25*, 5094.

DOI: 10.1021/acssynbio.9b00318
ACS Synth. Biol. 2019, 8, 2718−2725

(5) Stahelin, H. F., and Wartburg, A. V. (1991) The chemical and biological route from podophyllotoxin glucoside to etoposide: ninth cain memorial award lecture. *Cancer Res. 51*, 5−15.
(6) Zi, C. T., Yang, D., Dong, F. W., Li, G. T., Li, Y., Ding, Z. T., Zhou, J., Jiang, Z. H., and Hu, J. M. (2015) Synthesis and antitumor activity of novel per-butyrylated glycosides of podophyllotoxin and its derivatives. *Bioorg. Med. Chem. 23*, 1437−1446.
(7) Liu, H., Liao, J. X., Hu, Y., Tu, Y. H., and Sun, J. S. (2016) A highly efficient approach to construct (epi)-podophyllotoxin-4-*O*-glycosidic linkages as well as its application in concise syntheses of etoposide and teniposide. *Org. Lett. 18*, 1294−1297.
(8) Chang, A., Singh, S., Phillips, G. N., and Thorson, J. S. (2011) Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. *Curr. Opin. Biotechnol. 22*, 800−808.
(9) Peng, M., Shahzad, R., Gul, A., Subthain, H., Shen, S., Lei, L., Zheng, Z., Zhou, J., Lu, D., Wang, S., Nishawy, E., Liu, X., Tohge, T., Fernie, A. R., and Luo, J. (2017) Differentially evolved glucosyltransferases determine natural variation of rice flavone accumulation and UV-tolerance. *Nat. Commun. 8*, 1975.
(10) Marques, J. V., Kim, K. W., Lee, C., Costa, M. A., May, G. D., Crow, J. A., Davin, L. B., and Lewis, N. G. (2013) Next generation sequencing in predicting gene function in podophyllotoxin biosynthesis. *J. Biol. Chem. 288*, 466−479.
(11) Nagatoshi, M., Terasaka, K., Nagatsu, A., and Mizukami, H. (2011) Iridoid-specific glucosyltransferase from *Gardenia jasminoides*. *J. Biol. Chem. 286*, 32866−32874.
(12) Liang, D. M., Liu, J. H., Wu, H., Wang, B. B., Zhu, H. J., and Qiao, J. J. (2015) Glycosyltransferases: mechanisms and applications in natural product development. *Chem. Soc. Rev. 44*, 8350−8374.
(13) Yan, X., Fan, Y., Wei, W., Wang, P., Liu, Q., Wei, Y., Zhang, L., Zhao, G., Yue, J., and Zhou, Z. (2014) Production of bioactive ginsenoside compound K in metabolically engineered yeast. *Cell Res. 24*, 770−773.
(14) Caputi, L., Malnoy, M., Goremykin, V., Nikiforova, S., and Martens, S. (2012) A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. *Plant J. 69*, 1030−1042.
(15) Brazier-Hicks, M., Offen, W. A., Gershater, M. C., Revett, T. J., Lim, E. K., Bowles, D. J., Davies, G. J., and Edwards, R. (2007) Characterization and engineering of the bifunctional *N*- and *O*-glucosyltransferase involved in xenobiotic metabolism in plants. *Proc. Natl. Acad. Sci. U. S. A. 104*, 20238−20243.
(16) Brazier-Hicks, M., Gershater, M., Dixon, D., and Edwards, R. (2018) Characterization and engineering of the bifunctional *N*- and *O*-glucosyltransferase involved in xenobiotic metabolism in plants. *Plant Biotechnol. J. 16*, 337−348.
(17) Yonekura-Sakakibara, K., and Hanada, K. (2011) An evolutionary view of functional diversity in family 1 glycosyltransferases. *Plant J. 66*, 182−193.
(18) Nam, H., Lewis, N. E., Lerman, J. A., Lee, D. H., Chang, R. L., Kim, D., and Palsson, B. O. (2012) Network context and selection in the evolution to enzyme specificity. *Science 337*, 1101−1104.
(19) Leong, B. J., and Last, R. L. (2017) Promiscuity, impersonation and accommodation: evolution of plant specialized metabolism. *Curr. Opin. Struct. Biol. 47*, 105−112.
(20) Copley, S. D. (2017) Shining a light on enzyme promiscuity. *Curr. Opin. Struct. Biol. 47*, 167−175.
(21) Hasinoff, B. B., Chee, G. L., Day, B. W., Avor, K. S., Barnabe, N., Thampatty, P., and Yalowich, J. C. (2001) Synthesis and biological activity of a photoaffinity etoposide probe. *Bioorg. Med. Chem. 9*, 1765−1771.
