OXFORD

Chemical Biology

Semirational design and engineering of grapevine glucosyltransferases for enhanced activity and modified product selectivity

Rakesh Joshi[2,3,4], Johanna Trinkl[3], Annika Haugeneder[3], Katja Härtl[3], Katrin Franz-Oberdorf[3], Ashok Giri[4], Thomas Hoffmann[3] and Wilfried Schwab 🟢 [1,3]

[2] Institute of Bioinformatics and Biotechnology (IBB), Savitribai Phule Pune University (Formerly University of Pune), Ganeshkhind Road, Pune 411007, Maharashtra, India, [3]Biotechnology of Natural Products, Technische Universität München, 85354 Freising, Liesel-Beckmann-Str. 1, Germany, and [4]Division of Biochemical Sciences, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune 411 008, India

[1]To whom correspondence should be addressed: Tel: +498161712912; Fax +498161712950; e-mail: wilfried.schwab@tum.de

Abstract

Uridine diphosphate-dependent glycosyltransferases (UGTs) catalyze the transfer of a diversity of sugars to several acceptor molecules and often exhibit distinct substrate specificity. Modulation of glycosyltransferases for increased catalytic activity and altered substrate or product specificity are the key manipulations for the biotechnological use of glycosyltransferases in various biosynthetic processes. Here, we have engineered the binding pocket of three previously characterized *Vitis vinifera* glycosyltransferases, UGT88F12, UGT72B27 and UGT92G6, by structure-guided in silico mutagenesis to facilitate the interactions of active site residues with flavonol glucosides and thus modify substrate specificity and activity. Site-directed mutagenesis at selected sites, followed with liquid chromatography–mass spectrometry based activity assays, exhibited that mutant UGTs were altered in product selectivity and activity as compared to the wild-type enzymes. Mutant UGTs produced larger amounts of flavonol di-monosaccharide glucosides, which imply that the mutations led to structural changes that increased the volume of the binding pocket to accommodate a larger substrate and to release larger products at ease. Mutants showed increased activity and modified product specificity. Thus, structure-based systematic mutations of the amino acid residues in the binding pocket can be explored for the generation of engineered UGTs for diverse biotechnological applications.

Key words: flavonol, glycosyltransferase, in silico analysis, mutagenesis, selectivity

Introduction

Natural products produced by bacteria, fungi and plants are a rich source of bioactive metabolites. It is reported that bioactivities of these metabolites depend on regio- and stereospecific decoration of sugar moieties on its accessible groups (Newman et al. 2003; Luzhetskyy et al. 2008). Glycosyltransferases (GTs) catalyze this glycosidic bond formation between sugar moieties and natural products like polyphenols, alkaloids and terpenoids in plants (Liang et al. 2015). GTs transfer the sugar molecules from an activated donor molecule, e.g. uridine diphosphate glucose (UDP-Glc) to the nucleophilic oxygen, nitrogen, sulfur or carbon of the acceptor molecule (Lairson et al. 2008). GTs represent a superfamily of enzymes with over 105 structurally conserved classes consisting of numerous sequences with a high level of sequence diversity (Liang and Qiao 2007; www.cazy.
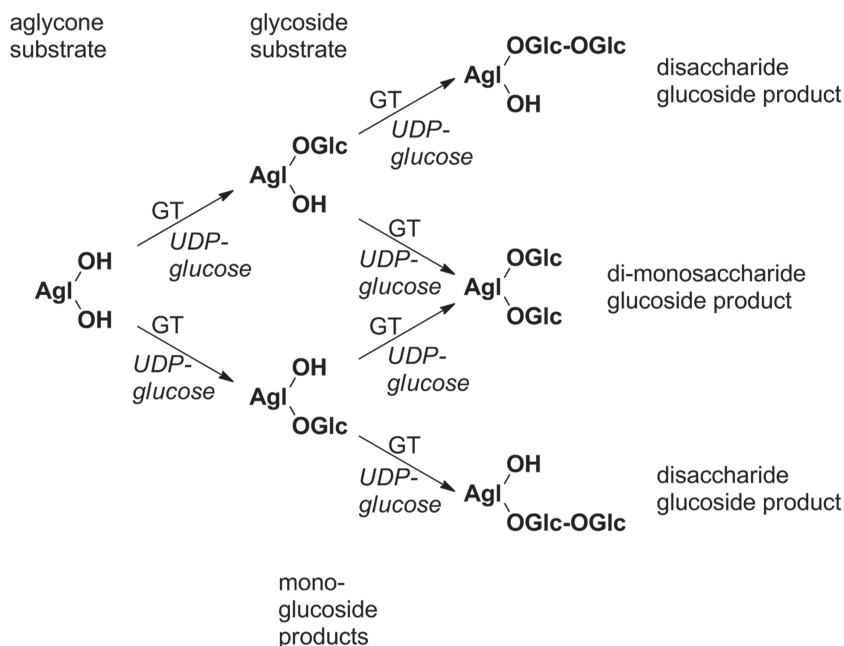
Fig. 1. Substrates and products of glucosyltransferases. Aglycones as well as glycosides were used as substrates for UGT88F16, UGT72B27 and UGT92G6. Glycoside substrates yield disaccharide and/or di-monosaccharide glucoside products. Agl: aglycone, Glc: glucose, GT: glycosyltransferase, UDP-glucose: uridine diphosphate glucose.

org/GlycosylTransferases.html). The majority of the family members adopt one of three conserved structural folds termed GT-A, GT-B and GT-C. Most of the plant GTs belong to the GT-1 family (UDP sugar-dependent UGTs) consisting of GT-B fold enzymes according to CAZy classification (Couthino et al. 2003; www.cazy.org/GlycosylTransferases.html). GT-B shows a canonical structure with two juxtaposed β/α/β Rossmann folds connected by a linker, and the reaction centre lies in the cleft between the two domains. At the N-terminus, the acceptor substrate recognition occurs, while the domain for the interaction with the nucleotide sugar donor is located at the C-terminus (Offen et al. 2006). In plants, GTs in class 1 with GT-B fold are majorly involved in recognition of a large array of different small molecules, including terpenoids, alkaloids, cyanohydrins and glucosinolates, as well as flavonoids and phenylpropanoids. Due to this substrate diversity, plants possess multiple GT genes. In grapevine, around 210 GT genes have been identified to date (http://genomes.cribi.unipd.it/grape/).

Polyphenols, including flavonoids and phenylpropanoids, are an essential group of metabolites contributing to the quality of grapevine and other fruit. In red wine, polyphenols are associated with color, bitterness and astringency. A growing body of research data also suggests that polyphenols help to reduce the risk of several diseases and thus affect human health (Chuang et al. 2011; Giovinazzo et al. 2015). Quercetin and myricetin are the main flavonols found in grapes but also kaempferol, laricitrin, isorhamnetin and syringetin are present (Mattivi et al. 2006). The majority of these flavonols occur in glycosylated form. In *Arabidopsis thaliana*, out of 91 recombinant enzymes that were analyzed for in vitro activity toward the flavonol quercetin, 29 were capable of glucosylating the substrate (Lim et al. 2004). Monoglucosides and di-mono-saccharide glucosides were produced (Figure 1; Kim et al. 2013). The position, the number of carbohydrates and the type of sugars bound to the aglycone contribute significantly to the diversity of these metabolites in the plant.

The aglycones are usually poorly soluble in water and unstable while some of them are even toxic (Ford and Høj 1998; Bowles et al. 2006; Bönisch et al. 2014a, b; Song et al. 2018). The mechanism of UGT-catalyzed glucosylation of natural products and their products have attracted much attention because glycosylation renders the substrates more water soluble, stable and reduce their toxicity (Schwab et al. 2015). Research on glycosylated products from plants has gained momentum since more and more genome sequences become publicly available, and plant genomes contain a multitude of UGT sequences. The sugar donor and receptor binding of a few plant UGTs have been elucidated, including a flavonoid GT from *Vitis vinifera* (Offen et al. 2006). Thus, information about binding residues can be utilized for tuning the regio-selectivity of UGT for various substrates. Recently, it has been shown that multiple points of mutants not only maintained the high product specificity but also significantly improved the catalytic efficiency in GT-1 from *Bacillus cereus* (BcGT1) (Chui et al. 2016). We concluded that similar approaches could be applied to modulate the substrate specificities and activity levels of *V. vinifera* UGTs as in silico analyses of plant flavonoid UGTs have revealed binding site features and substrate selectivity (Sharma et al. 2014).

Only recently, UGT92G6 from grapevine was shown to produce mono- and diglucosides in vitro from distinct compounds (Huang et al. 2018). However, the level of diglucosides was not elevated after agroinfiltration in *Nicotiana benthamiana* leaves but caffeoyl glucosides were produced. In addition, a resveratrol UGT (UGT72B27) was identified among three candidates (UGT88F12, UGT72B27 and UGT92G6) to be responsible for the production of phenolic glucosides in grapevine (Härtl et al. 2017). The three UGTs from *V. vinifera* were selected (Supplemental data Figure S1) for further analyses because they showed contrasting substrate tolerance and catalytic activity and produced diglucosides from glycoside substrates in addition to monoglucosides from aglycone substrates (Figure 1; Supplemental data Figure S2).

In the present study, we used in silico approaches for understanding molecular interaction between *V. vinifera* UGTs (VvUGTs) and substrates, followed by structure-based mutagenesis, and analyzed its implication on substrate interaction. We aimed to modify selectivity without loss of activity. The obtained results were validated by site-directed amino acid mutations in the substrate-binding pockets. UGT activity of the mutant proteins was analyzed experimentally by liquid chromatography–mass spectrometry (LC-MS) and results were compared with those of the wild-type UGTs for differential substrate specificity and activity. The mutagenesis-based semirational design of UGTs provides molecular insights into the structural basis of substrate activity and product specificity of VvUGTs and opens up possibilities for improving substrate activity and product specificity.

Results

VvUGTs showed structures cognate to GT-B fold

The tertiary structures of UGT proteins from grape were modeled with Modeller software to study the interactions that favor the formation of the protein–ligand complex. Crystal structures of GT1 from *V. vinifera* (PDB-2C1Z; Offen et al. 2006) with resolutions of 1.90 A° were used for modeling studies because they showed a coverage of more than 90% and shared identities of 35, 65 and 29% with UGT88F12, UGT72B27 and UGT92G6, respectively. The initial models were evaluated for their stereochemical parameters (Supplemental data Figures S3–S5). More than 99% of all residues were found to be in allowed regions of the Ramachandran plot. The root mean square deviation (RMSD) of the C-alpha atoms of the models and the templates was less than 1 Å. The model refinement phase involved reprocessing the initial models by adding hydrogens, assigning bond order and filling missing loops and side chains, followed by restrained minimization by applying the constraint to converge the non-hydrogen atoms to an RMSD of 0.3 Å using the OPLS 2005 force field. After that, the models were further subjected to 500 steps of steepest descent energy minimization followed by 1000 steps of conjugate gradient energy minimization using the same force field. Predicted models of UGTs showed a canonical structure of GT-B fold with two "Rossmann-like" (β/α/β) domains connected with the cleft (Figure 2A–C). At the N-terminus, carrying the catalytic His-residue, the protein interacts with the acceptor substrate (sugar acceptor domain), while the C-terminus contains the highly conserved plant secondary product glycosyltransferase (PSPG) motif, which binds the activated nucleotide sugar (sugar donor domain). These energy-minimized models were further used for docking studies and interaction analysis. Structural comparison between UGT88F12, UGT72B27 and UGT92G6 showed that sugar donor and substrate entry site is relatively more accessible in UGT72B27, followed by UGT88F12 and UGT92G6. Furthermore, it showed that there are differences in the loop conformations. Sequence and structural analysis showed that although the overall fold and sugar donor as well as substrate-binding conformation is mostly conserved (Supplemental data Figure S6), the positional variation of amino acids might leads to differential substrate specificity.

Mutation in binding site residues of VvUGTs might lead in altered substrate preference

Preprocessed structures of the ligands were docked with the predicted structures of VvUGTs using AutoDock Vina (Trott et al. 2010). Binding poses of the sugar donor with maximum interactions with the conserved residues of the PSPG motif of the GTs were selected for further analysis (Figure 2A–C). The bound ligand (UDP-sugar) of the template was used to mark the binding site by creating a grid around it. The flavonol kaempferol was docked in the acceptor-binding pocket of the protein–UDP sugar complex by placing the grid around the reference ligand, i.e. kaempferol and kaempferol-glucoside (Figure 2D–F). Docking studies showed that the environment surrounding the acceptor is conserved or similar for all docked structures of UGT88F12.

Binding site residues of UGT88F12, UGT72B27 and UGT92G6 were mutated in silico and computationally analyzed for their interaction with various aglycone and glycoside substrates (Figure 3). Mutations in some binding site residues led to significant changes in binding scores as compared to wild-type protein. Heatmaps reflect the change in binding scores in mutated proteins (Figure 3). Mutations R82A, R82K, D116E, Y147A and Y147S in UGT88F12 led to remarkable changes in glycoside substrate (flavonoid glucosides) binding (Figure 3D). Similarly, mutations like I81T, S273C, F311S and F311Y in UGT72B27 resulted in alteration in flavonoid glycoside substrate specificities (Figure 3E). In UGT92G6, also point mutation at a single residue, i.e. Y195S/F, caused a significant change in flavonoid glycoside substrate specificity (Figure 3F). Effects of mutations on the binding scores of aglycone substrates were less pronounced except for S139T and Y195S in UGT88F12 and UGT92G6, respectively (Figure 3A and C).

Change in substrate activity and product specificity because of site-directed mutagenesis

Mutations selected after in silico screening (UGT88F12_R82A, UGT72B27_I81T, UGT72B27_S273C, UGT72B27_F311Y, UGT92G6_Y195F and UGT92G6_Y195S) were generated in vitro by site-directed mutagenesis to test their predicted effects on enzyme specificity and activity. In the LC-MS-based assay, aglycones and glycosides served as acceptor substrates while UDP glucose was the donor substrate (Figure 4). The wild-type (WT) enzymes accepted simple phenolics and benzyl alcohol as aglycone substrates (Figure 4A). Furthermore, geraniol, citronellol, linalool, farnesol, menthol and nerol (non-aromatic terpenoids), octanol, decanol, *cis*-3-hexanol and 2-phenylethanol (primary alcohols) and the two anthocyanidins pelargonidin and malvidin were tested in the in vitro assays. However, none of the UGTs showed a significant activity towards one of the previously named substrates. Regarding the accepted substrates, the mutant UGT72B27_F311Y showed a significantly higher enzymatic activity than the WT enzyme for almost all aglycones (Figure 4A). Mutant UGT72B27_I81T displayed only a slightly different substrate profile, whereas the amino acid exchange in UGT72B27_S273C rendered the enzyme virtually inactive. The mutants of UGT92G6 and the WT enzyme showed similar catalytic activities and substrate preferences towards aglycone substrates (Figure 4A). A moderately altered activity was observed for UGT88F12_R82A, which produced higher amounts of kaempferol and quercetin glucosides (-glc) but lower levels of vanillin- and eugenol-glc than the WT protein.