(22) Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., Davis, B. G., Tarling, C. A., Ford, C. M., Bowles, D. J., and Davies, G. J. (2006) Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO J. 25*, 1396−1405.
(23) Yonekura-Sakakibara, K., Tohge, T., Matsuda, F., Nakabayashi, R., Takayama, H., Niida, R., Watanabe-Takahashi, A., Inoue, E., and Saito, K. (2008) Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in *Arabidopsis*. *Plant Cell 20*, 2160−2176.
(24) Yin, R. H., Han, K., Heller, W., Albert, A., Dobrev, P. I., Zazimalova, E., and Schaffner, A. R. (2014) Kaempferol 3-*O*-rhamnoside-7-*O*-rhamnoside is an endogenous flavonol inhibitor of polar auxin transport in *Arabidopsis* shoots. *New Phytol. 201*, 466−475.
(25) Yan, Y., Qi, B., Mo, T., Wang, X., Wang, J., Shi, S., Liu, X., and Tu, P. (2018) Research progress of rhamnosyltransferase. *Youji Huaxue 38*, 2281−2295.
(26) Janvary, L., Hoffmann, T., Pfeiffer, J., Hausmann, L., Topfer, R., Fischer, T. C., and Schwab, W. (2009) A double mutation in the anthocyanin 5-*O*-glucosyltransferase gene disrupts enzymatic activity in *Vitis vinifera* L. *J. Agric. Food Chem. 57*, 3512−3518.
(27) Kim, H. S., Kim, B. G., Sung, S., Kim, M., Mok, H., Chong, Y., and Ahn, J. (2013) Engineering flavonoid glycosyltransferases for enhanced catalytic efficiency and extended sugar-donor selectivity. *Planta 238*, 683−693.
(28) Song, C., Hong, X., Zhao, S., Liu, J., Schulenburg, K., Huang, F. C., Franz-Oberdorf, K., and Schwab, W. (2016) Glucosylation of 4-hydroxy-2,5-dimethyl-3(2H)-furanone, the key strawberry flavor compound in strawberry fruit. *Plant Physiol. 171*, 139−151.
(29) Jurcik, A., Bednar, D., Byska, J., Marques, S. M., Furmanova, K., Daniel, L., Kokkonen, P., Brezovsky, J., Strnad, O., Stourac, J., Pavelka, A., Manak, M., Damborsky, J., and Kozlikova, B. (2018) CAVER Analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. *Bioinformatics 34*, 3586−3588.
(30) Kress, N., Halder, J. M., Rapp, L. R., and Hauer, B. (2018) Unlocked potential of dynamic elements in protein structures: channels and loops. *Curr. Opin. Chem. Biol. 47*, 109−116.
(31) Brezovsky, J., Babkova, P., Degtjarik, O., Fortova, A., Gora, A., Iermak, I., Rezacova, P., Dvorak, P., Smatanova, I. K., Prokop, Z., Chaloupkova, R., and Damborsky, J. (2016) Engineering a *de novo* transport tunnel. *ACS Catal. 6*, 7597−7610.
(32) Yang, M., Fehl, C., Lees, K. V., Lim, E. K., Offen, W. A., Davies, G. J., Bowles, D. J., Davidson, M. G., Roberts, S. J., and Davis, B. G. (2018) Functional and informatics analysis enables glycosyltransferase activity prediction. *Nat. Chem. Biol. 14*, 1109−1117.
(33) Jia, K. Z., Zhang, Q., Sun, L. Y., Xu, Y. H., Li, H. M., and Tang, Y. J. (2016) *Clonostachys rosea* demethiolase STR3 controls the conversion of methionine into methanethiol. *Sci. Rep. 6*, 21920.
(34) Asada, K., Salim, V., Masada-Atsumi, S., Edmunds, E., Nagatoshi, M., Terasaka, K., Mizukami, H., and De Luca, V. (2013) A 7-deoxyloganetic acid glucosyltransferase contributes a key step in secologanin biosynthesis in *Madagascar Periwinkle*. *Plant Cell 25*, 4123−4134.
(35) Weiss, G. A., Watanabe, C. K., Zhong, A., Goddard, A., and Sidhu, S. S. (2000) Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. U. S. A. 97*, 8950−8954.
(36) Pal, G., Fong, S. Y., Kossiakoff, A. A., and Sidhu, S. S. (2005) Alternative views of functional protein binding epitopes obtained by combinatorial shotgun scanning mutagenesis. *Protein Sci. 14*, 2405−2413.
(37) Sidhu, S. S., and Kossiakoff, A. A. (2007) Exploring and designing protein function with restricted diversity. *Curr. Opin. Chem. Biol. 11*, 347−354.
(38) Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L., and Schwede, T. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res. 42*, W252−W258.

DOI: 10.1021/acssynbio.9b00318
ACS Synth. Biol. 2019, 8, 2718−2725