Using the flavonols kaempferol and quercetin as aglycone substrates, not only the formation of different monoglucosides (e.g. kaempferol-3-glc and -7-glc) could be observed (Figure 5A and B) but also the production of diglucosides (Figure 5C and D). Mutant UGT88F12_R82A produced a significantly higher amount of quercetin monoglucosides (Figure 5A) while mutants
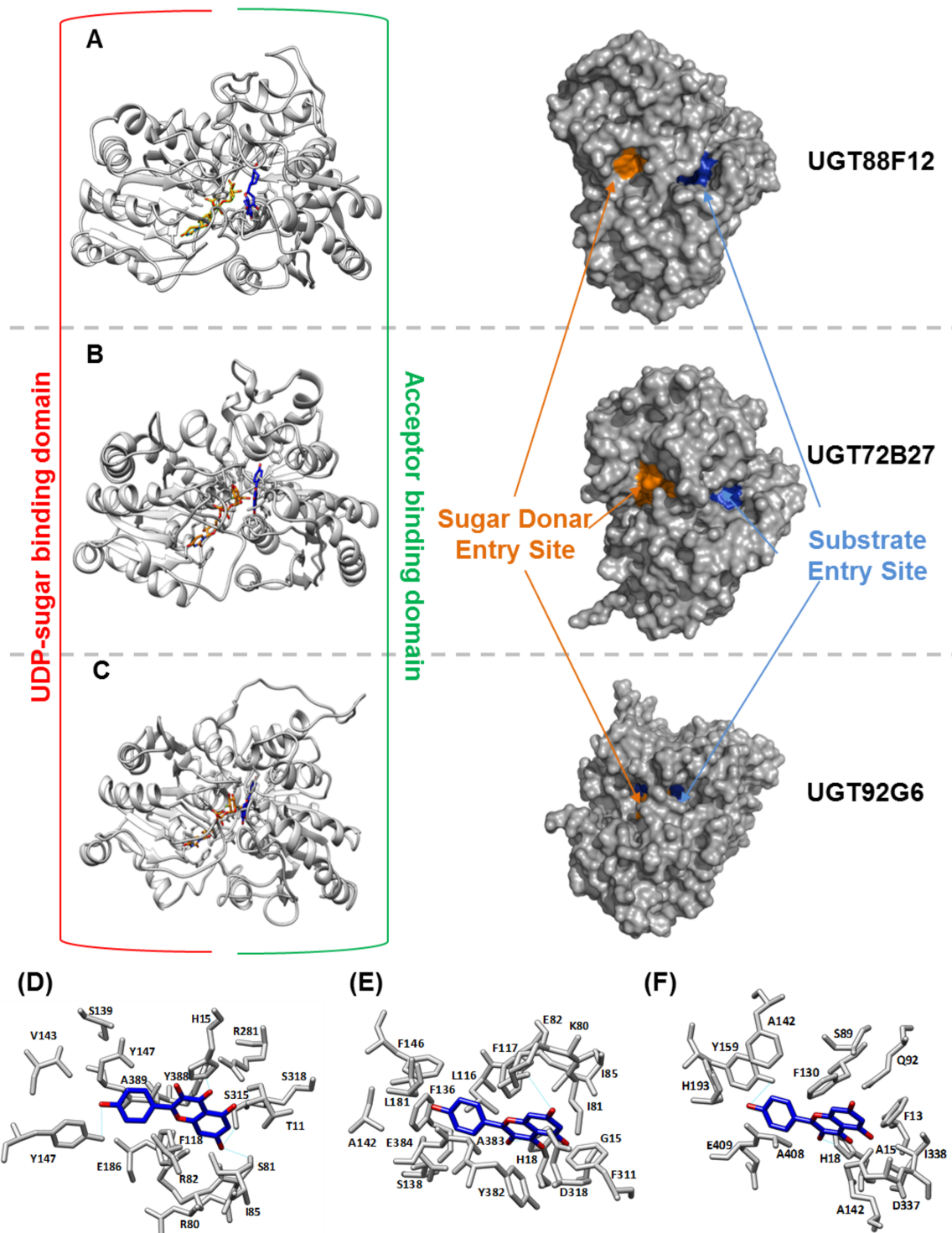
Fig. 2. Structural features of UGTs. Cartoon representations of (A) UGT88F12, (B) UGT72B27 and (C) UGT92G6 show distinct UDP sugar and acceptor-binding domains. The UDP sugar-binding domain containing the plant secondary product glycosyltransferase box (PSPG box) accepts activated nucleotide sugars (sugar donor), while the second substrate (sugar acceptor) binds at the acceptor-binding domain. Surface representations of these structure show that each protein has two entry sites, one for the sugar and another for substrate entry. (D–F) residues from the acceptor-binding domain interacting with candidate substrate (kaempferol).

UGT88F12_R82A, UGT72B27_F311Y and UGT92G6_Y195S formed larger quantities of quercetin diglucosides (Figure 5C). The same trend could be observed for the formation of kaempferol diglucosides (Figure 5D). These results indicate that glucoside, which is formed by the GTs in the first step, is used as a substrate in a second reaction step to form diglucosides.
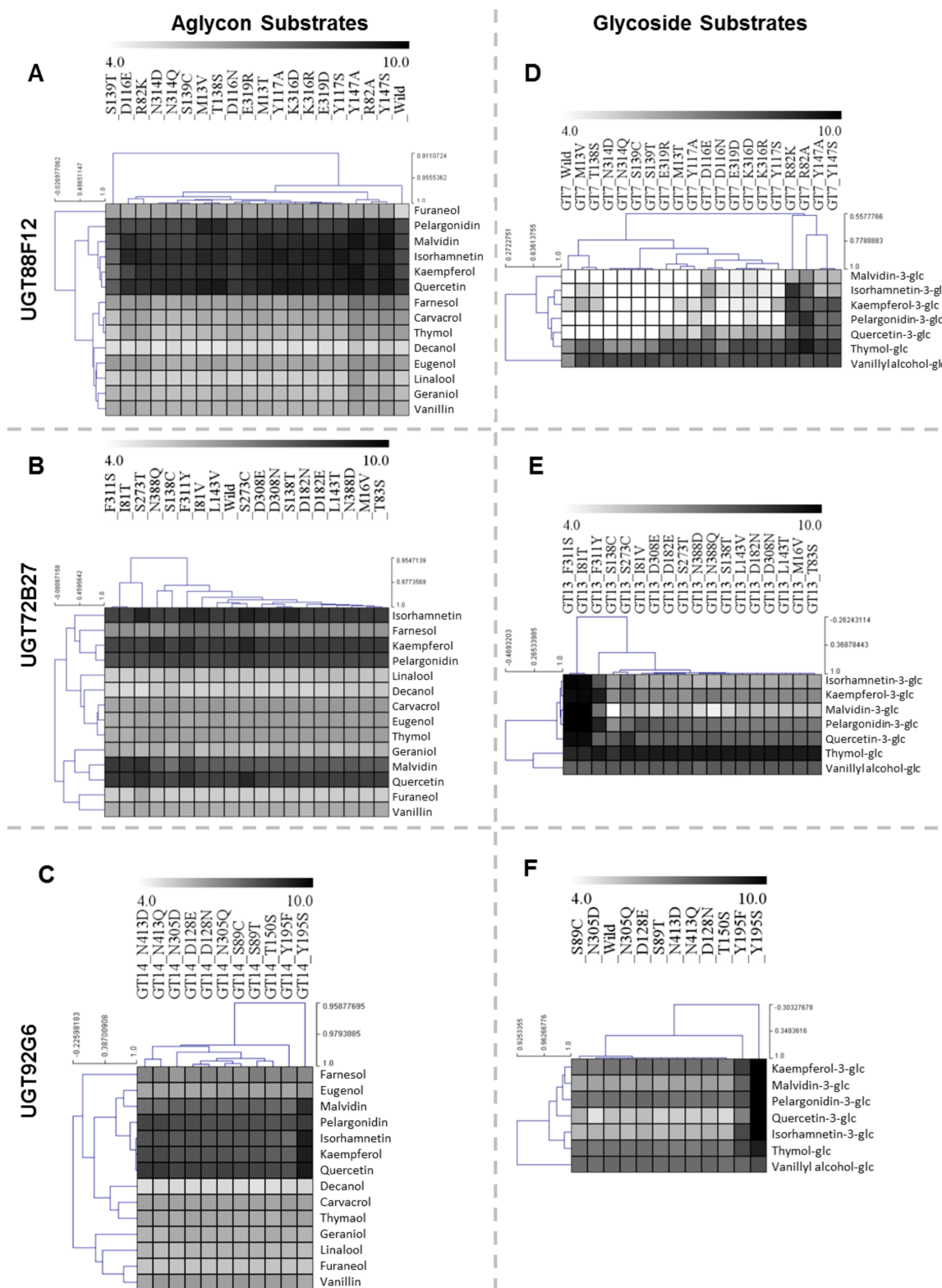
Fig. 3. Heatmap analysis of differential binding energy (converted in positive values for ease of representation) between enzyme and substrate of wild and mutant UGTs. White color represents weak binding, while grey represents stronger binding. Comparison of interaction scores for UGT88F12, UGT72B27 and UGT92G6 with aglycon (A–C) and glycoside substrates (D–F).
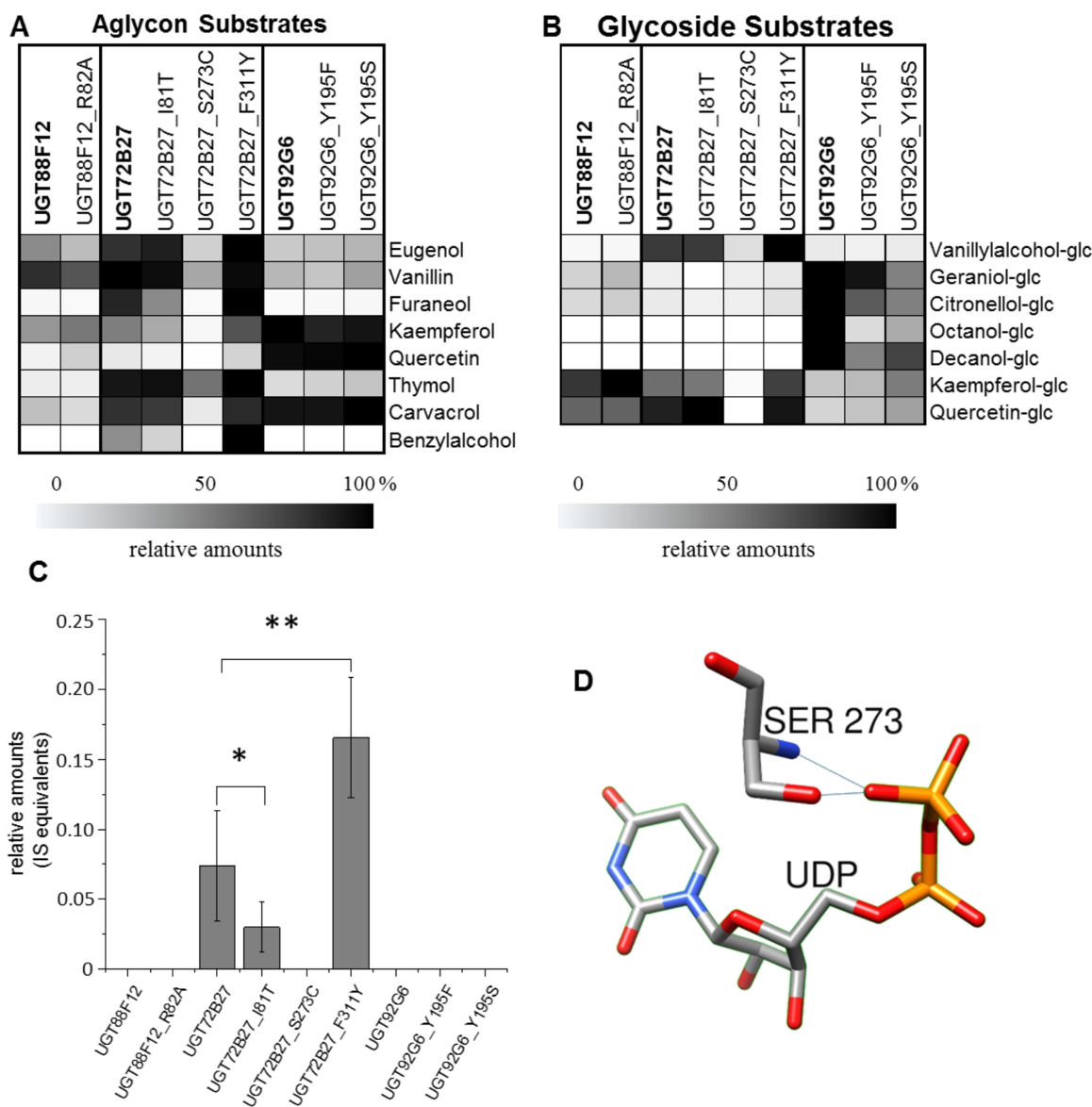
Fig. 4. Effect of mutations on enzyme activity. Product amounts produced from aglycone (A) and glycoside substrates (B) by purified wild-type and mutant proteins were determined via LC-MS and expressed as relative amount of the internal standard. Amount of products formed from benzyl alcohol (C) and proposed interaction of Ser 273 and UDP in the calculated active site of UGT72B27 (D). Values represent means of three biological and three technical replicates, and error bars show the standard deviations. Significant differences were determined by t-test ($P \leq 0.05$ (*) and $P \leq 0.01$ (**)).

In contrast, UGT72B27_I81T produced fewer amounts of kaempferol-mono- and -di-glucosides than the WT enzyme (Figure 5B and D) and UGT72B27_S273C showed significantly reduced GT activity for all substrates (Figure 5A–F). Strikingly, the mutant UGT72B27_F311Y yielded the highest quantity of quercetin- and kaempferol-di-glucosides (Figure 5C and D). In the latter case, especially the production of kaempferol diglucoside 2 was significantly increased (Figure 5D). The data show that the substrate specificity was virtually not affected by the different mutations (Figure 4A), except for UGT72B27_S273C, which was inactive, but the enzyme activity and product specificity were altered by specific exchanges of individual amino acids (Figure 5A–F).

Several glucosides were also accepted as substrates by the tested enzymes and mutant GTs (Figure 4B). Further tested glucoside substrates without significant activity were thymol glucoside and car-

vacrol glucoside. UGT92G6 and its derived mutants glucosylated the terpenoid glucosides geraniol-glc and citronellol-glc as well as alkanol-glc, octanol-glc and decanol-glc. Since free binding sites were present only on the sugar moiety, glycosides containing a disaccharide were generated. Interestingly, the aglycones geraniol, citronellol, octanol and decanol did not serve as substrates under identical conditions. In contrast, kaempferol-glc and quercetin-glc that act as GT substrates also have free nucleophilic acceptor groups on the aglycone residue in addition to the already bound glucose. To clarify the binding site of the second carbohydrate residue, additional experiments were performed using $^{13}C_6$-labeled UDP-glucose as donor substrate. UGT88F12 and UGT88F12_R82A formed three kaempferol diglucosides (Figure 5F), which showed distinct product ion spectra MS2 when kaempferol-3-O-glucoside and $^{13}C_6$-UDP-glucose were used as acceptor and donor substrate, respectively (Figure 6A–C).
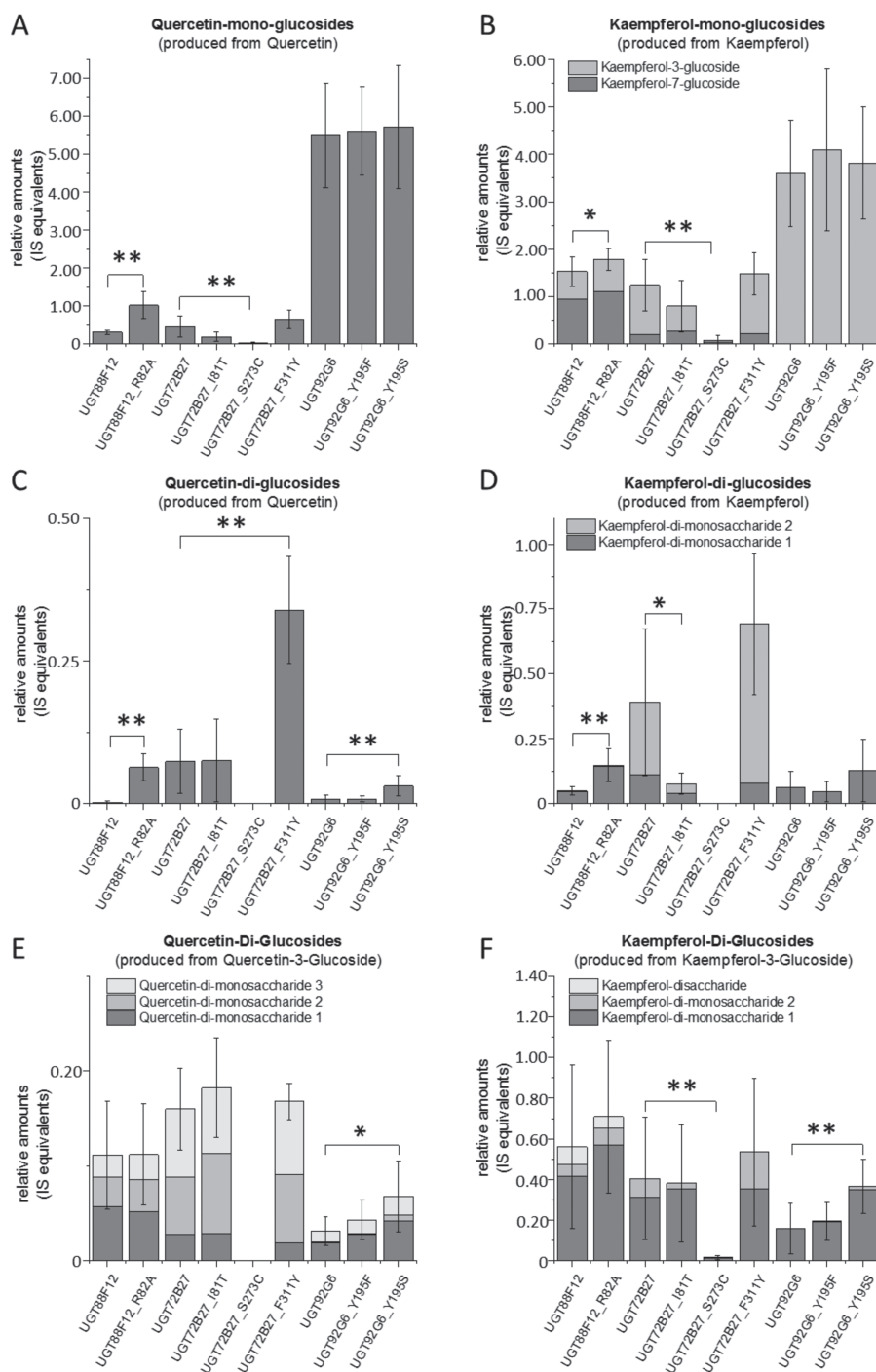
Fig. 5. Formation of monoglucosides and diglucosides produced from aglycons and glycoside substrates. Quercetin (A, C) and kaempferol (B, D) served as aglycon substrates and quercetin-3-glucoside (E) and kaempferol-3-glucoside (F) as glycoside substrates. Values represent means of three biological and three technical replicates and error bars show the standard deviations. Significant differences were determined by t-test ($P \leq 0.05$ (*) and $P \leq 0.01$ (**)).

As the unlabeled ($-m/z$ 162) and the newly bound $^{13}C_6$-glucose residue ($-m/z$ 168) were cleaved off singly in the MS2 spectra of products 1 and 2 (Figure 6B), they were identified as kaempferol-di-mono-glucosides (Figure 6D). The third product produced by UGT88F12 and UGT88F12_R82A (Figure 6A) showed a low intensity of the fragment ion m/z 447, indicating the loss of a bound $^{13}C_6$-glucose residue ($-m/z$ 168). In contrast, the fragment ion m/z 285 indicating the cleavage of a $^{13}C_6$-labeled di-hexoside is present

with high intensity (Figure 6B). Therefore, product 3 was tentatively assigned as kaempferol-disaccharide-glycoside (Figure 6D; Ablajan et al. 2006). Similarly, the production of quercetin-di-mono-glucosides and disaccharide-glycosides was analyzed by LC-MS analysis.

When using a terpineol or alkanol monoglucoside as a substrate, both UGT92G6 mutants produced fewer disaccharides compared to the WT (Figure 4B). However, UGT92G6_Y195S produced larger quantities of kaempferol-di-mono-glucosides (Figure 5D and F).
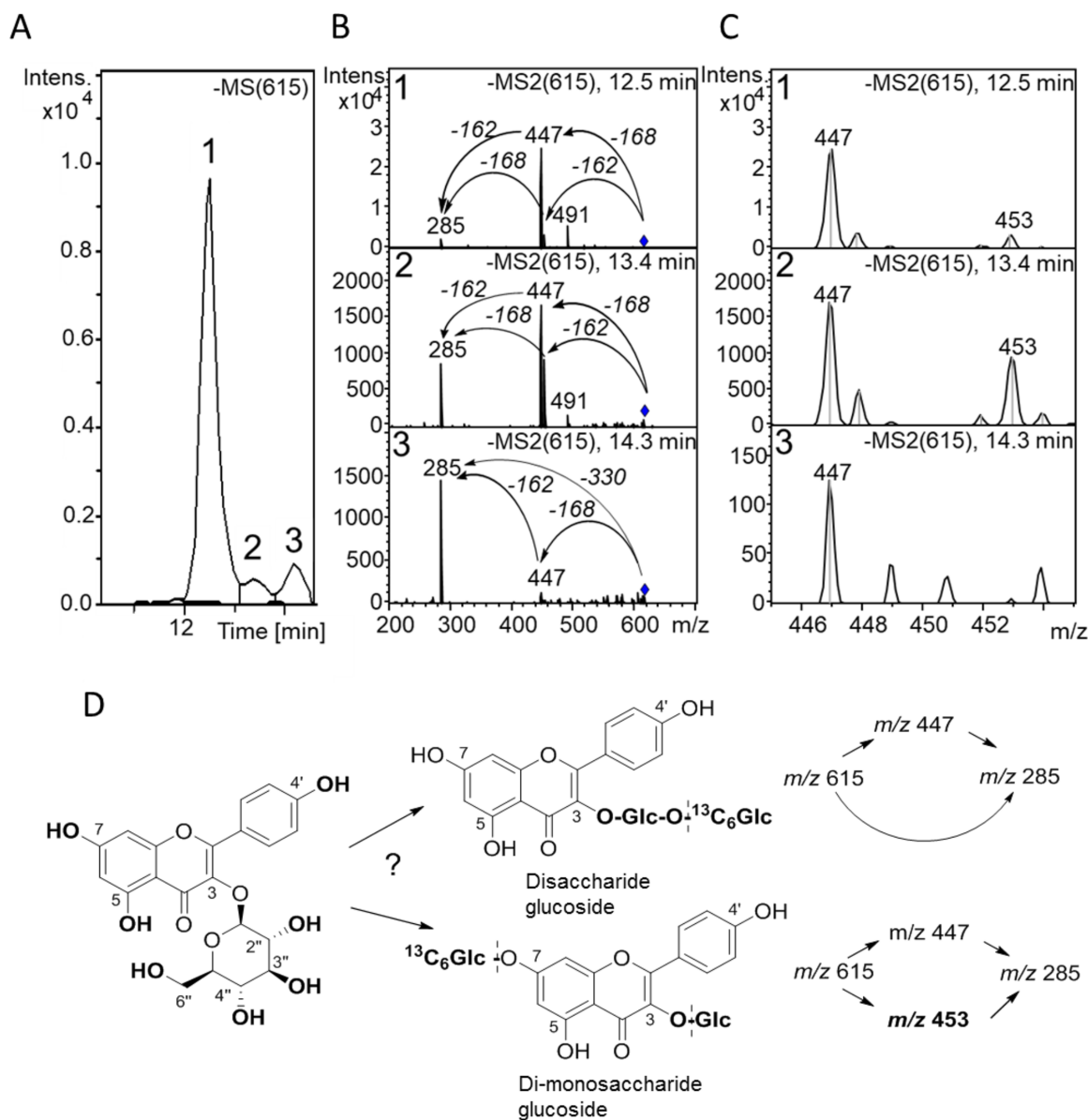
Fig. 6. Formation of kaempferol diglucosides from kaempferol-3-glucoside by UGT88F (A). UDP-^{13}C-glucose was used as sugar donor to localize the binding site of the second glucose residue. Product ion spectra (B), enlarged section (*m/z* 446–454) (C) and postulated fragmentation pattern (D).

Mutations in UGT88F12 did not alter enzyme activity towards glycoside substrates, except for UGT88F12_R82A, which showed increased activity for kaempferol glucoside (Figure 4B and 5F). The mutant UGT72B27_S273C was inactive while mutants UGT72B27_I81T and UGT72B27_F311Y displayed modified product specificity in comparison to the wild type (Figure 5F). UGT72B27 and its derived mutants exclusively catalyzed the formation of di-mono-glucosides for all tested glucoside substrates.

Discussion

Recently, three highly promiscuous *V. vinifera* UGTs were characterized that produced both mono- and diglucosides from aglycone and glycoside substrates, respectively (Härtl et al. 2017; Huang et al. 2018). Because these proteins already acquired the ability to transfer a second sugar onto glucosides, this study aimed to raise

the diglucoside-forming catalytic activity by semirational design and protein engineering.

Structure homology modeling combined with ligand docking studies (Figure 2) and in silico mutation analysis were performed to optimize the interaction of the enzymes with various aglycone and glucoside substrates (Figure 3). One, two and three mutant proteins of UGT88F12, UGT72B27 and UGT92G6, respectively, were generated and their catalytic properties analyzed (Figure 4). The three WT enzymes and their derived mutant proteins glucosylated simple phenolic compounds, terpenols, benzyl alcohol and furaneol and thus confirmed recent results (Härtl et al. 2017; Huang et al. 2018).

The mutated amino acids did not significantly alter the product yield when aglycone substrates were converted, except for benzyl alcohol in the case of UGT72B27_F311Y and UGT72B27_S273C, which rendered the enzyme inactive (Figure 4). However, the amount of kaempferol- and quercetin-di-glucosides formed from their respec-

tive monoglucosides by UGT88F12_R82A, UGT72B27_F311Y and UGT92G6_Y195S were significantly increased compared to the levels produced by the WT enzymes (Figures 4 and 5). Thus, increasing the interaction of VvUGTs with flavonoid glucoside substrates by in silico semirational design provided mutant proteins with superior GT enzymatic activity toward monoglucosides.

Similarly, mutants UGT88F12_R82A, UGT72B27_F311Y and UGT92G6_Y195S produced larger amounts of diglucosides from the aglycone substrates kaempferol and quercetin (Figure 5C and D). The mutations facilitate the additional transfer of a glucose molecule onto the primary monoglucoside, suggesting that monoglucosides fit better into the active site of the mutant proteins than into the WT enzymes. Recently, in a similar fashion, the regio-selectivity of UGTBL1 from *Bacillus licheniformis* was switched from 4′-OH to 3-OH glucosylation by semirational design including structure-guided alanine scanning and saturation mutation (Fan et al. 2018).

Among all tested mutations, UGT72B27_S273C holds a unique position as it showed rather weak activity towards all substrates. This result highlights the importance of serine at position 273. It plays a vital role in the stabilization of the UDP glucose in the binding pocket (Figure 4D). Exchanging serine by the less polar cysteine seems to have a negative effect on the stabilization of the UDP glucose, which resulted in reduced enzyme activity. Thus, when analyzing the interactions of substrate molecules with amino acids in the active site by in silico studies, the effects on the second substrate must also be considered and were not predicted by the used algorithms.

Structure- or sequence-guided GT engineering includes site-specific mutagenesis and the construction of chimeric GTs (Chang et al. 2011). Success in the production of functionally active chimeric GTs typically requires a high degree of structural and sequence homology between the GTs to be combined. Few general rules are known for engineering sugar donor specificity, e.g. a single R140W exchange changed the glucuronosyltransferase VvGT5 to a bifunctional glucosyl/galactosyltransferase (Ono et al. 2010). However, these amino acids are not conserved in GTs. General rules are also lacking for engineering the acceptor specificity probably due to the divergent nature of acceptors and acceptor binding. Structure-guided mutagenesis of UGT71G1 revealed two mutations (F148V and Y202A), which drastically changed the regioselectivity of quercetin glycosylation (He et al. 2006).

Our results indicate that the selected mutations affect enzymes activity and product specificity, while the substrate specificity was unchanged. Semirational design to enhance active site interaction in GT enzymes with flavonoid glucosides was successful as the modifications resulted in significantly increased quantities of diglucoside products for three structurally distinct GTs from *V. vinifera*.

The study showed that biocatalysts have a certain plasticity in their active site architecture, since several amino acid exchanges proposed by *in-silico* analysis, which should influence substrate specificity and activities, did not lead to changes in the biochemical properties of the enzymes. In addition, the experiments yielded amino acids essential for UGT catalytic activity and helped to identify positions, which affect the volume of the active site and therefore the size of potential substrates. Finally, amino acids were found that modulate the catalytic activity. Overall, semirational design contributed to explain the substrate, reaction and product preferences of biotechnologically relevant enzymes and will in the future provide strategies to improve different types of biocatalysts.

Materials and methods

Molecular modeling and docking of VvUGT proteins

Homology modeling was performed using Modeller software (https://salilab.org/modeller/) to predict the three-dimensional structure of VvUGTs. X-ray crystal structures of plant glycosyltransferase were used as a template for homology modeling (Webb et al. 2014). The predicted model was energy minimized using the GROMOS 43BI force field and assessed for its quality with Ramachandran plot analysis. The initial models were evaluated for stereochemical quality of the protein backbone and side chains using PROCHECK (www.ebi.ac.uk/thornton-srv/software/PROCHECK/). The environment of the atoms in the protein model was checked by the ERRAT server (http://servicesn.mbi.ucla.edu/ERRAT/). Verify3D plot showed the compatibility of the 3D structure concerning the protein sequence (http://servicesn.mbi.ucla.edu/Verify3d/). Errors in the model structures were also checked with the ProSA server (https://prosa.services.came.sbg.ac.at/prosa.php). After model validation, the initial models were refined using impref minimization of protein preparation wizard and Impact 5.8 minimization. These energy-minimized final models were further used for the binding studies with their substrates (Sharma et al. 2014).

Molecular docking analysis of VvUGT and flavonol substrate binding

The 2D sketcher utility of Maestro 9.3 (www.schrodinger.com/maestro) was used to design the structures of ligand molecules in two dimensions, which were then converted into three-dimensional structures in mol2 format. Before proceeding with the docking studies, water and other heteroatom groups were removed from the protein structures using protein preparation utility of Maestro. Hydrogens were added subsequently to carry out restrained minimization of the models. The minimization was done using Maestro in which the heavy atoms were restrained such that the strains generated upon protonation could be relieved. The RMSD of the atomic displacement for terminating the minimization was set as 0.3 A°. Similarly, ligands were refined with the help of LigPrep 2.5 (www.schrodinger.com/ligprep) to define their charged state and enumerate their stereo isomers. Ligands and receptor files were optimized for docking studies and converted from.pdb to.pdbqt format using AutoDock 4.2 version (Morris et al., 2009). Binding sites in VvUGT models were predicted using superimposition with the structure of GT co-crystallized with substrates.

Docking studies were performed using AutoDock 4.2 version (http://autodock.scripps.edu/). The grid was set around predicted binding site residue of VvUGT with dimensions of $20 \times 20 \times 20$ Å. A grid was made either by taking the reference ligand or by selecting the active site residues involved in the binding of the substrate. Flexible ligand docking was carried out using the standard precision option. The sugar donor was docked in the active site by creating a grid around the bound reference ligand, i.e. U2F (uridine-5′-diphosphate-2-deoxy-2-fluoro-α-D-glucose). Final products of the reaction were also docked in the *V. vinifera* GT crystal structure. The docking parameters were configured on a Lamarckian genetic algorithms calculation of 10,000 runs. The obtained conformations were later summarized, collected and extracted by using AutoDock Tool. The geometry of resulting complexes was studied using the PyMOL Molecular Viewer utility (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC).

In silico screening of mutations and interaction analysis

Virtual alanine scanning was done using the ABS-Scan server (Anand et al. 2013). Residues important for interaction were predicted. These residues were further mutated into amino acids with similar or opposite properties. Mutant proteins were further energy minimized, and each model was analyzed for its stability after mutation. Stable mutants were further used for docking analyses with various substrates. Binding scores obtained for WT and mutant protein for various substrates were compared and represented as a heatmap. Mutations showing a significant difference in the binding score as compared to wild proteins were further selected for in vitro studies.

Site-directed mutagenesis and mutant VvUGT expression

Site-directed mutagenesis of UGT88F12, 13 and 14 was carried out according to Schulenburg et al. (2016) following the QuikChange® protocol (Agilent Technology, Santa Clara, CA). Successful mutations were confirmed by sequencing. The plasmids were transformed into *Escherichia coli* BL21 (DE3) pLysS (Novagen, Darmstadt, Germany) for recombinant expression. The expression and purification of the proteins were performed according to Schulenburg et al. (2016) with small adjustments: LB medium for cultivation contained 100 μg mL^{-1} ampicillin and 34 μg mL^{-1} chloramphenicol. After growing the preparatory culture overnight, 200 mL of LB medium was inoculated with 2 mL of the preculture. The induction of protein expression and harvesting of the cultures were carried out following the protocol suggested by Schulenburg et al. (2016). To purify the recombinant glutathione S-transferase (GST) fusion proteins, a GST Bind Resin (Novagen) was used in accordance to Schulenburg et al. (2016). The binding buffer (4.3 mM Na$_2$HPO$_4$, 1.47 mM KH$_2$PO$_4$, 137 mM NaCl, 2.7 mM KCl, pH 7.3) contained 100 μM of phenylmethylsulfonyl fluoride as a proteinase inhibitor. After five elution steps with GST elution buffer (10 mM GST reduced, 50 mM Tris/HCl, pH 8.0), the total protein concentration of crude protein, wash fractions and all elution fractions were quantified via Bradford Assay (Schulenburg et al., 2016). For further activity assays, elution fractions with a total protein concentration $c > 450$ μg μL^{-1} were combined for each protein, respectively. The purity of the recombinant proteins was checked via SDS-PAGE (Supplemental data Figure S7).

LC-MS-based activity assay for wild and mutant VvUGTs

Substrate screens were carried out with a reaction mixture containing 7.5 μg recombinant protein, 300 μM of the respective substrate, 1 mM UDP glucose and 100 mM Tris–HCl buffer (pH 7.5). The reaction volume was made up to 200 μL and incubated at 30°C overnight at 400 rpm mixing. It was stopped by heating at 75°C for 10 min. After centrifugation at 13,200 rpm for 5 min, the supernatant was transferred to a fresh tube and centrifuged again. Finally, 50 μL of the supernatant was transferred to a LC-MS vial. As negative controls, all substrates were incubated with proteins of an empty vector control under the same conditions (Huang et al. 2018). After adding 5 μL of biochanin A (0.25 mM in MeOH) as standard for quantification, the mixture was subjected to LC-MS analysis. In order to distinguish between formed di-mono-glucosides and disaccharide-glycosides, additional assays were performed following the described

method but with 600 μM substrate and use of ^{13}C$_6$-labeled UDP glucose (Omicron Biochemicals, Inc.). Enzymatically formed products were quantified on an Agilent 1100 HPLC/UV system (Agilent Technologies) combined with a Bruker Esquire 300 plus ion trap mass spectrometer (Bruker Daltonics). Separation of the analytes was performed with a reverse-phase column (Luna 3u C18(2) 100 A, 150 × 2 mm; Phenomenex). As solvents, water (0.1% formic acid) and methanol (0.1% formic acid) were used. The gradient was operated according to Härtl et al. (2017).

Supplementary data

Supplementary data is available at *Glycobiology* online.

Conflict of interest

Authors have declared no conflict of interest.

Abbreviations

BcGT1, *Bacillus cereus*; glc, glucoside; GST, glutathione S-transferase; GTs, glycosyltransferases; LC-MS, liquid chromatography–mass spectrometry; PSPG, plant secondary product glycosyltransferase; RMSD, root mean square deviation; UDP-Glc, uridine diphosphate glucose; UGT, uridine diphosphate-dependent glycosyltransferases; *V. vinifera*, *Vitis vinifera*; VvUGTs, *V. vinifera* UGTs; WT, wild type

References

Ablajan K, Ablitz Z, Shang X-Y, He J-M, Zhang R-P, Shi J-G. 2006. Structural characterization of flavonol 3,7-di-O-glycosides and determination of the glycosylation position by using negative ion electrospray ionization tandem mass spectrometry. *J Mass Spectrom*. 41:352–360.

Anand P, Nagarajan D, Mukherjee S, Chandra N. 2013. ABS-Scan: In silico alanine scanning mutagenesis for binding site residues in protein-ligand complex. *F1000Research*. 3:214–214.

Bönisch F, Frotscher J, Stanitzek S, Rühl E, Wüst M, Bitz O, Schwab W. 2014a. A UDP-glucose: Monoterpenol glucosyltransferase adds to the chemical diversity of the grapevine metabolome. *Plant Physiol*. 165:561–581.

Bönisch F, Frotscher J, Stanitzek S, Rühl E, Wüst M, Bitz O, Schwab W. 2014b. Activity-based profiling of a physiologic aglycone library reveals sugar acceptor promiscuity of family 1 UDP-glucosyltransferases from grape. *Plant Physiol*. 166:23–39.

Bowles D, Lim EK, Poppenberger B, Vaistij FE. 2006. Glycosyltransferases of lipophilic small molecules. *Annu Rev Plant Biol*. 57:567–597.

Chang A, Singh S, Phillips GN Jr, Thorson JS. 2011. Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. *Curr Opin Biotechnol*. 6:800–808.

Chiu HH, Hsieh YC, Chen YH, Wang HY, Lu CY, Chen CJ, Li YK. 2016. Three important amino acids control the regioselectivity of flavonoid glucosidation in glycosyltransferase-1 from *Bacillus cereus*. *Appl Microbiol Biotechnol*. 100:8411–8424.

Chuang C-C, McIntosh MK. 2011. Potential mechanisms by which polyphenol-rich grapes prevent obesity-mediated inflammation and metabolic diseases. *Annu Rev Nutr*. 31:155–176.

Coutinho PM, Deleury E, Davies GJ, Henrissat B. 2003. An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol.* 328:307–317.

Fan B, Dong W, Chen T, Chu J, He B. 2018. Switching glycosyltransferase UGT$_{BL}$1 regioselectivity toward polydatin synthesis using a semi-rational design. *Org Biomol Chem.* 16:2464–2469.

Ford CM, Høj PB. 1998. Multiple glucosyltransferase activities in the grapevine *Vitis vinifera* L. *Aust J Grape Wine Res.* 4:48–58.

Giovinazzo G, Grieco F. 2015. Functional properties of grape and wine polyphenols. *Plant Foods Hum Nutr.* 70:454–462.

Härtl K, Huang FC, Giri AP, Franz-Oberdorf K, Frotscher J, Shao Y, Hoffmann T, Schwab W. 2017. Glucosylation of smoke-derived volatiles in grapevine *(Vitis vinifera)* is catalyzed by a promiscuous resveratrol/guaiacol glucosyltransferase. *J Agric Food Chem.* 65:5681–5689.

He XZ, Wang XQ, Dixon RA. 2006. Mutational analysis of the Medicago glycosyltransferase UGT71G1 reveals residues that control regioselectivity for (iso) flavonoid glycosylation. *J Biol Chem.* 281:34441–34447.

Huang FC, Giri A, Daniilidis M, Sun G, Härtl K, Hoffmann T, Schwab W. 2018. Structural and functional analysis of UGT92G6 suggests an evolutionary link between mono- and disaccharide glycoside-forming transferases. *Plant Cell Physiol.* 59:862–875.

Kim HJ, Kim BG, Ahn JH. 2013. Regioselective synthesis of flavonoid bisglycosides using Escherichia coli harboring two glycosyltransferases. *Appl Microbiol Biotechnol.* 97:5275–5282.

Lairson LL, Henrissat B, Davies GJ, Withers SG. 2008. Glycosyltransferases: Structures, functions, and mechanisms. *Annu Rev Biochem.* 77:521–555.

Liang DM, Liu JH, Wu H, Wang BB, Zhu HJ, Qiao JJ. 2015. Glycosyltransferases: Mechanisms and applications in natural product development. *Chem Soc Rev.* 44:8350–8374.

Liang D, Qiao J. 2007. Phylogenetic analysis of antibiotic glycosyltransferases. *J Mol Evol.* 64:342–353.

Lim EK, Ashford DA, Hou B, Jackson RG, Bowles D. 2004. Arabidopsis glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse quercetin glucosides. *Biotechnol Bioeng.* 87:623–631.

Luzhetskyy A, Bechthold A. 2008. Features and applications of bacterial glycosyltransferases: Current state and prospects. *Appl Microbiol Biotechnol.* 80:945–952.

Mattivi F, Guzzon R, Vrhovsek U, Stafanini M, Velasco R. 2006. Metabolite profiling of grape: Flavonols and anthocyanins. *J Agric Food Chem.* 54:7692–7702.

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS. Olson AJ. (2009) Autodock4 and AutoDockTools 4: Automated docking with selective receptor flexiblity. *J Comp Chem.* 16:2785–91.

Newman DJ, Cragg GM, Snader KM. 2003. Natural products as sources of new drugs over the period 1981-2002. *J Nat Prod.* 66:1022–1037.

Offen W, Martinez-Fleites C, Yang M, Kiat-Lim E, Davis BG, Tarling CA, Ford CM, Bowles DJ, Davies GJ. 2006. Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO J.* 25:1396–1405.

Ono E, Homma Y, Horikawa M, Kunikane-Doi S, Imai H, Takahashi S, Kawai Y, Ishiguro M, Fukui Y, Nakayama T. 2010. Functional differentiation of the glycosyltransferases that contribute to thechemical diversity of bioactive flavonol glycosides in grapevines *(Vitis vinifera)*. *Plant Cell.* 22:2856–2871.

Schulenburg K, Feller A, Hoffmann T, Schecker JH, Martens S, Schwab W. 2016. Formation of β-glucogallin, the precursor of ellagic acid in strawberry and raspberry. *J Exp Bot.* 67:2299–2308.

Schwab W, Fischer TC, Giri A, Wüst M. 2015. Potential applications of glucosyltransferases in terpene glucoside production: Impacts on the use of aroma and fragance. *Appl Microbiol Biotechnol.* 99:165–174.

Sharma R, Panigrahi P, Suresh CG. 2014. In-Silico analysis of binding site features and substrate selectivity in plant flavonoid-3-O glycosyltransferases (F3GT) through molecular modeling, docking and dynamics simulation studies. *PloS one.* 9(3):e92636.

Song C, Härtl K, McGraphery K, Hoffmann T, Schwab W. 2018. Attractive but toxic: Emerging roles of glycosidically bound volatiles and glycosyltransferases involved in their formation. *Mol Plant.* 11:1225–1236.

Trott O, Olson AJ. 2010. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 31:455–461.

Webb B, Sali A. 2014. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinformatics.* 47:5.6.1–5.6.32.